

# Insect Molecular Genetics

An Introduction to Principles  
and Applications



Third Edition



Marjorie A. Hay



# DNA, Gene Structure, and DNA Replication

## Chapter Outline

1.1 Overview	3
1.2 DNA is the Hereditary Material: A Brief History	4
1.3 The Central Dogma	6
1.4 The “RNA World” Came First?	8
1.5 The Molecular Structure of DNA	8
1.6 The Molecular Structure of RNA	11
1.7 The Double Helix	13
1.8 Complementary Base Pairing is Fundamental	14
1.9 DNA Exists in Several Forms	14
1.10 Genes	16
1.11 The Genetic Code for Protein-Coding Genes is a Triplet and is Degenerate	16
1.12 Gene Organization	18
1.13 Efficient DNA Replication is Essential	21
1.14 DNA Replication is Semiconservative	21
1.15 Replication Begins at Replication Origins	21
1.16 DNA Replication Occurs Only in the 5' to 3' Direction	22
1.17 Replication of DNA Requires an RNA Primer	23
1.18 Ligation of Replicated DNA Fragments	23
1.19 DNA Replication during Mitosis in Eukaryotes	25
1.20 Telomeres at the End: A Solution to the Loss of DNA during Replication	28
1.21 DNA Replication Fidelity and DNA Repair	28
1.22 Mutations in the Genome	29
1.23 Common Genetic Terminology	32
1.24 Independent Assortment and Recombination during Sexual Reproduction	33
General References	33
References Cited	34

## 1.1 Overview

Arthropod genes are made of deoxyribonucleic acid (DNA) and are located in chromosomes that consist of proteins, RNA, and DNA. DNA is a polymer of nucleotides (nt). Each nucleotide consists of a pentose sugar, one of four

nitrogenous bases, and a phosphoric acid component. DNA consists of two complementary strands in a helix form. Pairing of the nitrogenous bases adenine (A) with thymine (T) and cytosine (C) with guanine (G) on the two complementary strands occurs by hydrogen bonding. A pairs with T by two hydrogen bonds, and C pairs with G by three hydrogen bonds. DNA has chemically distinct 5' and 3' ends, and the two strands are antiparallel, with one strand running in the 5' to 3' direction and the other strand running in the 3' to 5' direction. The antiparallel orientation of the two strands creates a special problem when the DNA is duplicated or replicated during mitosis or meiosis.

Genetic information in protein-coding genes is determined by the sequence of nitrogenous bases (A, T, G, C) in one of the strands, with a three-base (triplet) codon designating an amino acid. The genetic code is degenerate, meaning that more than one codon specifies most amino acids. The genetic information is expressed when DNA is transcribed into pre-messenger RNA (pre-mRNA) that is processed into mRNA and then translated into polypeptides. Most insect genes have intervening noncoding sequences (introns) that must be removed from the primary RNA molecule before translation into the protein can occur.

Efficient and accurate replication of DNA must occur at each cell division, or the cell or organism may not survive. DNA replication is semiconservative, i.e., one of the nucleotide strands of each new DNA molecule is new and the other nucleotide strand is old in each “cell generation.” The new DNA strand is complementary to the parental (or template) strand. DNA replication occurs in one direction only, from the 5' to the 3' end of the strand, and thus replication takes place differently on the two antiparallel strands. Replication on the “leading strand” can occur in the 5' to 3' direction in a continuous manner. DNA replication on the other strand, the “lagging strand,” occurs in short segments (Okazaki fragments) because the DNA runs in the 3' to 5' direction. Subsequently, the Okazaki fragments must be ligated together. Replication of DNA in chromosomes begins at multiple sites called origins of replication along the chromosome, and it involves many enzymes and proteins. Although DNA replication is usually highly accurate, errors in DNA replication, or mutations, can result from duplications, deletions, inversions, and translocations of nucleotides, all of which may affect the functioning of the resultant polypeptide. New combinations of genes can occur through recombination during meiosis.

## 1.2 DNA is the Hereditary Material: A Brief History

Gregor Johann Mendel founded modern genetics in 1866 by publishing his studies on inheritance in garden peas. He confirmed that hereditary traits were transmitted from generation to generation, and he proposed the principles

of Segregation and Independent Assortment, which are discussed further in the description of meiosis and mitosis in Chapter 3. His work, however, was not widely known until 1900, when Hugo de Vries, Carl Correns, and Erich von Tschermak rediscovered these laws of inheritance. Mendel described traits in peas that were “dominant” or “recessive,” showed that peas could be selected for different traits, and showed that the traits were inherited in a stable manner.

The discovery that DNA is the hereditary material was first determined using a bacterium that causes pneumonia, *Streptococcus pneumoniae* ([Griffiths 1928](#)). Before this discovery, scientists speculated that the hereditary material might be composed of proteins or RNA. Proteins were considered the most likely hereditary material because they were known to be more variable (having 20 amino acids that could serve as the genetic code) than DNA. Furthermore, proteins are present in the nucleus in amounts nearly equal to DNA. DNA, by contrast, seemed to have only four types of structure (consisting of A, T, C, or G) that could serve as the genetic code. [Griffiths \(1928\)](#) found that nonvirulent forms of *S. pneumoniae* could be “transformed” to virulent forms by combining heat-treated virulent bacteria with nonvirulent bacteria. The reverse was true and led to the conclusion that the virulence traits were heritable and that the heritable material was capable of surviving mild heat treatment. Subsequently, [Avery et al. \(1944\)](#) conducted experiments in which the “transforming principle” was found to have the characteristics of DNA, and the transforming factors did not test positive for proteins or RNA. [Avery et al. \(1944\)](#) showed that enzymes that degrade proteins or RNA did not degrade the transforming principle but that enzymes that could degrade DNA did degrade the transforming principle. [Hershey and Chase \(1952\)](#) conducted experiments to further resolve whether protein or DNA was the hereditary material. They labeled DNA and protein from viruses that infect bacteria (bacteriophages) with different radioactive markers and monitored whether labeled DNA or labeled protein entered the bacterial host. Only labeled DNA entered the bacteria, confirming that the transforming principle, or genetic information, was contained in DNA.

The next big questions were how the DNA was structured, how the genetic information was encoded, and how the genetic information was replicated in a reliable manner. Answers to these questions were hotly pursued by several scientists, including Francis Crick, James Watson, Rosalind Franklin, Maurice Wilkins, Linus Pauling, and others. Rosalind Franklin and Maurice Wilkins provided critical information relevant to the solution of the structure of DNA with their X-ray diffraction pictures of purified DNA. The X-ray diffraction photographs provided an essential clue that allowed Watson and Crick to propose the correct structure

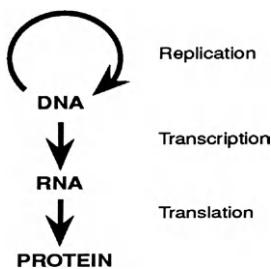
of DNA and to hypothesize how the genetic information was reliably replicated ([Watson and Crick 1953](#)). Previous proposals had been made that suggested that DNA “consists of three intertwined chains, with the phosphates near the fibre axis, and the bases on the outside.” Another three-chain structure also had been suggested in which “the phosphates are on the outside and the bases on the inside, linked together by hydrogen bonds.” [Watson and Crick \(1953\)](#) proposed that DNA “has two helical chains each coiled round the same axis... the bases on the inside of the helix and the phosphates on the outside...” and indicated the “novel feature of the structure is the manner in which the two chains are held together by the purine and pyrimidine bases...They are joined together in pairs, a single base from one chain being hydrogen-bonded to a single base from the other chain.” Watson and Crick stated, “It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material,” the accuracy of which is crucial to the transmission of genetic information from cell to cell and from generation to generation. The next big question to be answered involved the issue of how the purine and pyrimidine bases were able to encode the genetic information.

[Crick et al. \(1961\)](#) deciphered the genetic code, by showing that three bases of DNA code for one amino acid. As noted by [Crick et al. \(1961\)](#), “If the coding ratio is indeed 3 ... and if the code is the same throughout Nature, then the genetic code may well be solved within a year.” To resolve the question, [Crick et al. \(1961\)](#) found that mutations in a bacteriophage gene caused by the insertion or deletion of a single base pair resulted in a mutation that lead to a failure to produce a normal protein. The protein could be made functional again by inserting or deleting a total of three nucleotides, indicating that the genetic code uses a codon of three DNA bases that correspond to an amino acid and that the code for genes is not overlapping.

As more is learned about genomes, the concept of the gene has had to be modified. An early definition focused on protein-coding genes, but we now know that much of the DNA in an organism is transcribed into large and small RNAs that are not translated into proteins ([Collins and Penny 2009](#)). Furthermore, protein-coding genes have a variety of regulatory elements, including enhancers and promoters, whereas some genes code for RNAs that are used directly and other RNAs regulate development, as will be described in Chapters 2 and 3.

### 1.3 The Central Dogma

The **Central Dogma** of molecular biology, as proposed by Francis Crick ([Crick 1958](#)), stated that biological information is carried in DNA, this information



**Figure 1.1** The Central Dogma assumes that biological information transfers from DNA to RNA to proteins. Recent discoveries of viruses that transcribe information from RNA to DNA required modification in the Dogma. Three processes are involved in the Central Dogma: DNA replication, transcription of the genetic information into RNA, and translation of the mRNA into a polypeptide (protein).

subsequently is transferred to RNA (mRNA), and finally it is translated into specific proteins based on the code in the DNA. Initially, the Central Dogma stated that the flow of information is unidirectional, with proteins unable to direct the synthesis of RNA, and RNA unable to direct the synthesis of DNA ([Figure 1.1](#)).

The Central Dogma had to be amended in 1970 when certain viruses were found to transfer information from RNA to DNA. Subsequently, mutated proteins found in the membrane of brain cells of vertebrates were shown to be inherited. Although such aberrant proteins initially were thought to be caused by slow viruses or viroids, Stanley Prusiner discovered that the mutated proteins (called **prions**) could cause a group of fatal neurodegenerative diseases. The term *prion* refers to proteinaceous infectious particles ([Prusiner and Scott 1997](#)) that cause diseases such as bovine spongiform encephalopathy ("mad cow disease") in cattle, scrapie in sheep, and Creutzfeldt–Jakob disease or kuru in humans. These proteinaceous infective particles do *not* contain DNA, but they are able to transmit the disease to individuals who eat the altered proteins ([Prusiner and Scott 1997](#)). The altered protein acts as a template upon which the normal protein is refolded into a deformed molecule through a process facilitated by another protein ([Prusiner and Scott 1997, Tuite 2000](#)). Such abnormal proteins are transmitted to daughter cells, thereby propagating the mutant phenotype in the absence of any mutated nucleic acid.

The Central Dogma remains an important tenet of modern biology, although our knowledge of the roles of RNAs continues to expand and some have questioned its relevance ([Mattick 2009, Shapiro 2009](#)). In insects, genes (DNA) are found in complex structures called chromosomes that consist of proteins, RNAs, and DNA. This chapter reviews the structure of DNA and RNA, the basis of the

genetic code for proteins, the processes involved in DNA replication, and the changes in DNA that result in mutations.

## 1.4 The “RNA World” Came First?

It is thought that there could have been an era on early Earth during which RNA played the role of genetic material and also served as the main agent of catalytic activity, because RNA can serve as a ribozyme (DiGiulio 1997, Jeffares et al. 1998, Poole et al. 1998, Cooper 2000, Eddy 2001, Gesteland et al. 2006, Atkins et al. 2011, Darnell 2011). This role implies that enzymatic proteins in the modern world replaced RNA as the main catalysts. The “RNA organism” is thought to have had a multiple-copy, double-stranded RNA genome capable of recombination and splicing. The RNA genome was probably fragmented into “chromosomes” (Jeffares et al. 1998). RNA could have been the first genetic material because it can serve as a template for self-replication and can catalyze chemical reactions, including the polymerization of nucleotides (Johnston et al. 2001). It is thought that interactions between RNA and amino acids then evolved into the present-day world in which DNA is the more stable repository of genetic information. Knowledge of the number of RNAs and their very diverse functions has increased and an appreciation for the role of RNAs in gene regulation and development is reflected in many publications. First, however, let’s examine the structure and function of DNA in its role as a stable repository of genetic information.

## 1.5 The Molecular Structure of DNA

DNA is a long, double-stranded polymeric molecule consisting of individual monomers that are linked in a series and organized in a helix. Each monomer is called a **nucleotide**. Each nucleotide is itself a complex molecule made up of three components: a sugar, a nitrogenous base, and a phosphoric acid (Box 1.1).

In DNA, the sugar component is a pentose (with five carbon atoms) in a ring form that is called 2'-deoxyribose (Figure 1.2). The nitrogenous bases are single- or double-ring structures that are attached to the 1'-carbon of the sugar. The bases are **purines** (adenine and guanine) or **pyrimidines** (thymine and cytosine) (Figure 1.3). When a sugar is joined to a base it is called a **nucleoside**.

A nucleoside is converted to a nucleotide by the attachment of a **phosphoric acid group** to the 5'-carbon of the sugar ring (Figure 1.4). The four nucleotides that polymerize to form DNA are 2'-deoxyadenosine 5'-triphosphate, 2'-deoxyguanosine 5'-triphosphate, 2'-deoxycytidine 5'-triphosphate, and 2'-deoxythymidine 5'-triphosphate (Figure 1.5). These names usually are abbreviated as dATP, dGTP, dCTP, and dTTP, or shortened further to A, G, C and T, respectively.

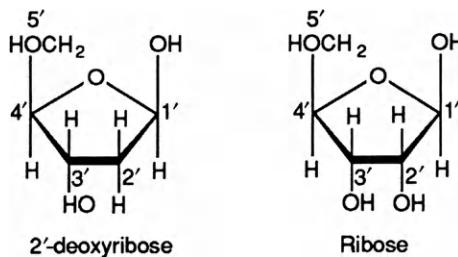
### Box 1.1 Key Points About DNA and RNA

#### DNA

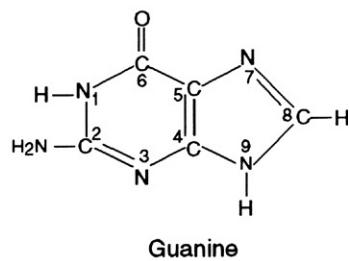
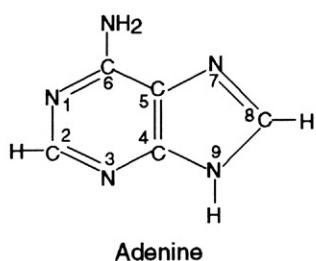
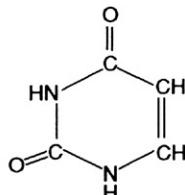
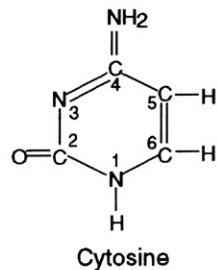
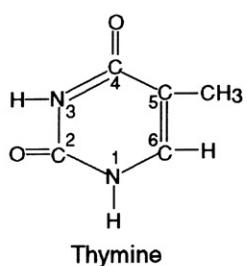
- DNA is a double helix with **antiparallel strands** consisting of **four types of nucleotides**.
- **Complementary base pairing** is essential to its structure and to maintaining an accurate code during replication.
- Adenine pairs with thymine (two hydrogen bonds) and guanine pairs with cytosine (three hydrogen bonds).
- **Replication is semiconservative**, a mechanism that helps to maintain the genetic information without errors. Daughter molecules contain one old and one new strand.
- DNA replication occurs only from the 5' to the 3' direction and requires an RNA primer.
- DNA replication begins at multiple origins of replication on the chromosome.
- Replication is continuous on the leading strand of DNA but is discontinuous on the lagging strand.
- The **genetic code consists of a triplet of bases** and is **degenerate**; 64 codons code for only 20 amino acids.
- **Protein-coding genes contain introns and exons in most eukaryotes**; the introns must be removed from pre-mRNA before the genetic information can be translated into proteins in eukaryotes.
- DNA is organized in **chromosomes with telomeres at their ends**, an organization that helps to preserve the ends of the chromosome.
- DNA can be mutated or damaged and repaired by multiple repair mechanisms.
- Mutations affect the phenotype of the insect in multiple ways.
- **DNA can be modified during meiosis by crossing over** and can result in new combinations of genes in the gametes.

#### RNA

- **RNA contains uracil** rather than thymine **and ribose sugar instead of deoxyribose**.
- **Pre-mRNA is processed in the nucleus** before the mRNA can move to the cytoplasm for translation.
- **Some RNAs are used directly** (ribosomal, transfer, and a variety of small RNAs) without being translated into proteins.



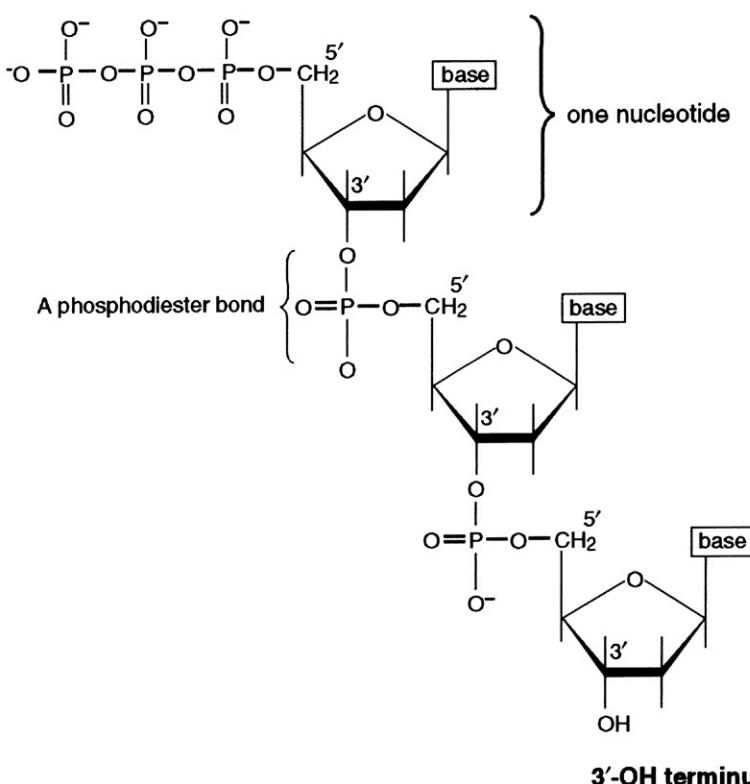
**Figure 1.2** Structure of sugars found in nucleic acids; 2'-deoxyribose is found in DNA and ribose is found in RNA.

**PURINES****PYRIMIDINES**

**Figure 1.3** Bases in DNA are purines (adenine and guanine) or pyrimidines (thymine and cytosine). Uracil is substituted for thymine in RNA.

Individual nucleotides are linked together by **phosphodiester bonds** to form polynucleotides (Figure 1.4). Polynucleotides have chemically distinct ends. In Figure 1.5, the top of the polynucleotide ends with a nucleotide in which the triphosphate group attached to the 5'-carbon has not participated in a phosphodiester bond. This end is called the 5' or 5'-P terminus. At the other end of the molecule, the unreacted group is not the phosphate, but the 3'-hydroxyl. This end is called the 3' or 3'-OH terminus. This distinction between the two ends (5' and 3') means that polynucleotides have an orientation that is very important in many molecular genetics concepts and applications.

Polynucleotides can be of any length and have any sequence of bases. The DNA molecules in chromosomes are probably several million nucleotides long.

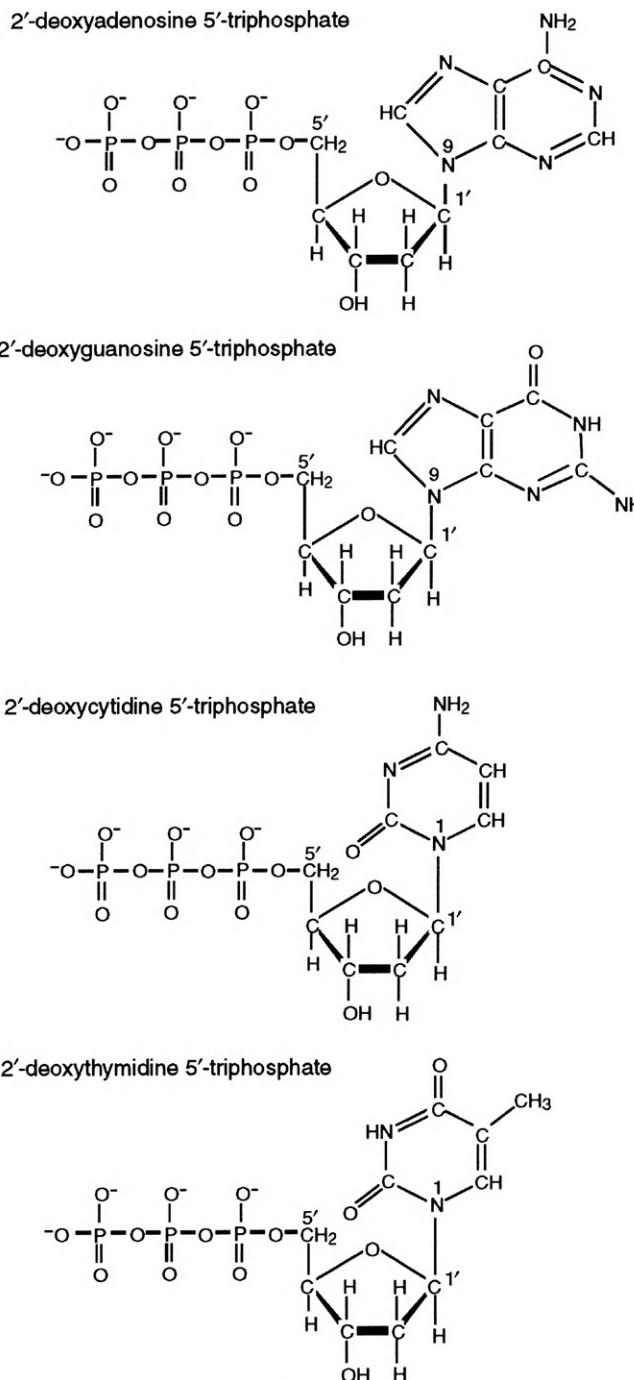
**5'-P terminus****3'-OH terminus**

**Figure 1.4** A nucleoside consists of a sugar joined to a base. It becomes a nucleotide (nt) when a phosphoric acid group is attached to the 5'-carbon of the sugar. Nucleotides link together by phosphodiester bonds to form polynucleotides.

Because there are no restrictions on the nucleotide sequence, a polynucleotide of just 10 nt long could have any one of  $4^{10}$  (or 1,048,576) different sequences. This ability to vary the sequence is what allows DNA to contain complex genetic information.

## 1.6 The Molecular Structure of RNA

RNA also is a polynucleotide and has multiple functions in the cell (Richter and Treisman 2011, Tuck and Tollervey 2011), including the role as mRNA. RNAs differ from DNA in two important ways. First, the sugar in RNA is **ribose** (Figure 1.2). Second, RNA contains the nitrogenous base **uracil** (U) instead of thymine (Figure 1.3). The four nucleotides that polymerize to form RNA are adenosine 5'-triphosphate, guanosine 5'-triphosphate, cytidine 5'-triphosphate, and uridine 5'-triphosphate, abbreviated as ATP, GTP, CTP, and UTP or A, G, C, and U,



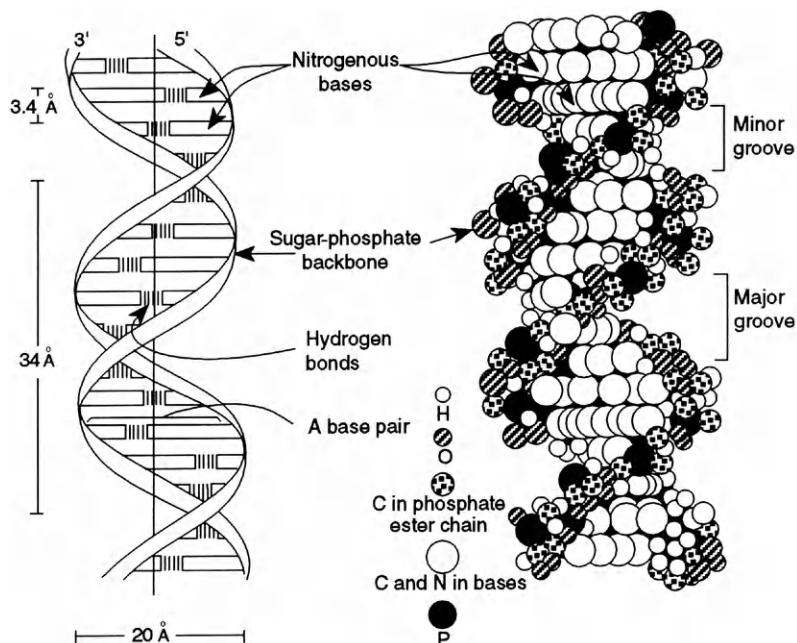
**Figure 1.5** The four nucleotides from which DNA is synthesized are 2'-deoxyadenosine 5'-triphosphate (dATP), 2'-deoxyguanosine 5'-triphosphate (dGTP), 2'-deoxycytidine 5'-triphosphate (dCTP), and 2'-deoxythymidine 5'-triphosphate (dTTP).

respectively. The individual nucleotides are linked together with 3' to 5' phosphodiester bonds. RNA is typically single-stranded, making it less stable than double-stranded DNA, although it can form complex structures (such as hairpins) or become double-stranded.

## 1.7 The Double Helix

The discovery, by [Watson and Crick \(1953\)](#), that DNA is a double helix of antiparallel polynucleotides ranks as one of the most important discoveries in biology because it provided a hypothesis as to how DNA replication could be achieved with a minimal error rate. Nitrogenous bases are located inside the double helix, with the sugar and phosphate groups forming the backbone of the molecule on the outside ([Figure 1.6](#)). The nitrogenous bases of the two polynucleotides interact by **hydrogen bonding**, with an A pairing to a T and a G to a C.

Hydrogen bonds are weak bonds in which two negatively charged atoms share a hydrogen atom between them. Two hydrogen bonds form between A and T and three hydrogen bonds form between G and C. Bonding between G and C is thus stronger, and more energy is required to break it. The hydrogen



**Figure 1.6** Two representations of the double-helix structure of DNA. The model on the left shows the hydrogen-bonding between nitrogenous bases that holds the two antiparallel strands together. The model on the right shows the relative sizes of the atoms in the molecule.

bonds, and other molecular interactions called stacking interactions, hold the double helix together.

The DNA helix turns approximately every 10 base pairs (bp), with spacing between adjacent base pairs of  $3.4\text{ \AA}$ , so that a complete turn requires  $34\text{ \AA}$  (Figure 1.6). The helix is  $20\text{ \AA}$  in diameter and right-handed, thus each chain follows a clockwise path. The strands run antiparallel to each other, with one strand running in the 5' to 3' direction and the other strand running in the 3' to 5' direction; this alignment has important implications for DNA replication and several molecular genetics techniques. The DNA helix has two grooves, a **major groove** and a **minor groove** (Figure 1.6) in which proteins involved in DNA replication and transcription interact with the DNA and with each other.

## 1.8 Complementary Base Pairing is Fundamental

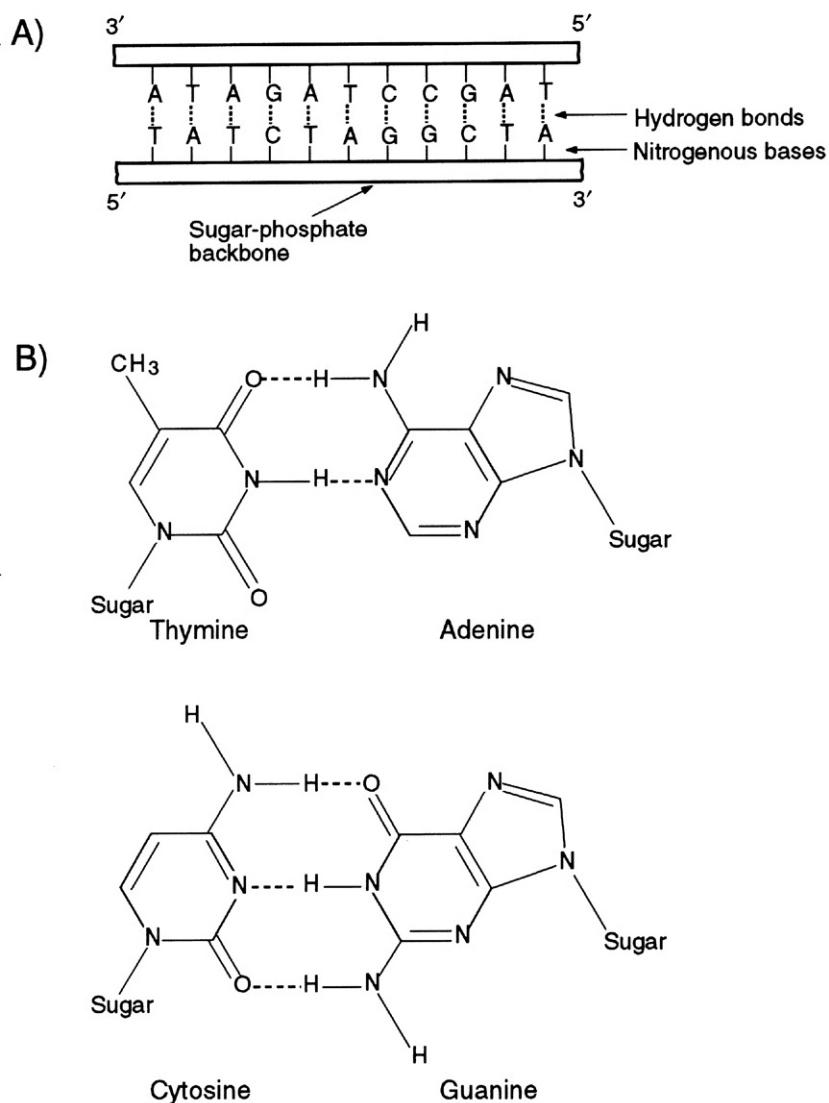
The principle of **complementary base pairing** is a fundamental element of DNA structure and of great practical significance in many techniques used in molecular biology. A pairs with T and G pairs with C. Normally, no other base-pairing pattern will fit in the helix or allow hydrogen bonding to occur (Figure 1.7, Box 1.1).

Complementary base pairing provides the mechanism by which the sequence of a DNA molecule is retained during replication of the DNA molecule, which is crucial if the information contained in the gene is not to be altered or lost during cell division. Complementary base pairing is important in transcription and expression of genetic information in the living insect and is important in several molecular techniques.

## 1.9 DNA Exists in Several Forms

DNA actually is a dynamic molecule in living organisms and has several different variations in form. In some regions of the chromosome, the strands of the DNA molecule may separate and later reanneal. DNA typically is right-handed, but it can form  $>20$  slightly different variations of right-handed helices. In some regions of the molecule, it can even form left-handed helices. If segments of nucleotides in the same strand are complementary, the DNA may even fold back upon itself in a hairpin structure.

DNA exists in different crystalline forms, depending upon the amount of water present in the DNA solution (Bustamante et al. 2003). B-DNA is the form in which DNA commonly occurs under most cellular conditions. A-DNA is more compact than B-DNA, with 11 bp/turn of the helix and a diameter of  $12\text{ \AA}$ .



**Figure 1.7** A) Complementary base pairing of polynucleotides by hydrogen bonds holds the two strands of the DNA molecule together. B) Thymine (T) pairs with adenine (A) with two hydrogen bonds, and guanine (G) pairs with cytosine (C) with three hydrogen bonds.

In addition, C-, D-, E-, and Z-DNA have been found. Z-DNA has a left-handed helix rather than a right-handed helix. A triple-helical form (H-DNA) also occurs. A, H, and Z forms of DNA are thought to occur in cells, and C, D, and E forms of DNA may be produced only under laboratory conditions.

## 1.10 Genes

The concept of a “gene” has evolved and has become increasingly difficult to define (Muller 1947, Maienschein 1992, Nelkin 2001, Pearson 2006, Pesole 2008, Brosius 2009). Genes can be a specific location on a chromosome (the bead on a string analogy), a particular type of biochemical material, and a physiological unit that directs development. Genes can consist of DNA sequences that can be spliced in alternative ways so that they are essentially coding for more than one protein (Pearson 2006). We also know that genes transcribed into RNA can regulate other genes, and many genes are never translated into proteins.

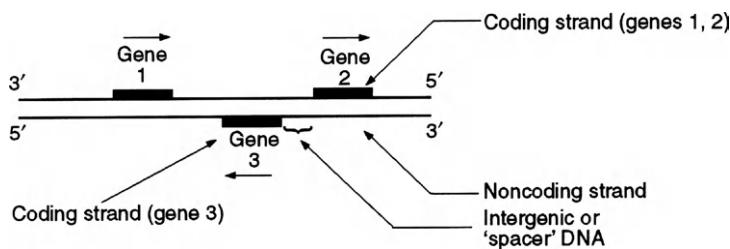
Genes are segments of a DNA molecule that may vary in size from as few as 75 nt to >200 kilobases (kb) of DNA. A kilobase is 1000 nt. Genes contain biological information by coding for the synthesis of an RNA molecule. The RNA may subsequently direct the synthesis of a protein molecule or the RNA may be the end product (e.g., transfer RNAs [tRNAs], ribosomal RNAs [rRNAs], regulatory RNAs). Proteins may regulate other genes, form part of the structure of cells, or function as enzymes. Expression of the information contained in protein-coding genes involves a two-step process of transcription and translation (Figure 1.1).

Genetic information is determined by one of the two strands of the double-helix DNA molecule. The DNA containing the genetic information is called the **coding strand**, and the other strand is the noncoding complement to it. Sometimes, the coding strand is known as the **sense** strand and the noncoding strand is known as the **antisense** strand. A few examples are known in which both strands are the “coding strand” for part of the length of the DNA molecule, but the genes occur in different specific regions. Thus, one strand of the double helix may be the sense strand over part of its length but be the antisense strand over other segments (Figure 1.8). A protein-coding gene typically includes a variety of regulatory structures and signals, as is described in Chapter 2.

Nonprotein-coding genes include genes that code for RNAs that are themselves the end products: the synthesized RNAs may be used directly as tRNAs, rRNAs, small nucleolar RNAs (snoRNAs), small nuclear RNAs (snRNAs), and other regulatory elements (Eddy 2001, Sharp 2009, Tuck and Tollervey 2011).

## 1.11 The Genetic Code for Protein-Coding Genes is a Triplet and is Degenerate

The genetic code for a protein-coding gene is based on the sequence of three nucleotides in the DNA molecule. The triplet sequence (or **codon**) determines which amino acids are assembled in a particular sequence into proteins. It is



**Figure 1.8** Genetic information is contained in genes carried on one of the two strands (coding strand). The complementary strand *in that region* is the noncoding strand. Genes can occur on different strands at different points of the DNA molecule. Noncoding DNA between genes is called intergenic or spacer DNA.

possible to order four different bases (A, T, C, G) in combinations of three into 64 triplets or codons. Because there are only 20 common amino acids, the question immediately arises as to what the other 44 codons do?

The answer is that the genetic code is **degenerate** with all amino acids, except methionine and tryptophan, determined by more than one codon (Table 1.1). A, U, C, and G represent the codons in Table 1.1 because the genetic information in DNA is transcribed into mRNA, which uses U instead of T.

The genetic code also contains punctuation codons. Three codons (UAA, UGA, and UAG) function as “stop” messages or **termination codons**; they occur at the end of a protein-coding gene to indicate where translation should stop. AUG serves as an **initiation or start codon** when it occurs at the front end of a gene. Because AUG is the sole codon for methionine, AUGs also are found in the middle of genes.

The genetic code is not universal, although it was assumed to be so initially. In 1979, it was found that mitochondrial genes use a slightly different code (Knight et al. 2001). For example, the codon AGA typically codes for arginine, but in *Drosophila* mitochondria the codon AGA codes for serine.

Eukaryotic genes have evolutionary histories and seem to have been derived from at least two sources. There are three domains of life: Archaea (archaeabacteria), Bacteria (eubacteria), and Eukarya (eukaryotes). **Eukaryotes** are organisms (including arthropods) that consist of cells with true nuclei bounded by nuclear membranes. Cell division in eukaryotes occurs by mitosis, reproductive cells undergo cell division by meiosis, and oxidative enzymes are packaged in mitochondria with its own circular chromosome. Evidence derived from analyses of genome sequences from the three domains strongly suggests that eukaryotic nuclear genes are derived from both archaeabacterial (informational genes) and

**Table 1.1: The 20 Amino Acids that Occur in Proteins and their Codons.**

Amino acid	Abbreviation	Codons					
Alanine	ala	GCU	GCC	GCA	GCG		
Arginine	arg	AGA	AGG				
Asparagine	asn	AAU	AAC				
Aspartic acid	asp	GAU	GAC				
Cysteine	cys	UGU	UGC				
Glutamic acid	glu	GAA	GAG				
Glutamine	gln	CAA	CAG				
Glycine	gly	GGU	GGC	GGA	GGG		
Histidine	his	CAU	CAC				
Isoleucine	ile	AUU	AUC	AUA			
Leucine	leu	UUA	UUG	CUU	CUC	CUA	CUG
Lysine	lys	AAA	AAG				
Methionine <sup>a</sup>	<u>met</u>	AUG					
Phenylalanine	phe	UUU	UUC				
Proline	pro	CCU	CCC	CCA	CCG		
Serine	ser	AGU	AGC				
Threonine	<u>thr</u>	ACU	ACC	ACA	ACG		
Tryptophan <sup>a</sup>	trp	UGG					
Tyrosine	try	UAU	UAC				
Valine	val	GUU	GUC	GUA	GUG		

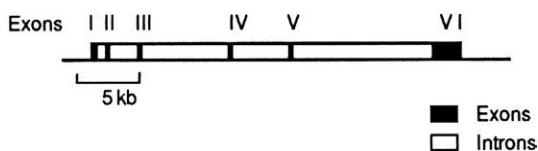
<sup>a</sup>Methionine and tryptophan are underlined because they are specified by only one codon.

eubacterial (operational genes) lineages, indicating that eukaryotic genomes are chimeric ([Lang et al. 1999](#), [Nesbo et al. 2001](#)).

## 1.12 Gene Organization

All genes are located on chromosomes. Each chromosome contains a single DNA molecule. These DNA molecules contain hundreds or thousands of genes in insects. For example, the fruit fly *Drosophila melanogaster* is estimated to have  $\approx$ 13,600 genes distributed on four chromosomes ([Adams et al. 2000](#)). Genes may be spaced out along the length of a DNA molecule, with DNA sequences intervening that do not code for proteins, or the genes may be grouped into clusters. Genes in a cluster may be related or unrelated to each other in structure and function. There are segments of DNA in eukaryotes in which the nucleotide sequences apparently do not code for anything; this DNA has been called “spacer” or intergenic DNA if it occurs between genes. Studies indicate that these noncoding sequences often are transcribed into RNAs that have gene regulatory functions.

**Multigene families** are clusters of related genes with similar nucleotide sequences. Multigene families may have originated from a single ancestral



**Figure 1.9** Protein-coding genes in eukaryotic organisms are divided into introns and exons. Introns are removed from the mRNA before it is translated into a polypeptide. In this example, there are six exons and four introns. The genetic message is present in exons I, II, III, IV, V, and VI.

gene that duplicated to produce two, or more, identical genes (Francino 2005). These identical genes could have diverged in nucleotide sequence through time to produce (two or more) related functional genes. In some cases, the genes of multigene families can be found at different positions on more than one chromosome after large-scale rearrangements (translocations or inversions) that occur both within and between chromosomes. Examples of multigene families in insects include *actins*, *tubulins*, *heat shock*, salivary glue, *chorion*, cuticle, and yolk protein genes. (Note that the name of a specific gene usually is italicized.)

**Pseudogenes** are DNA sequences that seem similar to those of functional genes, but the genetic information has been altered (mutated) so that the former gene is no longer functional. Once the biological information has been lost, a pseudogene can undergo rapid changes in nucleotide sequence and, given sufficient time, it may degrade to the point where it is not possible to identify the sequence as a former gene. At this point, it may be called “junk” DNA.

One of the interesting discoveries in genetics was the revelation in 1977 that most protein-coding genes in eukaryotes are discontinuous. Discontinuous genes contain coding and noncoding segments called **exons** and **introns**, respectively (Figure 1.9). Considerable discussion of the origin, evolution, and importance of introns has occurred previously (Herbert 1996, Gilbert et al. 1997, Trotman 1998). Introns have been maligned as examples of junk DNA because they may be considerably longer than the coding sequences (exons), and they were thought to have no function, although we now know that some introns contain regulatory sequences. Two hypotheses have been proposed to explain the origin of introns: the **introns-early** hypothesis and the **introns-late** hypothesis.

According to the **introns-early** hypothesis, many introns were present in the common ancestor of all life, but large or complete losses of the introns occurred in independent lineages. In addition, introns functioned in the primordial assembly of protein genes by promoting the recombination, or shuffling, of short exons, each encoding 15–20 amino acids (minigenes) into different

functional genes through fusion (Gilbert et al. 1997). It is likely that there has been an average of two or three acts of such fusions of minigenes into the larger exons of today (Gilbert et al. 1997). Some introns have been inherited for millions of years, making it possible to find a consistent location for the introns when homologous genes from different organisms are examined. The actual sequences of the introns in these homologous genes may have diverged through mutation to the point that they seem to have no sequence similarity. Trotman (1998) suggests that this consistent location of introns is evidence that introns may have been integral to the development of primordial genes, leading to the hypothesis that novel genes could arise from new combinations of exons, and thereby generate novel proteins and functions.

The **introns-late** hypothesis assumes that mechanisms for splicing introns out were not present in the common ancestor of life but that these mechanisms arose and spread within eukaryotes during their evolution. According to this hypothesis, introns could not have played a role in ancient gene and protein assembly. As is often the case with many “either/or” debates, the truth may be a combination of the two hypotheses (Tyshenko and Walker 1997). Both concepts may be correct; the introns in the *triosephosphate isomerase* genes of insects may be the result of the insertion of a transposable element relatively recently, whereas other introns may have been present for a very long time (Logsdon et al. 1995). DeSouza et al. (1998) suggest that 30–40% of the present-day intron positions in ancient genes correspond to the introns originally present in the ancestral gene. The rest of the intron positions are due to the movement or addition of introns over evolutionary time. Thus, introns may be both early and late, with ≈65% of the introns having been added to preexisting genes.

Introns are present in low frequency in prokaryotes and are rare in some eukaryotes, such as yeast. The number of introns and their lengths vary from species to species and from gene to gene. Some genes in eukaryotic organisms lack introns, whereas other genes in the same species may have as many as 50 introns. Introns may interrupt a coding region, or they may occur in the untranslated regions of the gene. Some eukaryotic genes contain numerous and very large introns, but introns typically range from 100 to 10,000 bp in length. A few introns contain genes themselves; how the genes got into the middle of an intron of another gene remains a mystery.

The presence of introns within many eukaryotic protein-coding genes requires that an additional step take place between transcription and translation in eukaryotes. Thus, when the DNA is transcribed into RNA, the initial RNA transcript is *not* mRNA. The synthesized RNA is a precursor to mRNA and is called

pre-mRNA. The pre-mRNA must undergo processing (splicing) in the nucleus to remove the introns before it can travel to the cytoplasm for translation into proteins. This process is described in Chapter 3, but first DNA replication is reviewed.

### 1.13 Efficient DNA Replication is Essential

Every living organism must make a copy of its genes in each cell each time the cell divides. Such replication ideally is both rapid and accurate. If not, the organism's survival and integrity are jeopardized. Even a very small error rate of 0.001% (one mistake/100,000 nt) can lead to detrimental changes or mutations. Although many mutations are detrimental, many apparently are neutral or nearly so, and a few are beneficial. The intrinsic structure of DNA helps to ensure that replication is accurate most of the time.

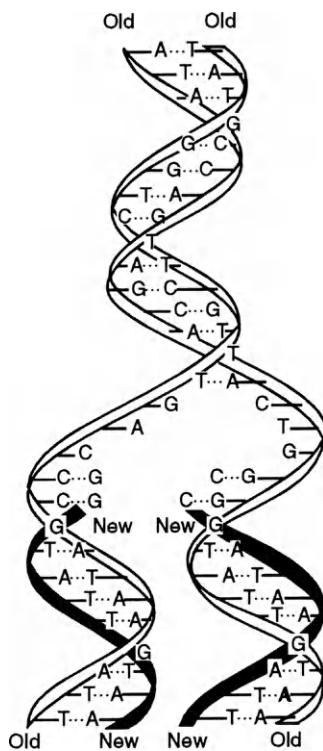
### 1.14 DNA Replication is Semiconservative

DNA replication is **semiconservative**, i.e., the daughter molecules each contain one polynucleotide derived from the original DNA molecule and one newly synthesized strand ([Figure 1.10](#), [Box 1.1](#)). Semiconservative DNA replication requires the hydrogen bonds that hold the two strands together be broken so that synthesis of new complementary strands can occur. Semiconservative replication of DNA increases the likelihood that replication error rates are very low.

### 1.15 Replication Begins at Replication Origins

During the replication of long DNA molecules, only a limited region of the DNA molecule is in an unpaired form at any one time. Replication occurs after the two strands separate; separation involves breaking the weak hydrogen bonds holding the bases of the opposite strands together. The separation of the two strands starts at specific multiple positions in the chromosome called **origins of replication** and moves along the molecule. Replication sites in *Drosophila*, for example, occur at thousands of sites throughout the genome ([Eaton et al. 2011](#)). Synthesis of the new complementary polynucleotides occurs as the double helix "unzips." The region at which the base pairs of the parent molecule are broken and the new polynucleotides are synthesized is the **replication fork** ([Figure 1.11](#)).

The two strands of the parent DNA molecule are broken apart by enzymes called **helicases**. Once the helicase has separated the two strands, **single-strand binding proteins** attach to the single strands to prevent them from immediately reannealing to each other ([Figure 1.11](#)). This attachment makes it possible for **DNA polymerase** to synthesize new complementary DNA strands. DNA

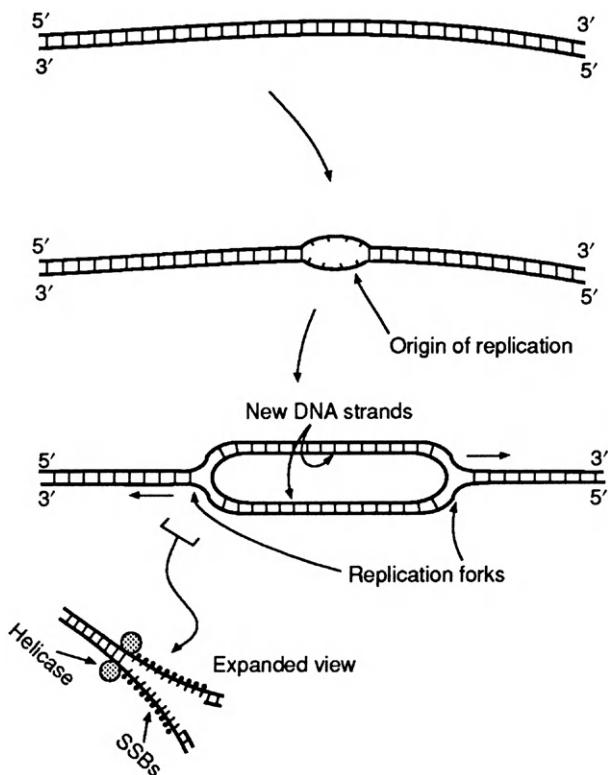


**Figure 1.10** DNA replication is semiconservative, meaning that each new DNA helix contains one old and one new complementary strand. DNA synthesis relies on complementary base pairing to replicate DNA accurately.

polymerases have two properties that complicate DNA synthesis. First, DNA polymerase can synthesize *only* in the 5' to 3' direction; and, second, DNA polymerase cannot initiate the synthesis of new DNA strands without a **primer**.

## 1.16 DNA Replication Occurs Only in the 5' to 3' Direction

Because DNA polymerases can synthesize DNA only in the 5' to 3' direction, the template strands must be read in the 3' to 5' direction (Box 1.1). This process is straight-forward for one of the DNA template strands, called the **leading strand**, and DNA synthesis can proceed in an uninterrupted manner the entire length of the leading strand. However, DNA synthesis cannot proceed uninterrupted on the other template strand, called the **lagging strand** (Figure 1.12). DNA synthesis on the lagging strand is discontinuous, occurring in short sections, and produces short fragments (100–200 nt in length) of DNA called **Okazaki fragments**, after their discoverer who identified them in 1968 (Ogawa and Okazaki 1980).



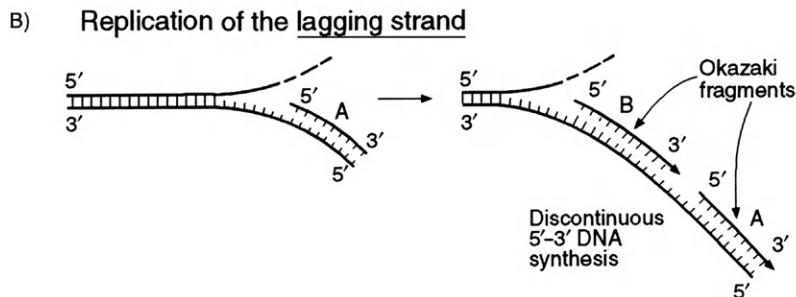
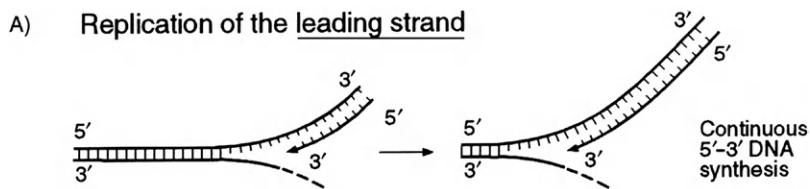
**Figure 1.11** During DNA replication, only part of the DNA molecule upzips to allow synthesis of new DNA strands. Helicases break the hydrogen bonds. In this example, replication begins at an origin of replication. Eukaryotes have many origins of replication along their chromosomes so that replication can occur rapidly. To keep the strands from reannealing at the replication forks where synthesis is occurring, single-strand binding proteins (SSBs) attach.

## 1.17 Replication of DNA Requires an RNA Primer

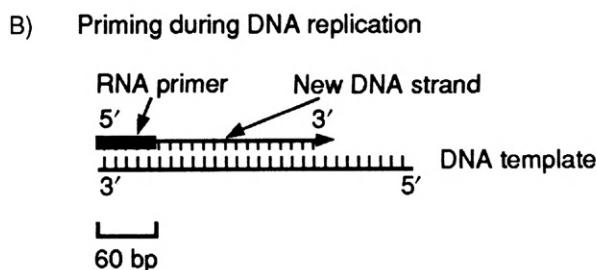
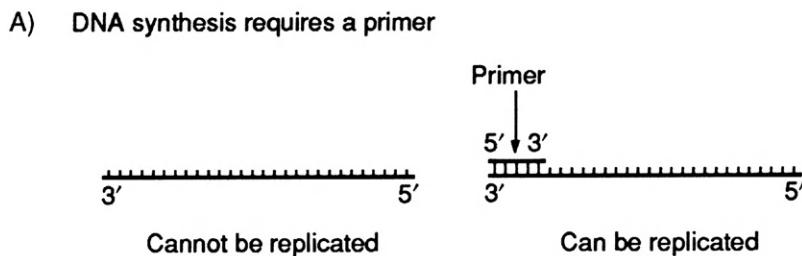
Another complication of DNA synthesis is that synthesis an RNA primer (Figure 1.13). Apparently, the first few (50–75) nucleotides attached to either the leading or lagging strands are not deoxyribonucleotides but rather ribonucleotides that are put in place by an RNA polymerase called **primase**. Once these ribonucleotides have been polymerized on the DNA template, the primase detaches, and DNA polymerase is able to synthesize DNA (Figure 1.14).

## 1.18 Ligation of Replicated DNA Fragments

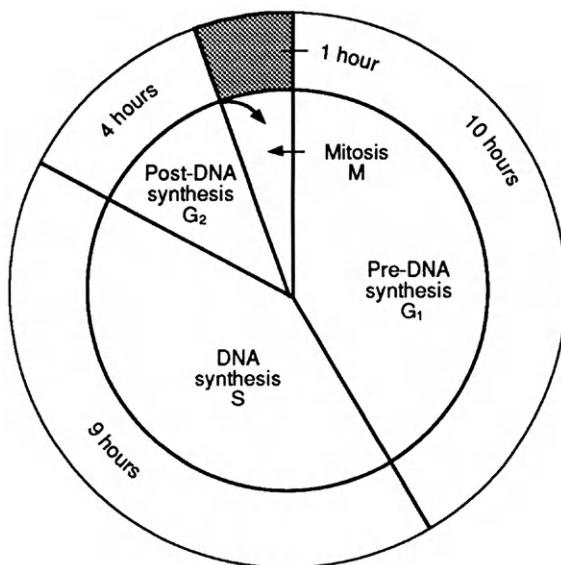
After the Okazaki fragments (sequences complementary to the lagging strand of DNA) are produced, they must be joined together to produce a continuous strand (Figure 1.12). On the lagging strand, DNA polymerase III of *Escherichia coli* stops



**Figure 1.12** DNA replication occurs in a different manner on the two strands. A) The leading strand is continuously copied, with synthesis occurring in the 5' to 3' direction. B) Synthesis on the lagging strand is discontinuous because the DNA strands are antiparallel. Synthesis occurs in short segments (Okazaki fragments) because DNA polymerase can only synthesize DNA in the 5' to 3' direction.



**Figure 1.13** A) DNA must be primed so that DNA polymerase is able to synthesize a complementary strand. B) A primer of ribonucleotides is attached to a strand by RNA polymerase. DNA polymerase can then attach deoxyribonucleotides (dNTPs) to the DNA template in a sequence that is determined by the template strand (complementary base pairing). DNA synthesis occurs in the 5' to 3' direction.



**Figure 1.14** The cell cycle of a eukaryotic cell with a generation time of  $\approx 24$  hours. DNA synthesis occurs during the S phase. During G<sub>1</sub> and G<sub>2</sub>, no DNA synthesis occurs. Mitosis (M) occurs after G<sub>2</sub>.

when it reaches the RNA primer at the 5' end of the next Okazaki fragment. Then, DNA polymerase I of *E. coli* removes the ribonucleotides from the Okazaki fragment and replaces them with deoxyribonucleotides. When all the ribonucleotides have been replaced, DNA polymerase I replaces nucleotides on a short distance into the DNA region, before it dissociates from the new double-helix molecule. The Okazaki fragments are then joined up by **DNA ligase** that catalyzes the formation of a phosphodiester bond between the neighboring nucleotides.

DNA replication also requires that the double helix be unwound, as well as unzipped. There are  $\approx 400,000$  turns in 400 kb of DNA. This unwinding is accomplished with the aid of enzymes called **DNA topoisomerases**. DNA topoisomerases unwind a DNA molecule without rotating the helix by causing short-term breaks in the polynucleotide backbone just in front of the replication fork. The reverse reaction is performed by DNA topoisomerases so that DNA molecules can be coiled.

## 1.19 DNA Replication during Mitosis in Eukaryotes

The goal is to replicate the genome once during mitosis and provide equivalent chromosomes to the daughter cells (Sclafani and Holzen 2007). The replication of prokaryotic and eukaryotic DNA is similar, but differs in several aspects, the details of which are not fully resolved (Gavin et al. 1995, Huberman 1995, Baker and Bell 1998, Leipe et al. 1999, Sutton and Walker 2001). DNA replication takes place during the eukaryotic cell cycle before the condensed metaphase

chromosomes become visible in mitosis or meiosis. DNA replication occurs at a rate of up to 1000 nt/second (Dixon 2009).

The **cell division cycle** consists of five distinct phases, if interphase is included. Nondividing cells are in interphase, the longest phase. Some cells remain in interphase and never divide. However, if cell division occurs, the chromosomes must be replicated in a precise manner, and any errors in replication must be detected and, ideally, corrected. There are two **gap phases** or periods ( $G_1$  and  $G_2$ ), when the cell is carrying out its normal metabolic activities (Figure 1.14), separated by the **S phase**, when DNA replication or synthesis occurs. **Mitosis** (M) occurs subsequent to the  $G_2$  phase (also known as the premitotic phase). Mitosis occurs when highly condensed duplicated chromosomes separate and segregate into daughter cells followed by cytokinesis in which the cell membrane forms around each daughter cell. During the  $G_2$  phase, cells grow rapidly and proteins and RNAs are synthesized. Mitosis is divided into several phases (prophase, metaphase, anaphase, and telophase), as described in Chapter 3.

To reduce the amount of time required to replicate the very long DNA molecule in eukaryotic chromosomes, DNA replication is initiated at a series of **replication origins**  $\approx$ 40 kb apart on the linear chromosome and proceeds in both directions (Figures 1.11, 1.12) (DePamphilis 1999). For example, replication in *D. melanogaster* occurs at a rate of  $\approx$ 2600 nt pairs/minute at 24 °C. The largest chromosome in *Drosophila* is  $\approx$ 8  $\times$  10<sup>7</sup> nt long, so, with  $\approx$ 8500 replication origins/chromosome,  $\approx$ 0.25–0.5 hour is required to replicate this chromosome. If replication occurred from a single replication fork, rather than from multiple replication origins, replication of a single chromosome would require  $\approx$ 15 days.

Origins of replication of eukaryotic chromosomes are recognized by a protein complex called the Origin of Replication Complex (ORC) that is essential for initiation of DNA replication at yeast origins. The protein complex opens the DNA, stabilizes the single-stranded DNA that is formed, and allows polymerases to copy the DNA. The ORC complex seems to recruit other proteins (including DNA helicases) to the origin of replication, leading to the start of replication. Proteins related to yeast ORC proteins have been identified in *Drosophila* and other eukaryotes (Gavin et al. 1995).

Eukaryotes contain several different DNA polymerases (Hubscher et al. 2002). One DNA polymerase is located in the mitochondria and is responsible for replication of mitochondrial DNA. The other DNA polymerases are in the nucleus and are involved in DNA replication, repair, and recombination. Polymerase  $\alpha$  complexes with **primase** (the RNA polymerase that primes DNA synthesis) and seems to function with primase to synthesize short RNA–DNA fragments. Two other polymerases then synthesize the leading and lagging strands, extending

the RNA–DNA primers initially synthesized by the polymerase  $\alpha$ -primase complex. A DNA polymerase fills the gaps between the Okazaki fragments after the primers are removed (Sutton and Walker 2001, Hubscher et al. 2002).

Proteins (called sliding-clamp proteins and clamp-loading proteins) act at the eukaryote replication fork to load the polymerase onto the primer and maintain its stable association with the template. The **clamp-loading proteins** (called replication factor C) recognize and bind DNA at the junction between the primer and template. The **sliding-clamp proteins** (proliferating cell nuclear antigen) in eukaryotes bind adjacent to the clamp-loading proteins, forming a ring around the template DNA. The clamp proteins then load the DNA polymerase onto the DNA at the primer-template junction.

The ring formed by the sliding clamp maintains the association of the polymerase with its template as replication progresses, allowing the uninterrupted synthesis of long DNA molecules. Helicases unwind the template DNA ahead of the replication fork. Single-stranded DNA-binding proteins (eukaryotic replication factor A) then stabilize the unwound template DNA so that the single-stranded DNA can be replicated. The enzymes involved in DNA replication, in combination with their accessory proteins, synthesize both leading and lagging strands of DNA simultaneously at the replication fork. The idea that DNA polymerases track like locomotives along the DNA template during DNA replication is pervasive and is probably based on the misperception that the polymerase is smaller than the DNA (Cook 1999). We now know that the DNA polymerase–protein complexes involved in DNA replication can be much larger than the DNA template.

An alternative model to the “movement” of polymerase along the DNA template has been proposed in which the fixed polymerase complexes “reel in their DNA templates” as they extrude newly made DNA in replication “foci” or replication factories within the cell. This “fixed” model assumes that the DNA polymerase complex is fixed and that the DNA rotates around it. This solution is a simple solution to the potential problem of untangling DNA strands that twine around each other if the DNA polymerase moves (Cook 1999).

Some DNA polymerases in eukaryotes have 3' to 5' exonuclease activity in addition to their polymerase activity, meaning that they can excise a misincorporated nucleotide by (proofreading) during DNA replication. DNA mismatch correction further minimizes replication errors by a survey of newly synthesized DNA strands. Furthermore, accessory factors such as DNA helicases apparently improve accuracy during DNA elongation, possibly due to resolution of stalled replication forks. Despite all these precautions, occasional misincorporated nucleotides or deletions, or insertions may remain, resulting in mutations.

## 1.20 Telomeres at the End: A Solution to the Loss of DNA during Replication

Because DNA synthesis occurs exclusively in the 5' to 3' direction and initiation requires a short RNA primer, the extreme 5' end of a linear DNA strand consists of an RNA primer (Figure 1.13B). If the RNA primer is not replaced by deoxyribonucleotides, the chromosome would gradually decrease in length after each replication during mitosis because the segment with the RNA primer would not be copied into DNA. Shortening of the chromosome by 50–200 bp at the 3' end of the lagging strand in each cell cycle could seriously affect gene function over time. However, linear chromosomes normally are stable because they have a specialized structure at their ends called a telomere (Zakian 1989). Without telomeres, chromosomes are sticky and could fuse with other chromosomes, resulting in growth arrest and cell death (Verdun and Karlseder 2007). The discovery of telomeres in 1978 resulted in a 2009 Nobel Prize in Physiology or Medicine to Elizabeth Blackburn, Carol Greider, and Jack Szostak.

**Telomeres** contain a series of species-specific repeated nucleotide sequences that are added to the ends of eukaryotic linear chromosomes by an enzyme called **telomerase**. In many arthropods, the highly repeated telomeric sequence has the motif of TTAGG and the telomeres may consist of 4–6 kb of sequence. Vitkova et al. (2005) found these sequences in Diplura, Collembola, crustaceans, myriapods, pycnogonids, and most chelicerates (except spiders). Telomerase is a **reverse transcriptase**, meaning that it can transcribe DNA from an RNA template. A few copies of a short repetitive sequence (called the telomere sequence) are required to prime the telomerase to add additional copies to form a telomere. There are also longer, moderately repetitive nucleotide sequences subterminal to the telomere sequences (subtelomeric region). Although the telomeres are maintained by telomerase during cell divisions, telomeres do shorten through time and, in vertebrates, shortened telomeres are correlated with aging (Aubert and Lansdorp 2008).

## 1.21 DNA Replication Fidelity and DNA Repair

Faithful maintenance of the genome is crucial to both the individual and the species (Lindahl and Wood 1999). When DNA is replicated inaccurately or is damaged by endogenous factors (such as water or oxygen) or exogenous factors (such as UV light, chemicals, and irradiation), death can ensue. Thus, there has been strong selection for multiple mechanisms to repair damaged DNA. Generally, the cell has two classes of mechanisms with which to repair DNA: 1) direct repair and 2) excision repair, or removal of the damaged bases followed by their replacement with newly synthesized DNA.

Two types of damage, DNA damaged caused by UV light and modifications of G by the addition of methyl or ethyl groups to the sixth oxygen position of the purine ring, are repaired directly (**direct repair**) (Cooper 2000). The most common repair mechanism in cells involves removal (excision) of damaged components of the DNA. Repair can be divided into the following excision repair systems: base-excision repair, nucleotide-excision repair, and mismatch repair. **Base-excision** repair involves removal of only the damaged base from the DNA strand. **Nucleotide-excision** repair operates mainly on damage caused by environmental mutagens and involves DNA synthesis and ligation to replace an excised oligonucleotide (Lindahl and Wood 1999). In **mismatch repair**, the mismatched bases that are incorporated during replication occasionally are not removed by the proofreading activity of DNA polymerase. The mismatched bases that are not removed are corrected by the mismatch repair system. If the DNA is not repaired before replication by the above-mentioned mechanisms, a postreplication repair system comes into play. **Postreplication repair** (recombinational repair) can repair several types of damage to DNA, including double-strand breaks introduced into DNA by irradiation.

## 1.22 Mutations in the Genome

Changes in the genetic material (**genotype**) of an organism occur if DNA repair is not successful. Such changes are **mutations**. Many kinds of mutations can occur: within an exon, within introns, or in the chromosomal regions (**intergenic regions**) located between the genes. If a mutation occurs in an intergenic region, it may be silent and have no detectable effect on the cell or individual. If a mutation occurs in an exon, it may alter protein product and cause a change in the organism's **phenotype** (or appearance). A mutation in an intron may not have an effect on the phenotype, but it could have an effect if there are regulatory elements in the intron that are important for proper gene function.

An organism with the "normal" appearance (phenotype) for that species is called the "**wild type**," whereas an organism with a phenotype that has been changed is a **mutant**. If the mutation is **dominant** (meaning that only a single copy is required to cause the change in phenotype), the name of the gene is capitalized. If the mutation is **recessive** (meaning that both copies of the gene carry the mutation), the name is not capitalized.

A **mutagen** is a chemical or physical agent that causes changes in bases. Mutagens include UV radiation, X-irradiation, ethyl methane sulfonate, base analogs such as 5-bromouracil, acridine dyes, and nitrous acid. Mutations occur spontaneously approximately once in every  $10^8$  bp/cell division, or they can be induced by the experimenter.

Mutations affect the DNA sequence, gene organization, gene regulation, or gene function (Table 1.2). A **point mutation** is the replacement of one nucleotide by another (substitution). A substitution can be either a transition or a transversion. **Transitions** involve changes between A and G (purines) or T and C

**Table 1.2: Mutations Affect DNA Sequence, Gene Function, Gene Regulation, and the Phenotype of the Organism.**

<b>Changes in DNA sequence</b>	
Point mutation	Replacement of one nucleotide by another.
Transition	A point mutation in which a purine is changed to a purine (A <--> G) or a pyrimidine to a pyrimidine (T<-->C).
Transversion	A point mutation in which the change is purine to pyrimidine, or vice versa (A or G <--> T or C).
<b>Changes in the gene</b>	
Silent mutation	Sequence changes in an intergenic region usually result in no phenotypic changes. Changes in a gene can be silent if a point mutation occurs in the third nucleotide of a codon that, because of the degeneracy of the code, does not alter the amino acid.
Nonsense mutation	A point mutation that alters a codon specifying an amino acid into a termination codon, which will prematurely terminate the polypeptide produced, changing the activity of the protein and altering the phenotype.
Frameshift mutation	Insertions or deletions that are not in multiples of three can cause changes in the amino acids downstream from the mutation, resulting in a mutant phenotype.
Insertions or deletions (indels)	When nucleotides are inserted or deleted (indels), the resulting mutations can be benign to lethal, depending on where these modifications occur.
Nonsynonymous or missense	Changes in codons that alter the amino acid specified.
<b>Changes in gene regulation</b>	
	Mutations in regulatory genes alter the organism's ability to control expression of a gene normally subject to regulation.
<b>Changes in the organism</b>	
Lethal mutation	Mutations that alter the function of an essential gene product so that the organism cannot survive.
Conditional lethal	Individuals with these mutations can survive under a particular set of conditions, such as a specific temperature range, but die if reared outside these conditions.
Back mutation	Organisms sometimes revert to the wild-type phenotype after a second mutation occurs that restores the original nucleotide sequence of the mutated gene.
Reversion	Mutations can be corrected by restoring the original phenotype, but not the original DNA sequence, in the mutated gene by altering a second site within the gene.
Suppression	The effects of a mutation can be altered by a new mutation that occurs in a different gene.

(pyrimidines), whereas **transversions** involve changes between a purine and a pyrimidine.

An **insertion** or **deletion** (also called **indels**) involves the addition or deletion of one or more nucleotides. An **inversion** is the excision of a part of the DNA molecule followed by its reinsertion into the same position but with a reversed orientation. An inversion in *Drosophila buzzatii* was caused by a transposable element called *Galileo*, and this mechanism may be the mechanism by which many inversions occur (Caceres et al. 1999).

Some mutations are lethal, whereas others have an effect on the organism that can range from phenotypically undetectable (silent) to lethal only under certain circumstances (**conditional lethal**). For example, many mutations are temperature-sensitive, and the organism can survive if reared within one temperature range but will die if reared at higher temperatures.

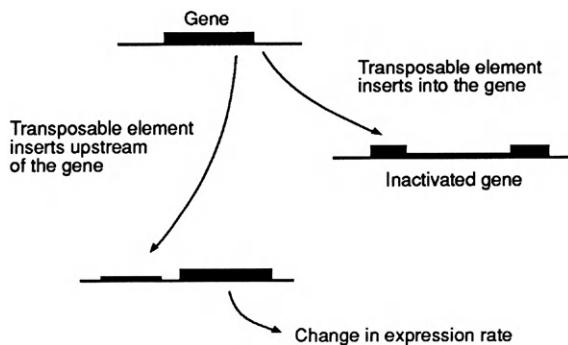
A **silent mutation** may occur if the third base in a codon is altered but, because the genetic code is degenerate, there is no change in the amino acid specified. These mutations are called **synonymous mutations** because there is no change in protein structure or function from a silent mutation.

Some changes in codons alter the amino acid specified, and they are called **nonsynonymous** or **missense mutations**. Most point mutations that occur at the first or second nucleotide positions of a codon will be missense, as will a few third-position changes. A polypeptide with an amino-acid change may result in a changed phenotype, depending on the precise role the altered amino acid plays in the structure or function of the polypeptide. Most proteins can tolerate some changes in their amino-acid sequence if the alteration does not change a segment of the polypeptide essential for its structure or function.

**Nonsense mutations** are point mutations that change a codon specifying an amino acid into a termination codon, producing a truncated gene that codes for a polypeptide that is terminated prematurely. In many cases, essential amino acids will be deleted, and the protein's activity will be altered, resulting in a mutant phenotype.

**Frameshift mutations** result if additions or deletions of base pairs occur that are not in a multiple of three. The polypeptide produced will likely have a complete new set of amino acids produced downstream of the frameshift. Frameshifts usually produce mutant phenotypes.

Occasionally, **back mutations** may occur to reverse a point mutation. **Reversions** sometimes occur when the original phenotype is restored by a new



**Figure 1.15** Movement (transposition) of transposable elements into chromosomes can result in mutations that inactivate genes or alter their expression. Transposable elements are also known as jumping genes.

change in the nucleotide sequence. In reversions, the original mutation is not restored to its previous unmutated form; rather, the second mutation restores the code for the original amino acid because the code is degenerate. **Regulatory mutations** are mutations that affect the ability to control expression of a gene.

The movement of a **transposable element** (TE) into a gene can also create mutations in genes. TEs are segments of foreign DNA that can move into genomes. When TEs move into a gene, as is shown in [Figure 1.15](#), the gene will be inactivated or the gene product will be altered and produce in a visible phenotype (mutation). TEs can cause other types of mutations, including inversions ([Caceres et al. 1999](#)). TEs are found in most eukaryotic organisms, and there are many types. TEs are important for understanding genome evolution and for genetic engineering, and they are discussed further in Chapters 3, 4, 9, and 14.

## 1.23 Common Genetic Terminology

A **wild-type gene** is normally identified only after a mutation has disrupted the phenotype of an organism. Mutations commonly are given a descriptive name, such as “white eyes.” The name of the gene usually is italicized (*white*) and is abbreviated using one, two, or three italicized letters (such as *w*). If the mutation is dominant, the name and abbreviation are capitalized (*White* and *W*); they are in lowercase (*white*) if the mutation is recessive. Individuals that are homozygous for the recessive *w* mutation are *w/w* and have white eyes. Heterozygous flies are *w/w<sup>+</sup>*, with the wild-type allele designated as *w<sup>+</sup>*, and their appearance (phenotype) should be wild type. The gene product is called the white product or white protein and is not italicized. The term for the gene product may be abbreviated as the *w* protein. Sometimes the protein product is designated by the gene name but capitalized to distinguish it from the gene (*WHITE*).

## 1.24 Independent Assortment and Recombination during Sexual Reproduction

For organisms to survive and evolve with changing environmental conditions, they need to be able to generate genetic variability. Mutations are one source of genetic variability and thus are not always undesirable. Another source of genetic variability is the result of sexual reproduction.

In sexually reproducing organisms, the progeny produced by parents that have different versions of genes (different alleles, AA or aa) will have a different combination of alleles (Aa). This shuffling of the genetic information during sexual reproduction is due to the **independent assortment** of chromosomes into the gametes during **meiosis**. Thus, an individual of genotype Aa Bb, in which the genes A and B are located on different chromosomes, will produce equal numbers of four different types of gametes: AB, Ab, aB, or ab.

**Crossing over** also leads to recombination between DNA molecules. Crossing over occurs between homologous chromosomes during the production of eggs or sperm in meiosis I, and results in an exchange of genetic material. Crossing over allows new combinations of different genes that are linked (located on the same chromosome). Thus, if a parent has one chromosome containing with A and B, and the homologous chromosome has a and b alleles, a physical exchange between the chromatids during meiosis I can lead to gametes that have the following combinations: A and B, A and b, a and B, and a and b. Nonhomologous recombination, crossing over between DNA lacking sequence homology, also may occur. Meiosis and mitosis are described further in Chapter 3.

## General References

- Brown, T.A., (2011). *Introduction to Genetics: A Molecular Approach*. Garland Science, New York.
- Carlson, E.A., (2004). *Mendel's Legacy. The Origin of Classical Genetics*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Cooper, G.M., and Hausman, R.E., (2009). *The Cell. A Molecular Approach*, 5th ed. ASM Press, Washington, DC.
- Deamer, D., and Szostak, J.W., (2010). *The Origins of Life*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- DePamphilis, M.L., and Bell, S.D., (2010). *Genome Duplication: Concepts, Mechanisms, Evolution, and Disease*. Garland Science, New York.
- Friedberg., E.C., (1997). *Correcting the Blueprint of Life: An Historical Account of the Discovery of DNA Repair Mechanisms*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Haga, S.B., (2006). Teaching resources for genetics. *Nat. Rev. Genet.* 7: 223–229.
- Hartl, D.L., and Jones, E.W., (1998). *Genetics: Principles and Analysis*. Jones and Bartlett, Sudbury, MA.
- Micklos, D.A., and Freyer, G., (2010). *DNA Science: A First Course*, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

- Perdew, G.H., Van den Heuvel, J.P., and Peters, J.M., (2006). *Regulation of Gene Expression*. Humana Press, Totowa, NJ.
- Russell, P.J., (2010). *iGenetics. A Molecular Approach*, 3rd ed. Benjamin Cummings, San Francisco, CA.
- Watson, J.D., Baker, T.A., Bell, S.P., Gann, A., Levine, M., and Losick, R., (2008). *Molecular Biology of the Gene*, 6th Edition. Benjamin Cummings, Menlo Park, CA.

## References Cited

- Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D., and Amanatides, P.G., et al. (2000). The genome sequence of *Drosophila melanogaster*. *Science* **287**: 2185–2195.
- Atkins, J.F., Gesteland, R.F., and Cech, T.R., (2011). *RNA Worlds: From Life's Origins to Diversity in Gene Regulation*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Aubert, G., and Lansdorp, P.M., (2008). Telomeres and aging. *Physiol. Rev.* **86**: 567–579.
- Avery, O.T., Macleod, C.M., and McCarty, M., (1944). Studies on the chemical nature of the substance inducing transformation of pneumococcal types. *J. Exp. Med.* **79**: 137–158.
- Baker, T.A., and Bell, S.P., (1998). Polymerases and the replisome: machines within machines. *Cell* **92**: 295–305.
- Brosius, J. (2009). The fragmented gene. In: *Natural Genetic Engineering and Natural Genome Editing*. G. Witzany, Ed. *Ann. N. Y. Acad. Sci.* **1178**: 186–193.
- Bustamante, C., Bryant, Z., and Smith, S.B., (2003). Ten years of tension: single-molecule DNA mechanics. *Nature* **421**: 423–426.
- Caceres, M., Ranz, J.M., Barbadilla, A., Long, M., and Ruiz, A., (1999). Generation of a widespread *Drosophila* inversion by a transposable element. *Science* **285**: 415–418.
- Collins, L.J., and Penny, D., (2009). The RNA infrastructure: dark matter of the eukaryotic cell? *Trends Genet.* **25**: 120–128.
- Cook, P.R., (1999). The organization of replication and transcription. *Science* **284**: 1790–1795.
- Cooper, G.M., (2000). *The Cell: A Molecular Approach*, 2nd ed. ASM Press, Washington, DC.
- Crick, F.H.C., (1958). On protein synthesis. *Symp. Soc. Exp. Biol.* **XII**: 139–163.
- Crick, F.H.C., Barnett, L., Brenner, S., and Watts-Tobin, R.J., (1961). General nature of the genetic code for proteins. *Nature* **4809**: 1227–1232.
- Darnell, J.E., (2011). *RNA: Life's Indispensable Molecule*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- DePamphilis, M.L., (1999). Replication origins in the metazoan chromosomes: fact or fiction? *Bioessays* **21**: 5–16.
- DeSouza, S.J., Long, M., Klein, R.J., Roy, S., Lin, S., and Gilbert, W., (1998). Toward a resolution of the introns early/late debate: only phase zero introns are correlated with the structure of ancient proteins. *Proc. Natl. Acad. Sci. USA* **95**: 5094–5099.
- DiGiulio, M., (1997). On the RNA world: evidence in favor of an early ribonucleopeptide world. *J. Mol. Evol.* **45**: 571–578.
- Dixon, N.E., (2009). Prime-time looping. *Nature* **462**: 854–856.
- Eaton, M.L., Prinz, J.A., MacAlpine, H.K., Tretyakov, G., Kherchenko, P.V., and MacAlpine, D.M., (2011). Chromatin signatures of the *Drosophila* replication program. *Genome Res.* **21**: 164–174.
- Eddy, S.R., (2001). Non-coding RNA genes and the modern RNA world. *Nat. Rev. Genet.* **2**: 919–929.
- Francino, M.P., (2005). An adaptive radiation model for the origin of new gene functions. *Nat. Genet.* **37**: 573–577.
- Gavin, K.A., Hidaka, M., and Stillman, B., (1995). Conserved initiator proteins in eukaryotes. *Science* **270**: 1667–1671.

- Gesteland, R.F., Cech, T.R., Atkins, J.F. (Eds.), 2006. *Cold Spring Harbor Monograph Series 43*, 3rd ed. Cold Spring Harbor Laboratory Press. Cold Spring Harbor, NY.
- Gilbert, W., DeSouza, S.J., and Long, M., (1997). Origin of genes. *Proc. Natl. Acad. Sci. USA* **94**: 7698–7703.
- Griffiths, F., (1928). The significance of pneumococcal types. *J. Hyg.* **27**: 113–159.
- Herbert, A., (1996). RNA editing, introns and evolution. *Trends Genet.* **12**: 6–8.
- Hershey, A., and Chase, M., (1952). Independent functions of viral protein and nucleic acid in growth of bacteriophage. *J. Gen. Physiol.* **36**: 39–56.
- Huberman, J.A., (1995). Prokaryotic and eukaryotic replicons. *Cell* **82**: 535–542.
- Hubscher, U., Maga, G., and Spadari, S., (2002). Eukaryotic DNA polymerases. *Annu. Rev. Biochem.* **71**: 133–163.
- Jeffares, D.C., Poole, A.M., and Penny, D., (1998). Relics from the RNA world. *J. Mol. Evol.* **46**: 18–36.
- Johnston, W.K., Unrau, P.J., Lawrence, M.S., Glasner, M.E., and Bartel, D.P., (2001). RNA-catalyzed RNA polymerization: accurate and general RNA-templated primer extension. *Science* **292**: 1319–1325.
- Knight, R.D., Landweber, L.F., and Yarus, M., (2001). How mitochondria redefine the code. *J. Mol. Evol.* **53**: 299–313.
- Lang, F., Gray, M.W., and Burger, G., (1999). Mitochondrial genome evolution and the origin of eukaryotes. *Annu. Rev. Genet.* **22**: 351–397.
- Leipe, D.D., Aravind, L., and Koonin, E.V., (1999). Did DNA replication evolve twice independently? *Nucleic Acids Res.* **27**: 3389–3401.
- Lindahl, T., and Wood, R.D., (1999). Quality control by DNA repair. *Science* **286**: 1897–1905.
- Logsdon Jr., J.M., Tyshenko, M.G., Dixon, C., Jafari, J.D., Walker, V.K., and Palmer, J.D., (1995). Seven newly discovered intron positions in the triose-phosphate isomerase gene: evidence for the introns-late theory. *Proc. Natl. Acad. Sci. USA* **92**: 8507–8511.
- Maienschein, J., (1992). Gene: historical perspectives. In: Fox-Keller, E., Lloyd, E.A. (Eds.), *Keywords in Evolutionary Biology* Harvard University Press. Cambridge, MA, pp. 122–127.
- Mattick, J.S., (2009). Deconstructing the dogma: a new view of the evolution and genetic programming of complex organisms. *Natural Genetic Engineering and Natural Genome Editing. Ann. N. Y. Acad. Sci* **1178**: 29–46.
- Muller, H.J., (1947). The gene. *Proc. R. Soc. Lond. B* **134**: 1–37.
- Nelkin, D., (2001). Molecular metaphors: the gene in popular discourse. *Nat. Rev. Genet.* **2**: 555–559.
- Nesbo, C.L., Boucher, Y., and Doolittle, W.F., (2001). Defining the core of nontransferable prokaryotic genes: the euryarchaeal core. *J. Mol. Evol.* **53**: 340–350.
- Ogawa, T., and Okazaki, T., (1980). Discontinuous DNA replication. *Annu. Rev. Biochem.* **49**: 421–457.
- Pearson, H., (2006). What is a gene? *Nature* **441**: 399–401.
- Pesole, G., (2008). What is a gene? An updated operational definition. *Gene* **417**: 1–4.
- Poole, A.M., Jeffares, D.C., and Penny, D., (1998). The path from the RNA world. *J. Mol. Evol.* **46**: 1–17.
- Prusiner, S.B., and Scott, M.R., (1997). Genetics of prions. *Annu. Rev. Genet.* **31**: 139–175.
- Richter, J.D., and Treisman, J.E., (2011). Not just the messenger: RNA takes control. *Curr. Opin. Gen. Dev.* **21**: 363–365.
- Sclafani, R.A., and Holzen, T.M., (2007). Cell cycle regulation of DNA replication. *Annu. Rev. Genet.* **41**: 237–280.
- Shapiro, J. A. (2009). Revisiting the central dogma in the 21st century. In: *Natural Genetic Engineering and Natural Genome Editing*. G. Witzany, Ed. *Ann. N. Y. Acad. Sci.* **1178**: 6–28.
- Sharp, P.A., (2009). The centrality of RNA. *Cell* **126**: 577–580.

- Sutton, M.D., and Walker, G.C., (2001). Managing DNA polymerases: coordinating DNA replication, DNA repair, and DNA recombination. *Proc. Natl. Acad. Sci. USA* **98**: 8342–8349.
- Trotman, C.N.A., (1998). Introns-early: slipping lately? *Trends Genet.* **14**: 132–134.
- Tuck, A.C., and Tollervey, D., (2011). RNA in pieces. *Trends Genet.* **27**: 422–432.
- Tuite, M.F., (2000). Sowing the protein seeds of prion propagation. *Science* **289**: 556–557.
- Tyshenko, M.G., and Walker, V.K., (1997). Towards a reconciliation of the introns early or late views: triosephosphate isomerase genes from insects. *Biochim. Biophys. Acta* **1353**: 131–136.
- Verdun, R.E., and Karlseder, J., (2007). Replication and protection of telomeres. *Nature* **447**: 924–931.
- Vitkova, M., Kral, J., Traut, W., Zrzavy, J., and Marec, F., (2005). The evolutionary origin of insect telomeric repeats, (TTAGG)<sub>n</sub>. *Chromosome Res.* **13**: 145–156.
- Watson, J.D., and Crick, F.H.C., (1953). Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Nature* **171**: 737–738.
- Zakian, O.A., (1989). Structure and function of the telomeres. *Annu. Rev. Genet.* **23**: 579–604.

# Transcription, Translation, and Regulation of Eukaryotic DNA

## Chapter Outline

- 2.1 Overview 37
- 2.2 Introduction 38
- 2.3 RNA Synthesis is Gene Transcription 41
- 2.4 Transcription Involves Binding, Initiation, Elongation, and Termination 43
- 2.5 RNA Transcripts of Protein-Coding Genes 45
- 2.6 RNA of Protein-Coding Genes Must Be Modified and Processed in Eukaryotes 45
- 2.7 Splicing Out the Introns 47
- 2.8 Translation Involves Protein Synthesis 48
- 2.9 RNA Surveillance: Damage Control 52
- 2.10 Import and Export from the Nucleus 53
- 2.11 Transport of Proteins within the Cytoplasm 54
- 2.12 mRNA Stability 55
- 2.13 Chaperones and the Proteosome 56
- 2.14 RNA Silencing or Interference and miRNAs 56
- 2.15 Gene Regulation in Eukaryotes 58
- 2.16 Insulators and Boundaries 63
- 2.17 Chromosome or Gene Imprinting in Insects 64
- 2.18 Eukaryotic Genomes and Evolution 66
- References Cited 67

## 2.1 Overview

Genetic information in an organism is expressed in three classes of genes: 1) structural (protein-coding) genes that are transcribed into messenger RNA (mRNA) and translated into polypeptides; 2) genes that code for ribosomal RNA (rRNA), transfer RNA (tRNA), or other classes of small RNAs, in which the transcription product is used directly; and 3) regulatory sites that are not transcribed but that serve as recognition sequences for proteins involved in DNA replication, transcription, and repair.

Protein-coding genes in eukaryotic organisms are transcribed from DNA into pre-messenger RNAs (pre-mRNAs) that then are processed into mRNA.

Processing of mRNA involves splicing to remove introns; mRNA also is capped and methylated at the 5' end, and most mRNAs are polyadenylated at the 3' end. The information in the mRNA then is translated into proteins via protein synthesis in ribosomes.

A ribosome begins protein synthesis once the 5' end of an mRNA is inserted into it. A lengthening polypeptide chain is produced and, once the 5' end of the mRNA emerges from a ribosome, the mRNA can attach to a second ribosome and a second identical polypeptide can be synthesized. The assembly of amino acids into a peptide starts at the amino end and terminates at the carboxyl end. Amino acids are carried to the ribosome by transfer RNAs (tRNAs). The tRNAs are held so their anticodons form base pairs with complementary codons of the mRNA. One tRNA may recognize more than one codon because the genetic code is degenerate. According to the "wobble hypothesis," the initial two bases of the mRNA codon pair according to base-pairing rules, but the third base may pair with any one of several bases.

Protein synthesis occurs in the ribosomes that are located in the endoplasmic reticulum of the cytoplasm. Once proteins are produced, they are transported into the Golgi apparatus where they are processed and transported to their ultimate destination. The protein must be folded and, sometimes, assembled into multiprotein complexes. Folding may require the assistance of other proteins called molecular chaperones.

Gene regulation in insects and other eukaryotes is complex, diverse, and the subject of intensive research. Much of the DNA once labeled as "junk" or noncoding we now know is transcribed into RNAs that are involved in gene regulation and development. Protein-coding genes may be amplified or rearranged to increase the amount of gene product. Protein-coding genes may be regulated by being methylated. Activator proteins, hormones, enhancers and noncoding RNAs (ncRNAs) affect transcription. Alternative splicing, alternative promoters, and translational control are used in gene regulation, as are ncRNAs. Insulators or boundary elements are naturally occurring DNA sequences that protect genes from position effects, establishing independent functional domains within the chromosome.

## 2.2 Introduction

The Central Dogma, that DNA is transcribed into RNA that subsequently is translated into proteins, describes the process by which information contained in the protein-coding DNA is made available to the cell and organism (Figure 1.1). Proteins have many functions in the cell. Structural proteins form part of the framework of the organism, such as the sclerotin in the exoskeleton of insects. Contractile proteins enable organisms to move. Catalytic proteins, or enzymes, regulate the diverse

biochemical reactions taking place within the cell. Transport proteins carry important molecules throughout the body. Regulatory proteins control and coordinate biochemical reactions in the cell and the organism as a whole. Protective proteins (antibodies) protect against infectious agents and injury. Storage proteins store products for future use. However, the dogma that “one gene equals one protein” needs modification because not all genes code for proteins.

Much of the genome (up to 80%) contains DNA that does not code for proteins, and much of this DNA has been considered to be junk or parasitic DNA. Debate has raged as to why the genome would retain this expensive DNA if it had no function. However, geneticists have determined that much of this so-called junk DNA is transcribed into ncRNAs and that these ncRNAs are involved in a variety of essential functions, including gene regulation and development (Storz 2002). The discovery that ncRNAs are important in a variety of cellular functions largely resolves why eukaryotes have so much DNA (Box 2.1).

**Box 2.1 RNAs Have Multiple Functions. Additional RNAs and Functions May Yet be Discovered**

Name	Function	Abbreviation	Size
<b>Protein synthesis</b>			
Pre-messenger RNA	Transcript from the DNA	pre-mRNA	Variable
Messenger RNA	Codes for proteins	mRNA	Variable
Ribosomal RNA	Part of ribosome structure, involved in translation	rRNA	Several
Transfer RNA	Involved in translation, transport of amino acids	tRNA	75–80 nt
<b>Posttranslational modification or DNA replication</b>			
Small nuclear	Splicing of RNA in nucleus	snRNA	60–300 nt
Small nucleolar	Modification of rRNAs	snoRNAs	
Telomerase RNA	Telomere synthesis		
<b>Regulatory RNAs</b>			
Long noncoding	Various	long ncRNA	Various
Micro	Gene regulation, inhibiting translation by binding to mRNAs	miRNA	21–26 nt
Piwi-interacting	Defense against transposons and ?	piRNA	29–30 nt
Small interfering	Gene regulation; one strand of ds siRNA incorporates into RISC, inhibiting transcription of mRNAs with complementary sequence	siRNA	20–25 nt

If we simply know the DNA sequence of all of the protein-coding genes in the insect genome, we do not know how to “make an insect.” The development of a functioning insect involves the coordinated activity of many different proteins, the information for which is encoded in the protein-coding DNA and in the regulatory machinery, much of which is RNA. DNA also carries the code for rRNAs, tRNAs, and many other small RNAs. Ribosomal, tRNA, and small RNA molecules are used directly without being translated into proteins. Furthermore, additional RNAs that are used after transcription are having their functions elucidated (He and Hannon 2004, Atkins et al. 2011). Research to decipher how genetic information in the protein-coding genes and the various RNAs is used by the cells and organism during development is rapidly advancing (Herbert and Rich 1999, Lee and Young 2000, Taft et al. 2009, Belles 2010, Belles et al. 2010). Thus, a reductionist approach to understanding genetics (focusing primarily on DNA sequences) is insufficient to understand the structure and function of insects and other organisms. An insect brain cell and an ovary cell contain the same DNA, but each clearly has different structure and function. How cells differentiate into different tissues and organs during the life stages of an insect (egg, larva, pupa, adult) is a complex process. Thus, geneticists and systems biologists need to work together to resolve this complexity (Alberts et al. 2010, Nurse and Hayles 2011).

Noncoding RNAs (ncRNAs) include microRNAs (miRNAs) and small interfering RNAs (siRNAs), small RNAs (small nucleolar RNAs [snoRNAs], small modulatory RNAs [smRNAs], and Piwi-interacting RNAs [piRNAs]), and medium and large RNAs (Box 2.1). These ncRNAs are designated by their length and function and are called noncoding because they do not code for proteins. They clearly code for important functions, however. miRNAs and siRNAs are composed of 20–25 nucleotides (nt) in length and are involved in gene silencing and RNA interference (RNAi). Small RNAs include snoRNAs, smRNAs, piRNAs, and others, and they range in size from 20 to 300 nt. snoRNAs (60–300 nt) are required for maturation of rRNA (Dieci et al. 2009). Small RNAs are involved in modification of target RNAs, synthesis of telomeric DNA at the ends of chromosomes, modification of chromatin structure, and modulation of transcription. Tiny RNAs (18 nt in length) are associated with promoter regions of genes that are highly expressed (Taft et al. 2009). Medium and large ncRNAs are 300–1000 nt in length and are involved in imprinting DNA, X-chromosome inactivation, DNA demethylation, gene transcription, and generation of miRNAs and small RNAs.

Eukaryotes differ from microorganisms (Eubacteria and Archaeabacteria) in several important ways. First, eukaryotes typically have more than one

chromosome and are genetically more complex (with  $\approx$ 15,000–35,000 genes). Furthermore, most eukaryotic protein-coding genes are split, with one or more noncoding **introns** interspersed among the coding **exons**. The presence of introns requires that transcribed RNA, called pre-mRNA, must have the introns removed before it becomes mRNA and can go from the nucleus into the cytoplasm, where translation can occur.

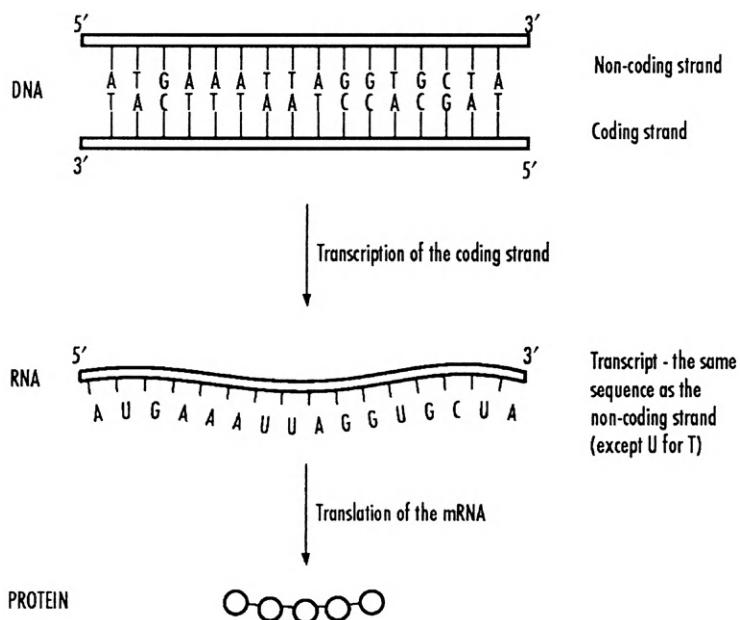
Control elements, such as **promoters** and **enhancers** and a variety of ncRNAs, are important components of gene regulation in eukaryotes. Furthermore, within a particular tissue, a mixture of active and inactive genes is present on each chromosome. Research on *Drosophila* indicates that specific sequences called **boundary elements** insulate the active from the inactive regions of chromosomes (Bell and Felsenfeld 1999, Bell et al. 2001). The location of specific chromosomal regions in the nucleus affects the ease with which specific genes are transcribed.

Finally, the nuclear membrane in eukaryotes separates the processes of transcription in the nucleus and translation in the cytoplasm in both time and space. The nuclear membrane controls which molecules enter and leave the nucleus. The nuclear membrane contains pores with a complex structure consisting of  $\approx$ 400 polypeptide chains (nucleoporins) forming a cylindrical structure. The nuclear pore allows the diffusion of ions, small molecules, and molecules  $<$ 40 kilodalton (kDa) (Kriwacki and Yoon 2011). Larger molecules containing markers indicating they are destined for the nucleus are actively transported through the membrane through mechanisms that remain under study. The intricacies of transcription and translation of eukaryotic DNA are still being unraveled, but molecules involved in transcription and translation must pass through this membrane.

## 2.3 RNA Synthesis is Gene Transcription

**Transcription** is the first stage of protein-coding gene expression (Figure 2.1). During transcription, the coding strand of DNA serves as a template for synthesis of a complementary RNA molecule. The sequence of the RNA molecule is determined by complementary-base pairing so that the RNA is a complementary transcript (copy) of the coding strand of DNA.

Transcription requires four ribonucleoside 5'-triphosphates: ATP, GTP, CTP, and UTP (uracil substitutes for thymine in RNA). A sugar–phosphate bond is formed between the 3'-OH group of one nucleotide and the 5'-triphosphate of a second nucleotide by the enzyme RNA polymerase. RNA polymerase can initiate RNA synthesis without requiring a primer. The sequence of bases in the RNA

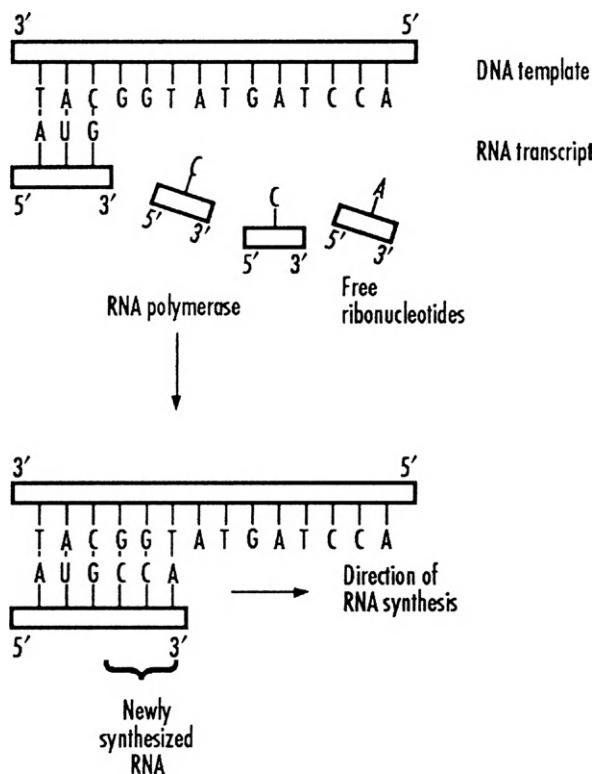


**Figure 2.1** Expression of protein-coding genes involves transcription of the coding strand of DNA to pre-mRNA by RNA polymerase and then processing the pre-mRNA to mRNA in the nucleus before it is transported to the cytoplasm where the mRNA is translated into proteins on the ribosomes.

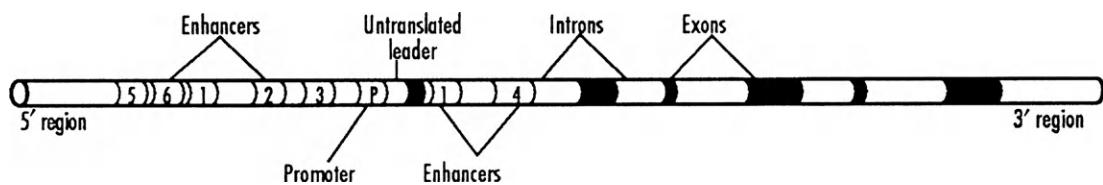
molecule is determined by the sequence of bases in the DNA coding strand. Base pairing occurs between the DNA bases and the newly forming single-stranded RNA molecule. Nucleotides are added to the 3'-OH end of the growing end of the RNA molecule; thus, synthesis of RNA proceeds in the 5' to 3' direction, as does DNA synthesis (Figure 2.2).

Eukaryotes have three types of nuclear RNA polymerases, called RNA polymerase I, II, or III, and each is responsible for transcribing the three different classes of genes (Carter and Drouin 2009). RNA polymerase I primarily is responsible for synthesis of class I genes; this class includes the large rRNAs in the nucleolus. Class II genes include all the DNA sequences that code for proteins and some small nuclear RNAs (miRNAs, small nuclear RNAs [snRNAs], snoRNAs). RNA polymerase III synthesizes tRNAs, 5S ribosomal RNA, a small ribosomal RNA and snRNAs not made by RNA polymerase II.

Figure 2.3 illustrates several of the elements of a typical eukaryotic class II gene. This gene consists of noncoding introns, amino-acid-coding exons, one or more promoters, and several (in this example, six) enhancer elements.



**Figure 2.2** RNA synthesis involves polymerization of free ribonucleotides by an RNA polymerase in the 5' to 3' direction. Thus, the DNA template is read in the 3' to 5' direction but does not require a primer (as does DNA polymerase).



**Figure 2.3** Components of a typical class II eukaryotic gene that codes for proteins include non-coding introns that are spliced out of the pre-messenger RNA, coding exons (shown in black), a promoter to which RNA polymerase II attaches to initiate RNA synthesis, and several enhancers (here numbered 1–6) that influence the level of gene transcription.

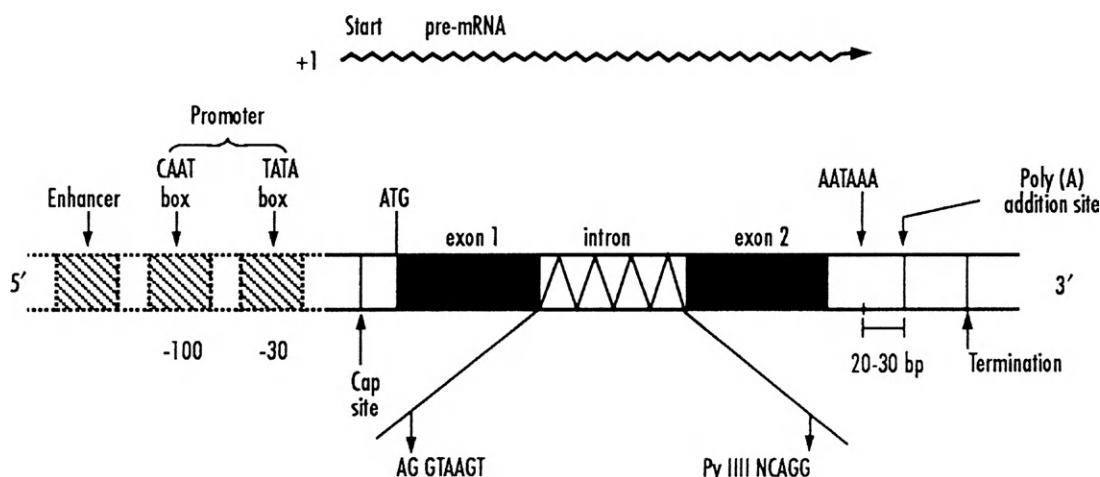
## 2.4 Transcription Involves Binding, Initiation, Elongation, and Termination

RNA polymerase II synthesizes all mRNA in eukaryotes and is a large and complex protein (Cramer et al. 2000, Lee and Young 2000). RNA polymerase II can unwind

DNA, polymerize RNA, and proofread the developing transcript. RNA polymerase II recognizes promoters and responds to regulatory signals (Cramer et al. 2000).

**Transcription** is a complex, multistep process involving binding of RNA polymerase to the DNA, initiation, elongation, and production of a pre-mRNA transcript (Hoffman et al. 1997, John and Workman 1998). To initiate transcription of a class II protein-coding gene, binding of RNA polymerase II must occur at a specific point upstream of the DNA to be transcribed. This process usually involves loading a protein called TFIID onto a promoter, followed by the recruitment of a group of proteins called general transcription factors (GTFs). These promoter-bound GTFs then recruit RNA polymerase II to form a preinitiation complex.

The specific attachment sites of the preinitiation complex are called **promoters**, and they are typically 20–200 nt long and are ≈50 base pairs (bp) upstream of the start site. Different eukaryotic promoter sequences are known but certain common, or consensus, patterns occur. For example, TATAAT and CAAT are promoter sequences for protein-coding genes (Figure 2.4), and they often are called the TATA and CAAT boxes. The location of the TATA sequence may vary. Housekeeping genes (i.e., genes that are expressed in all cells to maintain



**Figure 2.4** A more detailed view of a eukaryotic class II gene that codes for proteins. Promoters often have CAAT and TATA boxes upstream from the start site. The left junction (splice donor, AG GTAACT) and right junction (splice acceptor, NCAGG) sequences of the intron are shown. Splice sites are indicated by the arrow between AG and GT of the splice donor and between the two Gs of the splice acceptor. The pre-mRNA synthesis is initiated at the +1 (start) site and proceeds in the 3' direction.

fundamental activities) may lack the TATA box and have a GC-rich region  $\approx$ 33 nt upstream from the start site. The actual sequences of the promoter vary from gene to gene; the “strength” of the promoter affects the extent to which each gene is expressed.

Eukaryotic promoters often have DNA sequences called **enhancers** that influence the efficiency with which RNA polymerase II and accessory factors can assemble at a promoter to initiate transcription of the DNA (Figures 2.3, 2.4). Enhancers can be at a great distance relative to the RNA start site and can be upstream or downstream of the gene (Blackwood and Kadonaga 1998). For example, the enhancer of the *Drosophila* *cut* gene is 85 kilobases (kb) upstream from the promoter. Enhancer sequences vary in length from 50 to 1.5 kb. Enhancers activate their target gene in a specific cell type at a particular stage in development. Once RNA polymerase recognizes the specific attachment site, the next two phases in transcription can occur: initiation and elongation (John and Workman 1998). Initiation involves “melting” the DNA around the start site of transcription. Structural changes and movement of the DNA accompany the transition from an initial complex to a transcribing complex (Cramer et al. 2000).

## 2.5 RNA Transcripts of Protein-Coding Genes

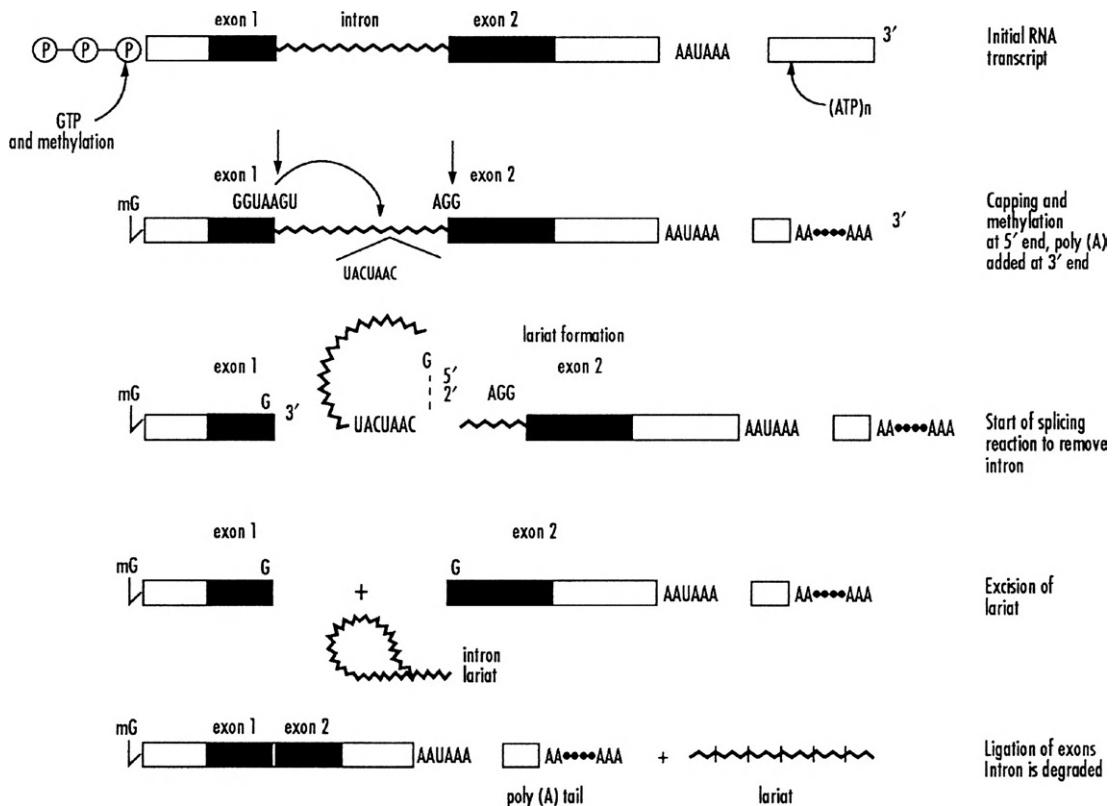
The actual RNA transcript (pre-mRNA) produced in eukaryotes is longer than the gene it is transcribing because RNA polymerase transcribes a **leader sequence**, the length of which varies from gene to gene. When the end of the gene has been reached, RNA polymerase continues to transcribe a **trailer segment** before terminating its activities.

Termination of class II (protein-coding) genes seems to occur hundreds or even thousands of nucleotides downstream of the 3'-end of the mRNA that, in turn, generally lies  $\approx$ 35 nt downstream from the site coding for a polyadenylation [poly(A)] signal, AAUAAA (Figure 2.4).

The number of noncoding introns in class II genes varies, as does their length. The boundaries between introns and exons often are determined by a consensus sequence to ensure that the introns are spliced out of the transcript in a precise manner (Figure 2.5).

## 2.6 RNA of Protein-Coding Genes Must Be Modified and Processed in Eukaryotes

In eukaryotes, RNA transcribed from DNA (pre-mRNA) must be modified and processed before it can function as mRNA (Figure 2.5). Processing the pre-mRNA



**Figure 2.5** The pre-mRNA must be processed before it becomes mRNA. The initial transcript is capped at the 5' end by adding a G and the G is then methylated. The intron is removed after the left end is cut and a lariat is formed between the G at the 5' end of the intron and the 3' A nucleotide with an unusual 5'-2' phosphodiester bond. The right end of the intron is cut, the lariat is released, and the exons are then joined. The released lariat is degraded later. The 3' end is polyadenylated after the signal AAUAAA; the length of the poly(A) tail affects the length of time the mRNA can be translated into a polypeptide.

involves two activities: 1) modifying both ends of the RNA molecule and 2) excising the noncoding sequences (introns) contained within the coding region.

Modifying the ends involves **capping** the 5' end and adding a polyA tail to the 3' end. Newly synthesized eukaryotic RNA molecules (pre-mRNAs) are capped at the 5' end by adding a terminal guanine (G) that has been methylated on the 7-position and linked to the start site by an unusual 5'-5' triphosphate linkage (Figure 2.5). Capping seems to be necessary to enable the ribosome to bind with the mRNA before protein synthesis can begin. The methylated G nucleotide is added in a two-step process, with methylation occurring after a standard G has been added. In some eukaryotes, additional methyl groups may be added to one or both of the next two nucleotides of the mRNA molecule.

The 3' end of eukaryotic RNA is modified by adding 40–200 adenine (A) residues to a region near the 3' end of the transcript to produce the **poly(A)** tail (Figure 2.5). The polyadenylation does not simply add the A residues to the end of the transcript. First, a cleavage occurs between 10 and 30 nt downstream of a specific polyadenylation signal; in insects, this signal is usually AAUAAA and is found in the 3' noncoding region of the RNA. This cleavage results in an intermediate 3' end to which the poly(A) tail is added by the enzyme poly(A) polymerase. The length of the poly(A) tail may determine how long the mRNA survives in the cytoplasm before being degraded.

## 2.7 Splicing Out the Introns

The third modification of the pre-mRNA involves splicing to remove any introns. Splicing takes place in two steps. Introns have a 5' donor and a 3' acceptor end with common consensus sequences (Figure 2.5). The 5' donor end typically has the sequence GGUAAGU. After a cut in the donor site, the G at the 5' end forms a loop by attaching to an A nucleotide a short distance upstream from the pyrimidines near the acceptor splice site. The consensus sequence of the 3' acceptor site is AGG. In the final step, a cut is made in the acceptor site, and the intron is freed. The exons are then joined together. The excised loop (intron) is released as a lariat-shaped structure and is later degraded (Figure 2.5).

RNA splicing occurs in large multicomponent complexes called **spliceosomes**. Spliceosomes are composed of >50 proteins and five types of snRNA molecules (U1, U2, U4, U5, and U6) (Reed 2000). The active catalytic components of the spliceosome are the snRNAs rather than the proteins, although the proteins are required and participate in both assembly of the spliceosome and the splicing reaction. In addition, several proteins play auxiliary roles in splicing and spliceosome assembly.

Because many pre-mRNAs contain multiple introns, the splicing machinery must be able to identify and join the appropriate 5' end and 3' splice sites to produce a functional mRNA. The specificity of the splicing operation is determined by the snRNAs that contain sequences that are complementary to the splice junctions (Cooper 2000, Hastings and Krainer 2001). Introns may have no function, and synthetic genes lacking introns can function quite well. However, some introns are important in gene regulation and determine when, or in what tissue, the gene will be transcribed. For example, sex determination in *Drosophila melanogaster* depends on a cascade of splicing and the pre-mRNA of the *double-sex<sup>+</sup>* gene of the female contains exon 4, whereas males lack it. Likewise, the splicing of *Sex lethal<sup>+</sup>* and *transformer<sup>+</sup>* varies by sex (see Chapter 10).

Mutations in introns can be neutral or can alter gene regulation. Mutations in the splicing signals may result in two classes of mutations. If an intron is not spliced out, a mutant protein can be produced that functions abnormally. An abnormal mRNA is produced if splicing occurs at a different site than normal, and a mutant protein will be produced as a result. More than half the introns in *Drosophila* and other invertebrates are <80 nt long (Guo and Mount 1995). Once mRNA is produced, it must be transported through the nuclear envelope to the cytoplasm where it is translated.

## 2.8 Translation Involves Protein Synthesis

Translation is the second stage of class II gene expression in which the information in the mRNA is used to direct the synthesis of a polypeptide, the amino-acid sequence of which is determined by the nucleotide sequence of the RNA.

The genetic code consists of a triplet of adjacent ribonucleotides that specify an amino acid (see Table 1.1). Translation requires ribosomes; tRNAs, a set of enzymes (aminoacyl tRNA synthetases) to catalyze the attachment of each amino acid to its corresponding tRNA molecule; and initiation, elongation, and termination factors. Translation occurs in ribosomes located in the cytoplasm. **Ribosomes** are cellular organelles, consisting of two subunits (a larger 60S subunit and a smaller 40S subunit), each composed of ribosomal RNA and proteins. (The S refers to the rate of sedimentation and is an indication of size.) The larger 60S unit consists of three rRNA molecules (25S, 5.8S, and 5S) and 46 proteins, whereas the smaller 40S subunit includes one rRNA (18S) and 33 proteins (Ben-Shem et al. 2011). The smaller subunit binds mRNA and the anticodon end of tRNAs and helps to decode the mRNA. The larger subunit interacts with the amino-acid–carrying end of tRNAs, catalyzes the formation of the peptide bonds, and contains the polypeptide exit tunnel (Ben-Shem et al. 2011, Klinge et al. 2011).

The mechanisms whereby ribosomes engage with the mRNA and select the start site for translation are more complicated in eukaryotes than in prokaryotes (Dever 1999, Kozak 1999, Preiss and Hentze 1999). However, the fundamental components of translation are conserved: rRNAs are strongly conserved in both primary and secondary structure among all organisms. The majority of ribosomal proteins are conserved, as well as the elongation factors, the tRNAs, and the aminoacyl-tRNA synthetases (Kyprides and Woese 1998).

Initiation sites in eukaryotic mRNAs are reached by a scanning mechanism that predicts translation should start at the AUG codon nearest the 5' end of the mRNA. The selection of the start codon sets the reading frame that is

maintained throughout subsequent steps in the translation process. Protein synthesis is often regulated at the level of initiation, making it an important step.

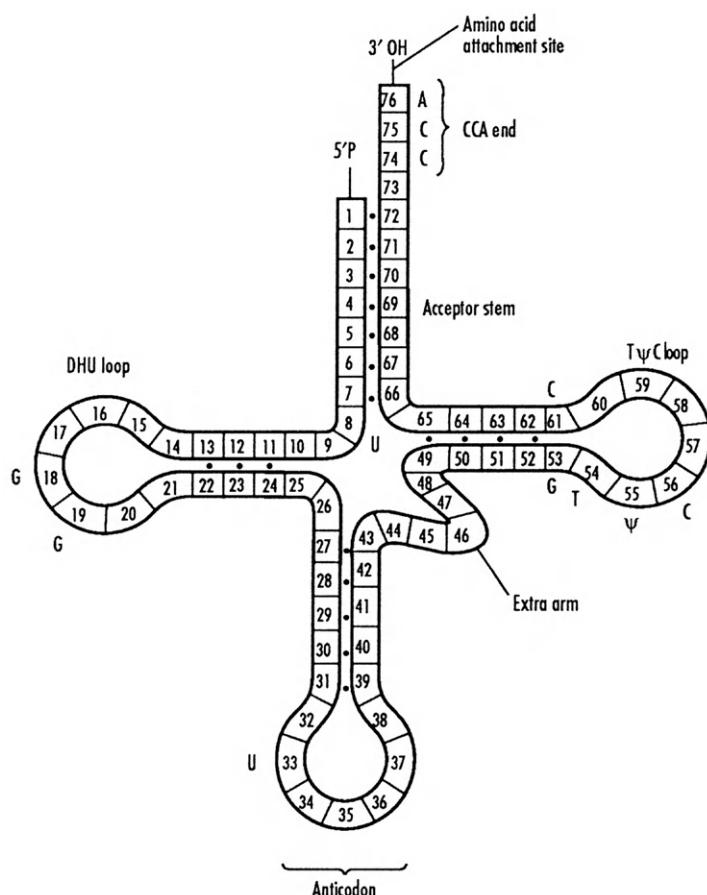
Because the RNA of the larger ribosomal subunit carries out the formation of the peptide bonds, the ribosome is a **ribozyme**, meaning that the RNA acts as an enzyme. The RNA carries out the key peptidyl transferase reaction (Cech 2000, Nissen et al. 2000). By contrast, the proteins in the ribosome are structural units and help to organize key catalytic RNA elements.

Translation of the genetic information in eukaryotes begins when an mRNA molecule binds to the surface of a ribosome and the initiation codon (AUG) is selected. The Met-tRNA interacts with the AUG start codon of the mRNA. More than nine eukaryotic initiation factors have been identified and several are composed of multiple polypeptide chains. The large number of polypeptides indicates that protein–protein interactions play an important role in initiation of translation (Dever 1999).

tRNAs carry an amino acid to the ribosome where they bind to the mRNA molecule attached to the ribosome. tRNAs have a three-base sequence, the **anticodon**, that is complementary to a specific codon in the mRNA. At the other end of the tRNA is a site to which a specific amino acid is bound (Figure 2.6). Binding between the mRNA codon and tRNA anticodon occurs by hydrogen bonds. Proteins within the ribosome function as cofactors, buttressing, stabilizing, and orienting the floppy ribosomal RNA into a specific, active ribozyme. **Peptide bonds** are made between the successively aligned amino acids until the stop codon at the end of the mRNA is reached (UAA, UAG, or UGA) and the completed protein is released. The polypeptide is thus synthesized from the amino end toward the carboxyl end.

tRNAs are small, single-stranded molecules ranging from 70 to 90nt (Figure 2.6). Internal complementary-base sequences allow the molecule to form short double-stranded regions, thereby yielding a folded molecule (sometimes called a cloverleaf structure) in which open loops are connected to each other by double-stranded stems. The three-dimensional structure of tRNA molecules is important. One significant region is the anticodon region; this region consists of three bases that can base-pair with the codon in the mRNA. A second critical site is the 3'OH end of the molecule where the amino acid attaches.

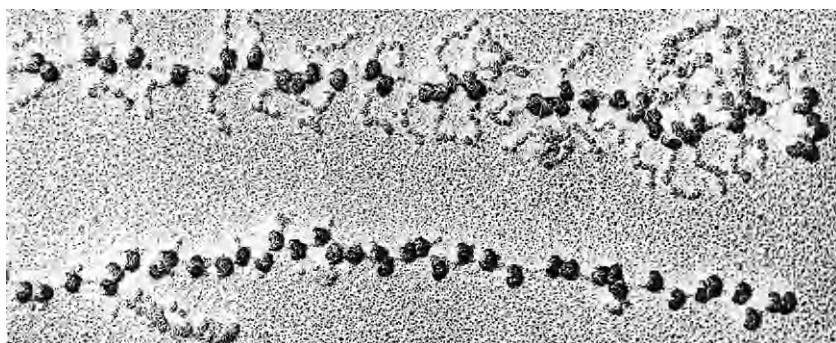
A specific enzyme called **aminoacyl tRNA synthetase** matches each amino acid with the tRNA attachment site. tRNA molecules and their synthetases are designated by giving the name of the amino acid that is specific to each particular tRNA molecule. Thus, leucyl-tRNA synthetase attaches leucine to tRNA<sup>Leu</sup>. If an amino acid is attached to a tRNA molecule, it is “charged.” Usually, one, and only



**Figure 2.6** A tRNA molecule has a complex shape. Bases are numbered from 1 to 76. A few bases that are present in almost all tRNA molecules are identified by letters. The Greek letter  $\psi$  is a symbol for the unusual base pseudouridine. The amino-acid attachment site is at the 3' end of the tRNA and the anticodon (the genetic code that determines the amino-acid order) is at the bottom of the diagram. The anticodon will base-pair in a complementary manner with the mRNA codon.

one, aminoacyl synthetase is found for each amino acid. However, there are fewer aminoacyl synthetases than there are codons for amino acids. Thus, aminoacyl synthetases must recognize more than one codon. The **wobble hypothesis** suggests that base pairing is most critical with the first two bases of the three-base codon but that pairing is extremely flexible in the third position.

Gene translation in eukaryotes usually involves multiple ribosomes processing along a single mRNA molecule. Thus, after  $\approx 25$  amino acids have been joined together into a polypeptide, the AUG initiation codon is free of the first



**Figure 2.7** Micrograph of polysomes from salivary gland cells of a midge larva (*Chironomus*). Here, the ribosomes are moving in order along the mRNA molecule, gradually extending their individual protein chains. More than one ribosome can be attached to the mRNA so that multiple polypeptides are produced. The start of the polysome is at the bottom and the end is at the top.

ribosome and another ribosome can attach and a second polypeptide can begin to form. When the second ribosome has joined  $\approx 25$  amino acids, a third ribosome can attach to the initiation site. This process can result in mRNA molecules with many ribosomes, all of which are moving in the same 5' to 3' direction. Such a large unit is called a **polysome**. Figure 2.7 illustrates a polysome isolated from a midge larva (*Chironomus*). This electron micrograph, magnified  $\approx 140,000$  times, shows the start of an mRNA molecule on the bottom right (Kiseleva 1989). The structure at the top shows the end of the molecule, with the growing proteins shown attached to the ribosomes.

Proteins are linear chains of amino acids that adopt unique three-dimensional structures that allow them to carry out their biological functions. A central tenet until recently has been that the function of a protein depends entirely on its fixed three-dimensional shape (Chouard 2011). However, new studies indicate that, although many proteins immediately and spontaneously fold into appropriate three-dimensional shapes, some protein segments can only function when able to change shape, and this unstable region may be essential to its function. However, truly malformed or defective proteins can be dangerous to organisms. It is thought that all the information needed to specify a protein's three-dimensional structure is contained within its amino-acid sequence (Denton and Marshall 2001), but the process of prediction remains complex and the subject of intense research (Chouard 2011). Protein folds are the basic units of proteins, each consisting of between 80 and 200 amino acids. Some proteins consist of a single fold, but most are a combination of two or more folds. Given suitable conditions, most small proteins fold spontaneously into their three-dimensional

**Box 2.2 Key Points Regarding Transcription and Translation of DNA****Transcription**

Protein-coding genes are transcribed into pre-mRNA that must be processed in the nucleus into mRNA before going to the cytoplasm for translation.

Transcription requires RNA polymerase, but no primer; transcription of the coding strand of DNA occurs from the 5' to the 3' end. Ribonucleotides include A, C, G, and U.

Processing of mRNA involves splicing out the introns in a spliceosome, capping and methylation of the 5' end, and adding a poly(A) tail at the 3' end.

mRNA for protein-coding genes also includes a leader sequence and a trailer segment.

Genes for rRNA, tRNA, and other small RNAs are used directly without translation into proteins.

**Translation**

**Protein-coding genes are translated in the cytoplasm on the ribosomes**, which consist of rRNAs and proteins. The rRNAs are ribozymes.

**tRNAs have an anticodon complementary to the codon of the mRNA.**

**tRNAs have a three-dimensional structure and contain an amino-acid attachment site** where the amino acids are attached by specific enzymes.

**tRNAs carry a specific amino acid to the ribosome** where they bind to the mRNA molecule based on complementary base pairing.

**mRNA is monitored and destroyed if it is damaged.**

The **stability of mRNA is regulated** so that specific proteins are produced in appropriate amounts at the appropriate location to ensure normal development and metabolic activities are accurately maintained.

form (Baker 2000). The ability to predict protein structure and folding mechanisms is very difficult, but this information helps in understanding how the protein functions (Lee et al. 2007, Chouard 2011).

See Box 2.2 for key points regarding transcription and translation of DNA.

## 2.9 RNA Surveillance: Damage Control

It is very important that mRNA be accurately produced; otherwise, damaged or truncated proteins are produced (Hilleren and Parker 1999, Maquat and Carmichael 2001, Doma and Parker 2007). **mRNA surveillance** systems in eukaryotes monitor pre-mRNA processing and RNA translation (Doma and Parker 2007). mRNAs that are translated aberrantly, or that have no stop codons, or that produce a stall in elongation, or that translate beyond the normal stop codon into the 3' untranslated region are detected by specific machinery that eliminates defective mRNAs (Doma and Parker 2007). mRNA surveillance

increases the fidelity of gene expression by degrading aberrant mRNAs. RNA surveillance systems also assess whether the transcripts possess the proper regulatory elements. Thus, mRNA surveillance is important in determining which mRNA molecules become available for translation and, hence, which proteins are produced within the cell. tRNAs also are monitored and destroyed if defective, as are defective rRNAs (Doma and Parker 2007).

## 2.10 Import and Export from the Nucleus

The presence of a nucleus distinguishes eukaryotes from prokaryotes. The nucleus contains the nuclear genetic information and is the cell's control center where DNA replication, transcription, and RNA processing take place. The final stage of gene expression, translation, occurs in the cytoplasm.

The nucleus is surrounded by a nuclear envelope consisting of two nuclear membranes constructed of phospholipid bilayers, an underlying lamina (a network of lamin filaments that provide structural support), and nuclear pore complexes. The nuclear lamina serves as a site to which the chromosomes attach. When cells are not undergoing mitosis or meiosis, the chromosomes are "unwound" and resemble a "plate of spaghetti." The location of these chromosomes is not random; chromosomal location is important in determining when and how well genes function (Pennisi 2011). The nucleolus, which consists of proteins and RNAs where the cell's ribosomes are made before export to the cytoplasm, has a designated location, as do other nuclear structures (Baker 2011). It seems that gene location varies within the nucleus depending upon the developmental stage or tissue type. Active genes are usually near the edges of chromosome territories, whereas silenced genes are deep inside these territories. Gene location within the nucleus plays a role in fully activating a gene and making transcription and RNA processing efficient (Pennisi 2011).

The inner and outer membranes of the nuclear membrane are joined at the nuclear pore complexes, the only channels through which small polar molecules and macromolecules are able to travel through the nuclear envelope. The nuclear membranes prevent the free movement of molecules between the nucleus and the cytoplasm, thus maintaining the nucleus as a distinct biochemical compartment. The outer nuclear membrane is similar in structure to the membranes of the endoplasmic reticulum, to which it attaches.

The **nuclear pore complex** has a diameter of  $\approx 120\text{ nm}$ . The pore functions as a semipermeable filter that allows diffusion of ions, small molecules, and larger molecules  $<60\text{ kDa}$  (Grunwald et al. 2011). Larger molecules that have a nuclear localization signal enter the nucleus by transport factors, but the method by

which this occurs is under investigation (Grunwald et al. 2011, Kriwacki and Yoon 2011). Approximately 400 polypeptide chains (nucleoporins) form the cylindrical nuclear pore (Kriwacki and Yoon 2011).

mRNA is transported from the nucleus to the cytoplasm through the nuclear pore after synthesis and processing within the nucleus. Proteins required for nuclear functions (such as transcription factors, histones, DNA polymerases, RNA polymerases, and splicing factors) are manufactured in the cytoplasm and transported into the nucleus. Transport of molecules through the pore can be passive or an active, energy-dependent process. The passive transfer of molecules through the nuclear pore complex involves only small molecules and proteins <50 kDa.

Most proteins and RNAs are transported actively into and out of the nucleus. During active transport, the nuclear pore can open to a diameter of >25 nm. Some proteins are recognized and transported because they carry specific signals (exportins) that bind to a receptor on the transport machinery (Nakielny and Dreyfuss 1999).

Pre-mRNAs and mRNAs are associated with a set of at least 20 proteins, forming a **heterogeneous nuclear ribonucleoprotein complex** (hnRNP) throughout their processing in the nucleus and transport to the cytoplasm. At least two of the hnRNP proteins contain nuclear export signals.

snRNAs function within the nucleus as components of the RNA processing machinery. These RNAs are initially transported from the nucleus to the cytoplasm where they associate with proteins to form functional small nuclear ribonucleoproteins, and then they return to the nucleus. Proteins that bind to the 5' caps of snRNAs seem to be involved in the export of the snRNAs to the cytoplasm, whereas other sequences are responsible for their transport from the cytoplasm into the nucleus (Cooper 2000).

## 2.11 Transport of Proteins within the Cytoplasm

Proteins are transferred into the endoplasmic reticulum when they are translated on membrane-bound ribosomes. The **endoplasmic reticulum** (ER) is a network of membrane-enclosed tubules and sacs (cisternae) that extends from the nuclear membrane throughout the cytoplasm. The entire ER is enclosed by a continuous membrane and is the largest organelle of most eukaryotic cells. Proteins destined to remain in the **cytosol** (the fluid portion of the cytoplasm, excluding organelles) or to be incorporated into the nucleus or mitochondria are synthesized on free ribosomes and released into the cytosol when their translation is complete.

Two types of ER, rough and smooth, perform different cellular functions. Rough ER is covered by ribosomes on its outer surface and functions in protein processing (Stutz and Rosbash 1998). Smooth ER functions in lipid metabolism (Cooper 2000). The details of ER transport are an active area of research.

Proteins move across the ER membrane through a hydrophilic channel that is evolutionarily ancient (Matlack et al. 1998). The ER membrane channel translocates proteins and also integrates membrane proteins into the lipid bilayer. Thus, the ER membrane channel must identify the signal sequences of the proteins, open in response to the signal, transport the protein from one side of the membrane to the other, and then close. For membrane-spanning proteins, some parts must be moved across the membrane, whereas others must be left in the cytoplasm. The ER has several mechanisms to ensure that only properly folded proteins enter the secretory pathway (Reddy and Corley 1998).

Proteins received from the ER are further processed and sorted for transport to their eventual destinations in the **Golgi complex** (Rothman 1994, Cooper 2000). The Golgi complex is a cytoplasmic organelle specialized for processing and sorting proteins and lipids before their transport to lysosomes, the plasma membrane, or secretion. The Golgi is made of flattened membrane-enclosed sacs (cisternae) and associated vesicles. Proteins from the ER enter one side of the Golgi and exit from the opposite side. As the proteins pass through the Golgi, they are modified; modifications that occur within the Golgi include **glycosylation**. Glycosylation involves adding carbohydrates to proteins. Proteins, lipids, and polysaccharides are transported from the Golgi complex to their final destinations via different kinds of transport vesicles (Featherstone 1998). The Golgi seems to be in constant flux, each cisterna emerging from the ER with its load of proteins and then carrying the proteins across the Golgi, while simultaneously putting the finishing touches on the proteins.

## 2.12 mRNA Stability

Many of the genes involved in development are expressed for a short time and in precisely defined domains of the body. Some genes affect the expression of genes that are downstream in the developmental pathway. Thus, gene expression must be precisely regulated if cell identity is to be specified. Modifying the rate of transcription of DNA is not always sufficient to maintain fine-tuned developmental processes. Sometimes existing mRNAs and proteins must be removed or deactivated, tasks that involve regulating **mRNA stability** (Surdej et al. 1994).

Several mechanisms are involved in regulating mRNA stability, including removal of the poly(A) tail; premature termination of translation due to a

premature termination codon; and, perhaps, mRNA localization. For example, during development of *Drosophila*, cellularization of the blastoderm can result in the destabilization of the *bicoid* mRNA.

## 2.13 Chaperones and the Proteosome

Maintaining the function of proteins within cells depends on more than transcription and translation. The initial folding of proteins and assembly of multiprotein complexes sometimes requires the assistance of molecular **chaperones**—proteins that catalyze protein folding. Chaperones prevent proteins from aggregating into insoluble, nonfunctional blobs and help them reach and maintain a stable functional state. Understanding of protein folding has increased dramatically in the past few years (Radford 2000). There seems not to be a single, specific folding pathway; rather, there are potentially a plethora of routes to the folded protein and which pathways are used will depend on the amino-acid sequence, the topology of the protein, and the environmental conditions of the cell. Different routes might be used or different intermediates and transition states might be observed as a consequence of relatively small alterations of in the cellular environment.

After initial folding and assembly, proteins may be damaged. Such proteins can be rescued (refolded) by chaperones or are destroyed by proteases. The efficiency and cost of protein quality depend on a balance among folding, refolding, and degradation (Wickner et al. 1999).

Controlled degradation of proteins inside the cell is essential. For example, protein degradation is important in cell cycle control, heat-shock response, programmed cell death, muscle atrophy, immune response, metamorphosis, development, and differentiation (Mykles 1999). Proteins targeted for destruction are marked by the attachment of a small protein called **ubiquitin** and, after unfolding, the protein is degraded within a large protein complex known as the **proteosome** (Stuart and Jones 1997). Protein degradation is potentially hazardous to the cell and must be restricted to specific sites and times to prevent the improper destruction of useful proteins. The proteosome provides a compartment to confine the proteolytic action to proteins that carry the degradation signal.

## 2.14 RNA Silencing or Interference and miRNAs

**RNA silencing** (also called **RNA interference** [RNAi]) is a mechanism for defending against the invasion of mobile DNA elements (transposable elements) that can cause mutations when they insert themselves into or close to a gene

(He and Hannon 2004). When double-stranded (ds) RNA is injected into eukaryotic cells by transposable elements (or viruses), the ds RNA seems to function as a signal that the cell is being invaded because most RNA in a cell is single-stranded. Thus, ds RNA elicits a defense response—RNAi (Matzke et al. 2001). RNAi initially was discovered in the nematode *Caenorhabditis elegans*, where it was found to modify larval development (Fire et al. 1998). The discovery of RNAi resulted in a Nobel Prize for Andrew Fire and Craig Mello in 2006. This discovery allows scientists to determine the function of many previously unknown genes through their silencing (Mito et al. 2011). RNAi is being developed as a method for analyzing gene functions in multiple species and orders of insects. RNAi also is being investigated as a tool for modifying the development of pest insects (Mito et al. 2011). RNAi involves the use of miRNAs that degrade specific RNAs (Figure 2.8).

The sequencing of multiple genomes has allowed scientists to confirm that miRNAs are evolutionarily conserved and are found in molluscs, sea urchins, flies, mice, and humans. Such an evolutionary conservatism in gene sequence indicates that these small RNAs are important in regulating the development of many eukaryotic organisms (Cernilgar et al. 2011). miRNAs are 21–25 nt and are derived from longer ds RNAs that form imperfect stem-loop structures. The miRNA is usually formed from one arm of the hairpin and is released after two enzymatic processes involving the RNase III enzymes Drosha and Dicer followed by subsequent processing in a protein complex (RISC) that includes the Argonaute protein. Argonaute cleaves and discards the sense strand of the ds RNA. The antisense strand of the RNA and the RISC protein complex then cleave the homologous messenger RNA, resulting in a *loss of a specific gene function*. The two-step reaction is efficient because each molecule of ds RNA primes several RNase molecules, so the cell can mount a large-scale response to only a few ds RNA molecules (Baulcombe 2001). Cernilgar et al. (2011) noted that RNAi pathways not only modulate gene expression by degrading RNA target molecules in the cytoplasm but also have a function in the nucleus of *D. melanogaster*. They demonstrated that key RNAi components (Dicer and Argonaute) associate with euchromatin, especially transcriptionally active loci, and interact with the core transcription machinery. These RNAi components apparently control the processivity of RNA polymerase II in the nucleus.

After the genome of the silkworm, *Bombyx mori*, was sequenced, Zhang et al. (2009) identified 354 miRNA genes and found that they were important in embryogenesis and metamorphosis. Furthermore, the sequences of these miRNAs were sufficiently different from those found in mammals and the nematode *C. elegans*, to suggest that there are many miRNAs unique to insects. Skalsky et al. (2010) evaluated miRNAs in the mosquitoes *Aedes albopictus* and *Culex*

double-stranded RNA (ds RNA) is introduced into the organism or produced naturally.



Dicer, an RNase III enzyme, cleaves the ds RNA into fragments called small interfering RNAs (siRNAs) that are 21–23 nt in length.



The siRNAs unwind and the antisense strand binds to the RNA-induced silencing complex (RISC), which includes Argonaute proteins. Argonaute cleaves and eliminates the sense strand of the siRNA duplex, leading to an active RISC.



The complex of small interfering RNAs (siRNAs) and RISC couples to the target homologous mRNA.



The mRNA is cleaved and is unable to be translated.

**Figure 2.8** Simplified outline of the steps in RNAi. RNAi is an evolutionarily conserved response to the invasion of mobile elements (transposable elements or viruses). It also allows loss-of-function phenotypes to be studied when no mutants are available. In some organisms, feeding or injection of ds RNA into the body can induce the RNAi response systemically. Some insects seem to lack the ability to respond to RNAi in a systemic manner. *Drosophila melanogaster*, for example, does not show a robust RNAi response.

*quinquefasciatus* and found there were >60 conserved and seven novel miRNAs, some of which are implicated in viral infections in these mosquito vectors.

## 2.15 Gene Regulation in Eukaryotes

At any one time, only  $\approx$ 15% of all the genes in an insect cell are turned on ([Harshman and James 1998](#)). Thus, insect development, behavior, and reproduction are determined by the expression of different genes at different times in different tissues. Research on the control of gene expression in *Drosophila*

### Box 2.3 Key Points Regarding Gene Regulation

#### How Can a Single Genome Code for So Many Cell and Tissue Types?

##### Regulation

**Multiple mechanisms regulate genes.** Some genes are expressed consistently (constitutively) at a specific level in all cells (housekeeping genes), but others must be turned off and on during development or in response to environmental challenges.

Regulatory mechanisms include

- regulating transcription,
- alternative splicing of mRNA transcripts,
- DNA amplification,
- methylation of cytosine bases, and
- translational control.

Transcriptional regulation can be achieved by

- a repressor binding to noncoding DNA on the strand close to or overlapping the promoter,
- the effects of activators enhancing the interaction of RNA polymerase and a promoter, and
- enhancers bound by activators loop DNA to bring a promoter to the initiation complex.

##### Epigenetic Regulation

Although DNA is the molecule that stores genetic information that is transferred to the next generation through the germ line and soma, some inheritance is non-Mendelian (i.e., not related to the DNA sequence in the genome).

Non-Mendelian inheritance can result from the expression of only one of the two alleles within the genome. The DNA sequence itself is not altered, but the expression of the DNA is reversibly modified.

DNA may be reversibly altered by histone modifications, methylation of cytosines, and chromatin remodeling and by targeting of small ncRNAs.

A fertilized egg is thought to develop into an adult organism because the single genome present in the eggs becomes epigenetically programmed to generate distinct epigenomes in the different cell types. Transient modification of the genome during development allows genes to be turned on and off during development.

is helping to resolve the fundamental principles of eukaryotic gene regulation (Harshman and James 1998).

Some eukaryotic genes code for essential metabolic enzymes or cell structural components and are expressed *constitutively* at a specific level in all cells. Such genes are often called **housekeeping genes**, but most genes are not expressed continuously in eukaryotic organisms. After cells differentiate, gene regulation may be influenced by environmental cues such as hormones, nutrients, or temperature. The control of gene expression is called **gene regulation**.

Different genes are regulated differently and gene regulatory mechanisms are often surprisingly complex, using more than one method (Box 2.3).

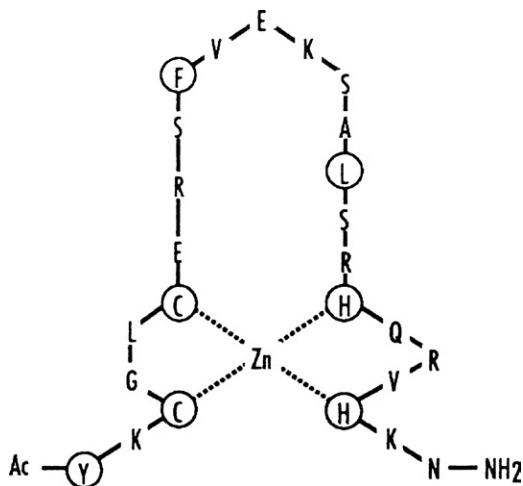
Regulation of transcription is the most common method of gene regulation in eukaryotes. mRNA molecules generally are short-lived, probably persisting a few minutes or hours. Rapid turnover means that by adjusting the rate of transcription the amount of a particular mRNA in the cell can be controlled. Both negative and positive transcriptional regulation can occur.

Transcriptional regulation involves **transcriptional activator proteins** that bind with an upstream DNA sequence to prepare a gene for transcription. They may help assemble a transcriptional complex, or they may initiate transcription by an already-assembled transcriptional complex. Some transcriptional activator proteins have a helix–turn–helix structure, i.e., a sequence of amino acids that form a pair of  $\alpha$ -helices separated by a bend. These helices fit into the grooves of a ds DNA molecule and allow the proteins to bind to the DNA, although the specificity of the binding is determined by other parts of the protein (Harrison 1991). Examples of helix–turn–helix DNA-binding proteins in insects include the **homeo domain**, a domain found in genes such as *engrailed*<sup>+</sup> and *Antennapedia*<sup>+</sup> that are important in regulating development of *D. melanogaster* (described in more detail in Chapter 4).

**Zinc-finger proteins** are a second type of transcriptional activator protein. They are characterized by loops (fingers) of repeating amino-acid sequences each associated with a zinc atom. Zinc-finger proteins bind in the major groove of the DNA helix (Figure 2.9) of an upstream DNA sequence to prepare a gene for transcription.

**Leucine zippers** are a third type of DNA-binding protein. Leucine zippers are DNA-binding proteins that contain four to five leucine residues separated from each other by six amino acids. The leucines on two protein molecules can interdigitate and dimerize in a specific interaction with a DNA recognition sequence (Abel and Maniatis 1989).

Hormones may turn on the transcription of specific sets of genes. For example, steroid hormones penetrate a target cell through diffusion because steroids pass freely through the cell and nuclear membranes. The nuclei of target cells contain specific receptor proteins that form complexes with the hormone; these complexes then undergo modification in their three-dimensional forms, enabling the receptor–hormone complexes to bind with particular sequences in the DNA and stimulate or repress transcription. In the lepidopteran *Manduca sexta*, **ecdysone**, a steroid, acts directly on the genome both to activate and repress genes. Ecdysone initiates and coordinates the molting process and thus the sequential expression of stage-specific genes (Riddiford et al. 1990).



**Figure 2.9** Amino-acid sequence of a *portion* of a protein containing zinc fingers drawn to illustrate the finger motif. Zinc-finger proteins are able to recognize specific DNA sequences. The fingers can intercalate into the DNA molecule, and they are important in gene regulation.

Hormone–receptor complexes and transcriptional activator proteins bind with specific DNA sequences called enhancers. Enhancer sequences can be found in a variety of sites in relation to the target gene (Figure 2.3). Enhancers can be long or short distances upstream (5') from the target gene; they may be included in introns within the coding region of the target gene, or even at the 3' end of the gene. Some enhancers respond to molecules produced inside the cell during development, and many genes are under the control of several enhancers, so they can respond to a variety of internal and external molecular signals.

In the living cell, the DNA of the chromosome is tightly bound up with proteins called histones and other proteins that can make transcription of the genes difficult by preventing the association of proteins with the DNA that are needed for gene transcription (Grunstein 1997, Pennisi 1997, Workman and Kingston 1998, Henikoff and Shilatifard 2011). At least four different enzymes, called nuclear histone acetylating enzymes, are associated with the transcription complex and add simple chemical groups known as acetyls to the histones to open the DNA up to transcription. In addition, there are five more enzymes that can undo the reaction and remove acetyls from the histones, thus making transcription more difficult.

The physical location of a gene within the nucleus during interphase is important in gene regulation (Cockell and Gasser 1999, Baker 2011, Henikoff and Shilatifard 2011). Furthermore, the locations of genes may shift as the cells

change their function (Baker 2011). Correlations have been found between gene silencing and the gene's proximity to a heterochromatic region or to the periphery of the nucleus (Singer and Green 1997). Heterochromatic DNA is usually condensed (making it stain intensely) during interphase, and any genes within heterochromatic regions usually are inactive.

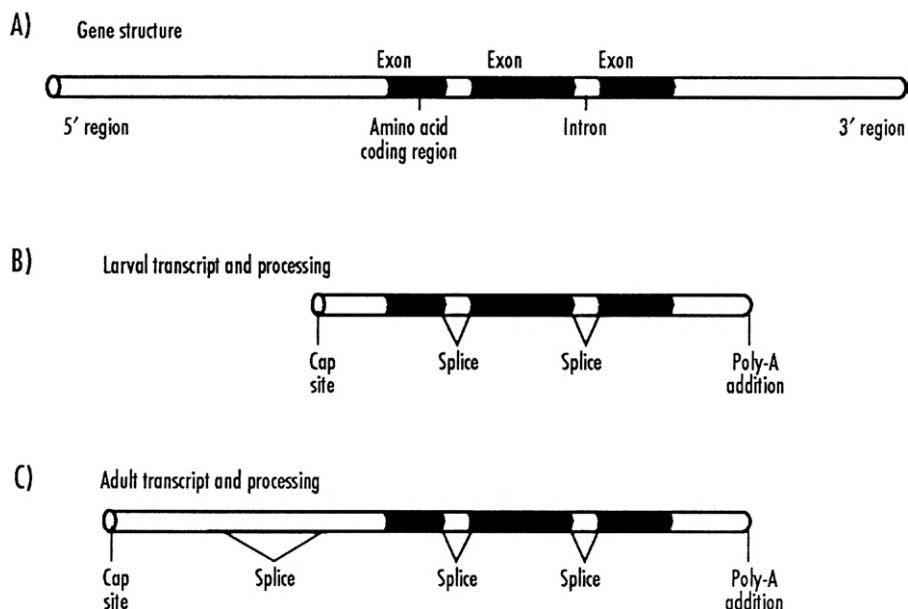
The components in a typical class II eukaryotic gene are illustrated in Figure 2.3. A transcriptional complex binds to the promoter to initiate RNA synthesis. The coding regions of the gene (the exons) are interrupted by introns that are eliminated in RNA processing. Transcription is regulated by enhancer elements (numbered 1–6) that respond to different molecules. Because enhancers may respond to different signals or cell conditions, genes can be regulated by a combination of different enhancers. Combinatorial control of gene transcription makes it possible to increase the complexity of gene regulation. If transcription is determined by which pattern of binding state occurs, then a small number of regulatory molecules can yield a large number of different regulatory patterns for different life stages or tissues.

Some genes have two or more promoters. Each promoter usually is active in different cell types, thereby allowing for independent regulation of transcription (Figure 2.10). The different promoters yield different primary transcripts that code for the same polypeptide. Thus, one promoter could control transcription in insect larvae (Figure 2.10B) and another promoter could control transcription in adults (Figure 2.10C).

Different cell types can produce different quantities of a protein or different proteins by **alternative splicing** of mRNA. Differential splicing of exons can result in different rates of synthesis, although the proteins may be identical. In other cases, the proteins produced are different as a result of alternative splicing. Alternative splicing is involved in many aspects of development, including determining the sex of an insect (see Chapter 10). Alternative splicing also makes it possible for protein diversity to be greater than the number of genes in a genome (Graveley 2001).

**RNA editing** involves altering the sequence of the RNA after it has been transcribed and before it has been translated. RNA editing apparently is widespread in both prokaryotes and eukaryotes, and results in functionally distinct proteins produced from a single gene (Maas and Rich 2000).

Gene regulation also can occur by controlling translation of the message into protein. **Translational regulation** can occur in several ways: 1) inability to translate an mRNA molecule unless a particular signal is present, 2) regulation of the



**Figure 2.10** Gene regulation can be achieved with the use of alternative promoters. A) In this gene, there are two introns within the coding region. B) The mRNA transcript in larvae uses the promoter nearest the 5' end of the coding region. C) In adults, the promoter further upstream is used and much of the leader sequence used in larvae is removed by splicing.

longevity of a particular mRNA molecule, 3) regulation of the rate of protein synthesis, and 4) localization of transcripts where translation is required ([Lipshitz and Smibert 2000](#)).

An example of inability to translate an mRNA unless a particular signal is present is found in unfertilized eggs, which are biologically static. After fertilization, many new proteins are synthesized, including the mitotic apparatus and cell membranes. However, unfertilized eggs can store large quantities of mRNA for months or years in an inactive form that abruptly and rapidly become active within minutes after fertilization. The timing of translation is thus regulated.

## 2.16 Insulators and Boundaries

Each gene is embedded within a chromosomal environment of other DNA sequences that have the potential to affect its expression ([Bell et al. 2001, Cai and Shen 2001](#)). For example, regulatory elements (enhancers or silencers) associated with nearby genes could disrupt normal expression of a gene. To combat the intrusion of extraneous regulatory elements upon a specific gene, specialized DNA sequences called **insulators** provide a barrier against influences from

surrounding DNA sequences. Insulators regulate gene activity by blocking enhancer–promoter interactions when positioned between the enhancer and promoter (Cai and Shen 2001). Insulators also have the ability to protect against “position effects.” When genes are moved from their normal site in the chromosome, they may be expressed differently in the new location. This differential expression is often the case with **transgenes** (foreign genes artificially introduced into an organism) (Geyer 1997). This variability in expression may be due to the nearness of an enhancer or silencer or the presence of nearby inactive heterochromatin.

## 2.17 Chromosome or Gene Imprinting in Insects

**Imprinting** is a *reversible*, differential marking of genes or chromosomes that is determined by the sex of the parent from whom the genetic material is inherited. Imprinting is one type of a broader phenomenon called **epigenetics**, a term that indicates there are changes in the phenotype and gene expression patterns that cannot be explained by changes at the DNA sequence level or by classical genetics (Lyco and Maleszka 2011). “Epi” means above, so epigenetics means that nongenetic changes that affect development and phenotype can be heritable, self-perpetuating, and reversible (Allis et al. 2007). Epigenetics may involve DNA methylation, histone modification, nucleosome location, or noncoding RNA. Imprinting is one type of epigenetic change to genomes (Boxes 2.3, 2.4).

Imprinting was first observed in the study of insect chromosomes. In the citrus mealybug, *Planococcus citri* (Hemiptera: Pseudococcidae), for example, both males and females develop from fertilized eggs, and there are no sex chromosomes. In females, all of the chromosomes remain functional, but in male embryos one haploid set of chromosomes becomes inactivated and the set that is inactivated is derived from the father (Brown and Nelson-Rees 1961). As a result, the males are functionally haploids. In these mealybugs, the chromosomes derived from the male parent are *undermethylated* compared with the chromosomes derived from female parents (Bongiorni et al. 1999, Buglia et al. 1999). Thus, one method of imprinting DNA involves methylation of DNA.

**Methylation** of cytosines at the carbon 5 positions of CpG dinucleotides is common in prokaryotes and eukaryotes (Colot and Rossignol 1999, Ng and Bird 1999). In prokaryotes, methylation is apparently part of a defense system against invading DNA parasites. However, in eukaryotes, methylation can be associated with inhibiting transcription initiation, arresting transcript elongation, serving as a signal for imprinting, and suppressing homologous recombination (Colot and Rossignol 1999). Usually, methylated DNA is inactive or

**Box 2.4 Genetics and Epigenetics of DNA**

Genetics	Epigenetics
Mutations in the DNA are inherited in the soma and germ line.	Changes in chromatin structure such as histone modifications or effects of small noncoding RNAs are potentially reversible. Some modifications are limited to the soma, whereas others may be transmitted through the germ line. These modifications regulate access to or processivity of transcription of the DNA. Nuclear inheritance that is not based on differences in DNA sequences.

expressed at a very low level. However, DNA methylation is not always the cause of gene inactivity.

Methylation of DNA in insects seems to occur much less often than in mammals (Lyko and Maleszka 2011). DNA methylation has been reported in hemipterans, including *Megoura viciae*, *Planococcus lilacius*, *Pseudococcus calceolariae*, *Pseudococcus obscurus*, and *Myzus persicae*; orthopterans, including *Locusta migratoria*, *Eyprepocnemis plorans*, *Pyrgomorpha conica*, *Gryllotalpa fossor*, and *Baetica ustalata*; and the lepidopteran *B. mori* (Manicardi et al. 1994, Regev et al. 1998). Differential expression of alleles due to imprinting has been observed in *D. melanogaster* (Golic et al. 1998, Regev et al. 1998), but the subject of DNA methylation in *D. melanogaster* has been controversial (Gowher et al. 2000, Lyko et al. 2000). Gowher et al. (2000) estimated that *D. melanogaster* DNA has ≈50 times less methylcytosine than DNA of mammals, thereby explains why it is difficult to detect methylation in *Drosophila*. DNA methylation seems to be lacking in some Diptera (*Drosophila virilis*, *Sciara coprophila*, *Musca domestica*, and *Sarcophaga bullata*) or is present at only low levels (*Culex bitaeniorhynchus*, *Chironomus plumosus*, *Anopheles maculipennis*, and *Aedes albopictus*) (Regev et al. 1998, Tweedie et al. 1999).

The sequencing of whole genomes of insects has resulted in the recognition that a conserved gene family, consisting of three subfamilies (DNA methyltransferases), involved in methylation is found in some insects. The standard DNA methyltransferase set consists of one copy of subfamilies 1 and 2 and three versions of subfamily 3. However, some insects have lost one or more of the subfamilies. For example, the silkworm has the first two subfamilies, but it has lost the third subfamily, whereas other insects have expanded the number of their copies in subfamily 1 (Lyko and Maleszka 2011). Species within the same order may vary in the number and type of the three subfamilies of DNA

methyltransferases. The hymenopterans *Apis mellifera* and *Nasonia vitripennis* have different numbers of subfamily 1 in their genomes (two and three, respectively). When additional arthropod genomes are completely sequenced and their methyltransferase genes discovered, we understand more about the role of methylation as an epigenetic mechanism in arthropods.

The ability to modify gene expression through epigenetics adds another method by which an insect can respond to environmental conditions. For example, in honeybees, DNA methylation is important in differential gene expression in castes (Elango et al. 2009), in its social interactions, and in complex brain functions (Wang et al. 2006, Kronforst et al. 2008, Kucharski et al. 2008, Lyko et al. 2010). Queen and worker honeybees are behaviorally and reproductively different despite their identical genome sequences. However, the intake of royal jelly and the methylation of DNA make queens very different. Methylation patterns of 550 genes in the brains of queens and workers were found to be different (Lyko et al. 2010). Thus, despite workers and queens having identical genomes, their phenotype, biology, and behavior are dramatically different. The methylome of the silkworm has been analyzed and, as seems common in insects, a relatively low level of DNA methylation is found (Xiang et al. 2010). The silkworm has only two DNA methyltransferase genes.

## 2.18 Eukaryotic Genomes and Evolution

The discovery of split genes and RNA splicing was a critically important discovery and has elicited considerable thought regarding the origin and evolution of eukaryotic genomes (Deamer and Szostak 2010, Atkins et al. 2011, Darnell 2011). Gene regulation, and especially RNA splicing and DNA methylation, is probably central to understanding the development of complex multicellular eukaryotic organisms. Alternative RNA splicing produces multiple mRNAs that encode different proteins (Sharp 1994). The spliceosome process for excising introns is probably as old as the ribosomal process for translation. Thus, the eukaryotic cell has two compartments: the nucleus, where the spliceosome processes pre-mRNAs by RNA catalysis; and the cytoplasm, where the ribosome translates mRNAs by RNA catalysis.

Evolution in eukaryotes by changes in RNA processing is aided by the availability of intron-rich, repeat-rich genomes, allowing new gene products to evolve based on changes in RNA processing (Herbert and Rich 1999). The eukaryotic genome can be thought of as a “junkyard” in which solutions to any number of problems can occur with the assembly of old components into new combinations by changes in RNA processing.

Unlike prokaryotes, eukaryotic organisms make RNA that differs from the DNA in the genome. This difference allows genomic information to be affected by information derived from the environment through epigenetic modifications (Lyko and Maleszka 2011). For example, numerous gene products can act on an mRNA; some affect nucleotide sequence, whereas others can change the half-life and translatability. Thus, the way in which pre-mRNA is processed in eukaryotes has led to new evolutionary opportunities (Herbert and Rich 1999). A full understanding of the evolution of eukaryotes seems to require knowledge of the continual interplay between RNA and DNA. Box 2.2 provides an overview of the types of RNAs important in the functioning of a typical eukaryotic cell. This list may have to be modified as scientists learn more about the roles of RNAs in cell biology. The organization of DNA in nuclear chromosomes and in cytoplasmic mitochondria is described in Chapter 3.

## References Cited

- Abel, T., and Maniatis, T., (1989). Gene regulation. Action of leucine zippers. *Nature* **341**: 24–25.
- Alberts, B., Bray, D., Hopkin, K., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P., (2010). *Essential Cell Biology*, 3 ed. Taylor & Francis, New York.
- Allis, C.D., Jenuwein, T., and Reinberg, D., (2007). Overview and concepts. In: Ellis, C.D., Jenuwein, T., Reinberg, D., and Caparros, M.L., Eds., *Epigenetics*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 23–58.
- Atkins, J.F., Gesteland, R.F., and Cech, T.R., (2011). *RNA Worlds: From Life's Origins to Diversity in Gene Regulation*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Baker, D., (2000). A surprising simplicity to protein folding. *Nature* **405**: 39–42.
- Baker, M., (2011). Genomes in three dimensions. *Nature* **470**: 289–294.
- Baulcombe, D., (2001). Diced defense. *Nature* **409**: 295–296.
- Bell, A.C., and Felsenfeld, G., (1999). Stopped at the border: boundaries and insulators. *Curr. Opin. Genet. Dev.* **9**: 191–198.
- Bell, A.C., West, A.G., and Felsenfeld, G., (2001). Insulators and boundaries: versatile regulatory elements in the eukaryotic genome. *Science* **291**: 447–495.
- Belles, X., (2010). Beyond *Drosophila*: RNAi in vivo and functional genomics in insects. *Annu. Rev. Entomol.* **55**: 111–128.
- Belles, X., Cristina, A.S., Tanaka, E.D., Rubio, M., and Piulachs, M.D., (2010). Insect microRNAs: from molecular mechanisms to biological roles. In: Gilbert, L.I. Ed., *Insect Molecular Biology and Biochemistry* Academic. San Diego, CA, pp. 30–56.
- Ben-Shem, A., de Loubrese, N.G., Melnikov, S., Jenner, L., Yusopova, G., and Usupov, M., (2011). The structure of the eukaryotic ribosome at 3.0 Å resolution. *Science* **334**: 1524–1529.
- Blackwood, E.M., and Kadonaga, J.T., (1998). Going the distance: a current view of enhancer action. *Science* **281**: 60–63.
- Bongiorni, S., Cintio, O., and Prantero, G., (1999). The relationship between DNA methylation and chromosome imprinting in the coccid *Planococcus citri*. *Genetics* **151**: 1471–1478.
- Brown, S.W., and Nelson-Rees, W.A., (1961). Radiation analysis of a lecanoid genetic system. *Genetics* **46**: 983–1007.
- Buglia, G., Predazzi, V., and Ferraro, M., (1999). Cytosine methylation is not involved in the heterochromatization of the paternal genome of mealybug *Planococcus citri*. *Chromosome Res.* **7**: 71–73.

- Cai, H.N., and Shen, P., (2001). Effects of *cis* arrangement of chromatin insulators on enhancer-blocking activity. *Science* **291**: 493–495.
- Carter, R., and Drouin, G., (2009). Structural differentiation of the three eukaryotic RNA polymerases. *Genomics* **94**: 388–396.
- Cech, T.R., (2000). The ribosome is a ribozyme. *Science* **289**: 878.
- Cernilgar, F.M., Onorati, M.C., Kothe, G.O., Burroughs, A.M., Parsi, K.M., and Breiling, A., et al. (2011). Chromatin-associated RNA interference components contribute to transcriptional regulation in *Drosophila*. *Nature* **480**: 391–395.
- Chouard, T., (2011). Breaking the protein rules. *Nature* **471**: 151–153.
- Cockell, M., and Gasser, S.M., (1999). Nuclear compartments and gene regulation. *Curr. Opin. Genet. Dev.* **9**: 199–205.
- Colot, V., and Rossignol, J.L., (1999). Eukaryotic DNA methylation as an evolutionary device. *BioEssays* **21**: 402–411.
- Cooper, G.M., (2000). *The Cell: A Molecular Approach*. ASM Press, Washington, DC.
- Cramer, P., Bushnell, D.A., Fu, J., Gnatt, A.L., Maier-Davis, B., and Thompson, N.E., et al. (2000). Architecture of RNA polymerase II and implications for the transcription mechanism. *Science* **288**: 640–649.
- Darnell, J., (2011). *RNA. Life's Indispensable Molecule*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Deamer, D., and Szostak, J.W., (2010). *The Origins of Life*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Denton, M., and Marshall, C., (2001). Laws of form revisited. *Nature* **410**: 417.
- Dever, T.E., (1999). Translation initiation: adept at adapting. *Trends Biochem. Sci.* **24**: 398–403.
- Dieci, G., Preti, M., and Montanini, B., (2009). Eukaryotic snoRNAs: a paradigm for gene expression flexibility. *Genomics* **94**: 83–88.
- Doma, M.K., and Parker, R., (2007). RNA quality control in eukaryotes. *Cell* **131**: 660–668.
- Elango, N., Hunt, B.G., Goodisman, M.A.D., and Yi, S.V., (2009). DNA methylation is widespread and associated with differential gene expression in castes of the honeybee, *Apis mellifera*. *Proc. Natl. Acad. Sci. USA* **106**: 11206–11211.
- Featherstone, C., (1998). Coming to grips with the Golgi. *Science* **282**: 2172–2174.
- Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., and Mello, C.C., (1998). Potent and specific genetic interference by double stranded RNA in *Caenorhabditis elegans*. *Nature* **391**: 806–811.
- Geyer, P.K., (1997). The role of insulator elements in defining domains of gene expression. *Curr. Opin. Genet. Dev.* **7**: 242–248.
- Golic, K.G., Golic, M.M., and Pimpinelli, S., (1998). Imprinted control of gene activity in *Drosophila*. *Curr. Biol.* **8**: 1273–1276.
- Gowher, H., Leismann, O., and Jeltsch, A., (2000). DNA of *Drosophila melanogaster* contains 5-methylcytosine. *EMBO J.* **19**: 6918–6923.
- Graveley, B.R., (2001). Alternative splicing: increasing diversity in the proteomic world. *Trends Genet.* **17**: 100–107.
- Grunstein, M., (1997). Histone acetylation in chromatin structure and transcription. *Nature* **389**: 349–352.
- Grunwald, D., Singer, R.H., and Rout, M., (2011). Nuclear export dynamics of RNA-protein complexes. *Nature* **475**: 333–341.
- Guo, M., and Mount, S.M., (1995). Localization of sequences required for size-specific splicing of a small *Drosophila* intron *in vitro*. *J. Mol. Biol.* **253**: 426–437.
- Harrison, S.C., (1991). A structural taxonomy of DNA-binding domains. *Nature* **353**: 715–719.
- Harshman, L.G., and James, A.A., (1998). Differential gene expression in insects: transcriptional control. *Annu. Rev. Entomol.* **43**: 671–700.
- Hastings, M.L., and Krainer, A.R., (2001). Pre-mRNA splicing in the new millennium. *Curr. Opin. Cell Biol.* **13**: 302–309.

- He, L., and Hannon, G.J., (2004). MicroRNAs: small RNAs with a big role in gene regulation. *Nat. Rev. Genet.* **5**: 522–531.
- Henikoff, S., and Shilatifard, A., (2011). Histone modification: cause or cog? *Trends Genet.* **27**: 389–396.
- Herbert, A., and Rich, A., (1999). RNA processing and the evolution of eukaryotes. *Nat. Genet.* **21**: 265–269.
- Hilleren, P., and Parker, R., (1999). Mechanisms of mRNA surveillance in eukaryotes. *Annu. Rev. Genet.* **33**: 229–260.
- Hoffman, A., Oelgeschlager, T., and Roeder, R.G., (1997). Considerations of transcriptional control mechanisms: do TFIID–core promoter complexes recapitulate nucleosome-like functions? *Proc. Natl. Acad. Sci. USA* **94**: 8928–8935.
- John, S., and Workman, J.L., (1998). Just the facts of chromatin transcription. *Science* **282**: 1836–1837.
- Kiseleva, E.V., (1989). Secretory protein synthesis in *Chironomus* salivary gland cells is not coupled with protein translocation across endoplasmic reticulum membranes. *FEBS Lett.* **257**: 251–253.
- Klinge, S., Voigts-Hoffmann, F., Leibundgut, M., Arpagaus, S., and Ban, N., (2011). Crystal structure of the eukaryotic 60 S ribosomal subunit in complex with initiation factor 6. *Science* **334**: 941–948.
- Kozak, M., (1999). Initiation of translation in prokaryotes and eukaryotes. *Gene* **234**: 187–208.
- Kriwacki, R.W., and Yoon, M.K., (2011). Fishing in the nuclear pore. *Science* **333**: 44–45.
- Kronforst, M.R., Gilley, D.C., Strassmann, J.E., and Queller, D.C., (2008). DNA methylation is widespread across social Hymenoptera. *Curr. Biol.* **19**: R287–R288.
- Kucharski, R., Maleszka, J., Foret, S., and Maleszka, R., (2008). Nutritional control of reproductive status in honeybees via DNA methylation. *Science* **319**: 1827–1830.
- Kyrpides, N.C., and Woese, C.R., (1998). Universally conserved translation initiation factors. *Proc. Natl. Acad. Sci. USA* **95**: 224–228.
- Lee, T.I., and Young, R.A., (2000). Transcription of eukaryotic protein-coding genes. *Annu. Rev. Genet.* **34**: 77–137.
- Lee, D., Redfern, O., and Orengo, C., (2007). Predicting protein function from sequence and structure. *Nat. Rev. Mol. Cell Biol.* **8**: 995–1005.
- Lipshitz, H.D., and Smibert, C.A., (2000). Mechanisms of RNA localization and translational regulation. *Curr. Opin. Genet. Dev.* **10**: 476–488.
- Lyko, F., Ramasahoye, B.H., and Jaenisch, R., (2000). DNA methylation in *Drosophila melanogaster*. *Nature* **408**: 538–539.
- Lyko, F., Foret, S., Kucharski, R., Wolf, S., Falckenbach, C., and Maleszka, R., (2010). The honey bee epigenomes: differential methylation of brain DNA in queens and workers. *PLoS Biol.* **8**: e1000506.
- Lyko, F., and Maleszka, R., (2011). Insects as innovative models for functional studies of DNA methylation. *Trends Genet.* **27**: 127–131.
- Maas, S., and Rich, A., (2000). Changing the genetic information through RNA editing. *BioEssays* **22**: 790–802.
- Manicardi, G.C., Bizzaro, D., Assoni, P., and Bianchi, U., (1994). Cytological and electrophoretic analysis of DNA methylation in the holocentric chromosomes of *Megoura viciae* (Homoptera, Aphidiidae). *Genome* **37**: 625–630.
- Maquat, L.E., and Carmichael, G.G., (2001). Quality control of mRNA function. *Cell* **104**: 173–176.
- Matlack, K.E.S., Mothes, W., and Rapoport, T.A., (1998). Protein translocation: tunnel vision. *Cell* **92**: 381–390.
- Matzke, M., Matzke, A.J.M., and Kooter, J.M., (2001). RNA: Guiding gene silencing. *Science* **293**: 1080–1083.
- Mito, T., Nakamura, T., Bando, T., Ohuchi, H., and Oji, S., (2011). The advent of RNA interference in entomology. *Entomol. Sci.* **14**: 1–8.

- Mykles, D.L., (1999). Structure and function of arthropod proteasomes. *Mol. Biol. Rep.* **26**: 103–111.
- Nakielny, S., and Dreyfuss, G., (1999). Transport of proteins and RNAs in and out of the nucleus. *Cell* **99**: 677–690.
- Ng, H.H., and Bird, A., (1999). DNA methylation and chromatin modification. *Curr. Opin. Genet. Dev.* **9**: 158–163.
- Nissen, P., Hansen, J., Ban, N., Moore, P.B., and Steitz, T.A., (2000). The structural basis of ribosome activity in peptide bond synthesis. *Science* **289**: 920–930.
- Nurse, P., and Hayles, J., (2011). The cell in an era of systems biology. *Cell* **144**: 850–854.
- Pennisi, E., (1997). Opening the way to gene activity. *Science* **275**: 155–157.
- Pennisi, E., (2011). Does a gene's location in the nucleus matter? *Science* **334**: 1050–1051.
- Preiss, T., and Hentze, M.W., (1999). From factors to mechanisms: translation and translational control in eukaryotes. *Curr. Opin. Genet. Dev.* **9**: 515–521.
- Radford, S.E., (2000). Protein folding: progress made and promises ahead. *Trends Biochem. Sci.* **25**: 611–618.
- Reddy, P.S., and Corley, R.B., (1998). Assembly, sorting and exit of oligomeric proteins from the endoplasmic reticulum. *BioEssays* **20**: 546–554.
- Reed, R., (2000). Mechanisms of fidelity in pre-mRNA splicing. *Curr. Opin. Cell Biol.* **12**: 340–345.
- Regev, A., Lamb, M.J., and Jablonka, E., (1998). The role of DNA methylation in invertebrates: developmental regulation or genome defense? *Mol. Biol. Evol.* **15**: 880–891.
- Riddiford, L.M., Palli, S.R., and Hiruma, K., (1990). Hormonal control of sequential gene expression in  *Manduca epidermis*. In: Epple, A., and Scanes, C.G., Eds., *Progress in Comparative Endocrinology*, Wiley-Liss, New York, pp. 226–231.
- Rothman, J.E., (1994). Mechanisms of intracellular protein transport. *Nature* **372**: 55–63.
- Sharp, P.A., (1994). Split genes and RNA splicing. Nobel Lecture. *Cell* **77**: 805–815.
- Singer, R.H., and Green, M.R., (1997). Compartmentalization of eukaryotic gene expression: causes and effects. *Cell* **91**: 291–294.
- Skalsky, R.L., Vanlandingham, D.L., Scholle, F., Higgs, S., and Cullen, B.R., (2010). Identification of microRNAs expressed in two mosquito vectors, *Aedes albopictus* and *Culex quinquefasciatus*. *BMC Genomics* **11**: 119.
- Storz, G., (2002). An expanding universe of noncoding RNAs. *Science* **296**: 1260–1262.
- Stuart, D.I., and Jones, E.Y., (1997). Cutting complexity down to size. *Nature* **386**: 437–438.
- Stutz, F., and Rosbash, M., (1998). Nuclear RNA export. *Genes Dev.* **12**: 3303–3319.
- Surdej, P., Riedl, A., and Jacobs-Lorena, M., (1994). Regulation of mRNA stability in development. *Annu. Rev. Genet.* **28**: 263–282.
- Taft, R.J., Glazov, E.A., Cloonan, N., Simons, C., Stephen, S., and Faulkner, G.J., et al. (2009). Tiny RNAs associated with transcription start sites in animals. *Nat. Genet.* **31**: 572–578.
- Tweedie, S., Ng, H.H., Barlow, A.L., Turer, B.M., Hendrich, B., and Bird, A., (1999). Vestiges of a DNA methylation system in *Drosophila melanogaster*? *Nat. Genet.* **23**: 389–390.
- Wang, Y., (2006). Functional CpG methylation system in a social insect. *Science* **314**: 645–647.
- Wickner, S., Maurizi, M.R., and Gottesman, S., (1999). Posttranslational quality control: folding, refolding and degrading proteins. *Science* **286**: 1888–1893.
- Workman, J.L., and Kingston, R.E., (1998). Alteration of nucleosome structure as a mechanism of transcriptional regulation. *Annu. Rev. Biochem.* **67**: 545–579.
- Xiang, H., Zhu, J., Chen, Q., Dai, F., Li, X., and Li, M., et al. (2010). Single base-resolution methylome of the silkworm reveals a sparse epigenomic map. *Nat. Biotechnol.* **5**: 516–520.
- Zhang, Y., Zhou, X., Ge, X., Jiang, J., Li, M., and Jia, S., et al. (2009). Insect-specific microRNA involved in the development of the silkworm *Bombyx mori*. *PLoS ONE* **4**: e4677.

# Nuclear and Extranuclear DNA in Insects

## Chapter Outline

- 3.1 Overview 71
- 3.2 Introduction 72
- 3.3 C-Value Paradox: Is it Real? 73
- 3.4 Repetitive DNA is Common in Insects 74
- 3.5 Composition of Insect DNA 76
- 3.6 Chromosomes are DNA Plus Proteins 76
- 3.7 Packaging Long, Thin DNA Molecules into Tiny Spaces 77
- 3.8 Structure of the Nucleus 79
- 3.9 Euchromatin and Heterochromatin 80
- 3.10 Centromeres 81
- 3.11 Telomeres 82
- 3.12 Chromosomes during Mitosis and Meiosis 84
  - 3.12.1 Mitosis 84
  - 3.12.2 Meiosis 87
- 3.13 Chromosome Damage 90
- 3.14 Polyteny 92
- 3.15 Chromosomal Puffing 93
- 3.16 B Chromosomes 94
- 3.17 Sex Chromosomes 94
- 3.18 Extranuclear Inheritance in Mitochondrial Genes 95
- 3.19 Transposable Elements are Ubiquitous Agents that Alter Genomes 99
- References Cited 99

## 3.1 Overview

DNA is found in chromosomes, complex structures in the nucleus of insect cells. Each chromosome contains a single linear DNA molecule combined with a variety of proteins, including histones. DNA and histones form structures called nucleosomes. Nucleosomes are arranged in a higher level of organization that serves to condense the DNA. Chromosomes visible by light microscopy may have discrete structures called centromeres to which spindle fibers attach

so that chromosomes are distributed to the daughter cells in an orderly manner during mitosis and meiosis, although some insect chromosomes have diffuse centromeres along the entire chromosome. The centromeres of eukaryotic chromosomes are complex regions that play a fundamental role in chromosome movement. Chromosomes always have telomeres at the ends; the telomeres are specialized structures that help maintain the ends of chromosomes in a stable state. In polyteny tissues such as *Drosophila* salivary glands, banding patterns are visible by light microscopy after staining. The bands staining lightly are called euchromatin, and the bands staining darkly are called heterochromatin. Euchromatin and heterochromatin represent active and less active regions, respectively, during somatic interphase. In most insects, there are two copies of each chromosome and hence two copies of each gene ( $2n$  = diploid complement), although polyteny or gene amplification can increase the copy number of chromosomes or genes. The reproductive cells (eggs, sperm) contain a single copy (haploid complement =  $n$ ) of chromosomes.

Chromosomes contain single-copy DNA (i.e., DNA present only once in the genome that codes for a polypeptide), highly repetitive DNA, moderately repetitive DNA, intergenic regions, centromeres, and telomeres. Intergenic regions contain transcription and regulatory information. Many types of transposable elements have been identified in both the coding and noncoding regions of arthropod chromosomes. Transposable elements are DNA sequences capable of moving within and among chromosomes, and they may make copies of themselves as they move. Transposable elements, or their defective derivatives, make up a significant portion of the middle-repetitive DNA in insect chromosomes. Arthropods also contain chromosomal DNA in the mitochondria that are located in the cytoplasm of the cell. Mitochondria are considered remnants of a microbial endosymbiont, and much of the original genetic information in this symbiont, turned organelle, has been lost or transferred to the nuclear genome of its host.

### 3.2 Introduction

DNA in insects is organized into chromosomes, which are complex structures. The number of chromosomes in insects varies, with the haploid number ranging from 1 to 221. Insect chromosomes were among the first chromosomes to be investigated, and studies of insect chromosomes have provided fundamental advances in genetics, including the initial proof that genes are on chromosomes and that spindle fibers exist in living cells and are not fixation artifacts (Ault 1996).

The complete sequence of the *Drosophila* genome, only the third eukaryotic genome to be sequenced after the yeast *Saccharomyces cerevisiae* and

the nematode *Caenorhabditis elegans*, became available in 2000 (Adams et al. 2000, Celniker 2000). The *Drosophila* genome is  $\approx$ 180 megabases (Mb) and contains  $\approx$ 13,600 genes and  $\approx$ 60 Mb of noncoding DNA (Celniker 2000). Because much of the noncoding DNA cannot be cloned, the genome sequences obtained primarily are from the euchromatin regions that contain protein-coding genes. The function of the noncoding, heterochromatic DNA is being deciphered now. Much of this so-called “junk” DNA is actually involved in gene and developmental regulation.

### 3.3 C-Value Paradox: Is it Real?

There seemed to be a great deal more DNA in eukaryotic organisms than is actually needed to code for the number of genes estimated to occur in the nuclear genome. This discrepancy is known as the **C-value paradox**. Genome size varies widely *among* insect species, with up to 250-fold differences in C values known (Petrov et al. 2000). For example, the locust *Schistocerca gregaria* has a C value of 9,300,000 kilobases (kb), 52-fold more DNA than *Drosophila melanogaster*, but is unlikely to have 52 times as many genes (Wagner et al. 1993). Among 37 species of tenebrionid beetles, nuclear DNA content varies by five-fold (Juan and Petitpierre 1991).

Genome size also can vary *within* species. For example, diploid cells in the mosquito *Aedes albopictus* contain 0.18–6 picograms (pg) of DNA, and C values vary by three-fold (0.62–1.6 pg) among different populations of *Ae. albopictus* (Kumar and Rai 1990). The amount of DNA in insect cells may vary because many tissues are polyploid, with different tissues having different degrees of ploidy.

**Polyplody** occurs when the amount of DNA in an organism increases over the usual diploid ( $2n$ ) amount, usually by duplicating the **number of chromosomes**, perhaps to  $3n$  or  $4n$  or more. Polyplody can occur throughout an organism’s cells or in just some tissues. A few insects are polyploid in all tissues (Otto and Whitton 2000), but many insects only have polyploid tissues within a diploid body. For example, the diploid blood cells of the silkworm, *Bombyx mori*, contain 1 pg of DNA/blood cell, but a polyplloid silk-gland cell in the same insect contains 170,000 pg of DNA. DNA content within cells also varies with developmental stage. At metamorphosis, the amount of DNA in *B. mori* declines by 81% after adults emerge from the pupal stage, which is probably due to histolysis of the larval silk glands and other polyploid cells.

Noncoding DNA can constitute 30% to >90% of the insect genome. This noncoding DNA has been called “**junk**” or “**parasitic**,” or “**selfish**,” and the role of this DNA has been the subject of intense study in the past 15 years and is still in

the process of being analyzed. There are several hypotheses to explain its persistence in genomes. One hypothesis suggested that the noncoding DNA performs essential functions, such as global regulation of gene expression and development. According to this hypothesis, the junk DNA is functional and deletions of such DNA would have a deleterious effect. A second hypothesis was that the noncoding DNA is useless, but that it is maintained because it is linked physically to functional genes; the excess DNA is not eliminated because it does not affect fitness of the organism and can be maintained indefinitely in the population. A third hypothesis suggested that the noncoding DNA is a functionless parasite that accumulates and is actively maintained by selection. A fourth hypothesis is that the DNA has a structural function, perhaps for compartmentalizing genes within the nucleus, or for maintaining a structural organization (nucleoskeleton) within the nucleus (Manuelidis 1990, Manuelidis and Cher 1990). Recent studies indicate that much of the so-called junk DNA does have a function in gene expression and development as well as playing a role in chromosome organization in the nucleus. Other junk is probably the remnants of parasitic DNA (transposable elements) that have been degraded but not yet eliminated from the genome.

The C-value paradox remains a topic of study because not all noncoding DNA has been shown to have a function and, unless the noncoding DNA has a function, such DNA constitutes a “load” upon the insect and should be lost over evolutionary time. Petrov et al. (1996) provided evidence that nonessential DNA is lost at a higher rate in *Drosophila* species than in mammalian species, suggesting that differences in genome size may result from persistent differences between organisms. Petrov et al. (2000) provided additional support for this hypothesis by comparing DNA loss in two insect genera (*Laupala* [crickets] and *Drosophila*) with different genome sizes. The crickets have a genome size an order of magnitude larger than that of *Drosophila* and eliminate nonessential DNA one-fortieth as quickly.

### 3.4 Repetitive DNA is Common in Insects

Much of the noncoding, heterochromatic junk DNA in insects is **repetitive DNA**—specific nucleotide sequences that are repeated several times to millions of times. Repetitive DNA has been classified as “**highly repetitive**” (sequences repeated several hundred to several million times per genome) or as “**moderately repetitive**.” Highly repetitive DNA is found in and near centromeres, telomeres, and other heterochromatic regions. Middle-repetitive DNA sequences are repeated 100–10,000 times and include genes that code for ribosomal RNA (rRNA) and transfer RNA (tRNA). Middle-repetitive sequences are found in euchromatic regions, as well as in heterochromatic regions of the genome.

Species vary in the number of repeated elements in their genome. For example, *Drosophila melanogaster* has  $\approx$ 30% of its genome as heterochromatic (repetitive) DNA, but  $\approx$ 60% of the genome of *Drosophila nasutoides* is repetitive DNA. More than 90% of a genome can be repetitive DNA. Some insects, such as aphids, have small amounts of repetitive DNA, and this DNA could be associated with a faster developmental time (Ma et al. 1992).

Repetitive and single-copy DNA are present in two different patterns in insect genomes. The **short-period interspersion pattern** has single-copy DNA, with a length of 1000–2000 base pairs (bp) long, alternating with short (200–600 bp) and moderately long (1000–4000 bp) repetitive sequences. The house fly *Musca domestica*, the Australian sheep blowfly *Lucilia cuprina*, and the wild silk moth *Antheraea pernyi* have this pattern. **Long-period interspersion patterns** have long ( $>5600$  bp) repeats alternating with very long ( $>12$  kilobases [kb]) uninterrupted stretches of unique DNA sequences. This pattern is often found in species with small genomes (0.1–0.5 pg of DNA/haploid genome), including *D. melanogaster*, as well as in the aphid *Schizaphis graminum* (Ma et al. 1992), the midge *Chironomus tentans* (Wells et al. 1976); the flesh-fly *Sarcophaga bullata* (Samols and Swift 1979), the honeybee *Apis mellifera* (Crain et al. 1976) and the red flour beetle *Tribolium castaneum* (Brown et al. 1990).

Even within an insect family, genome organization can vary. Total DNA in the genome of four mosquito species varies from 0.186 to 0.899 pg, and the amount of repetitive elements varies from 0.009 to 0.150 pg of foldback DNA (Black and Rai 1988). The mosquito *Anopheles quadrimaculatus* has a long-period interspersion type of genome organization, but *Culex pipiens*, *Ae. albopictus*, and *Aedes triseriatus* have the short-period interspersion type. Generally, the amounts of fold-back, highly repetitive, and middle-repetitive DNA increase linearly with genome size in these mosquitoes. Intraspecific variation in the amount of highly repetitive DNA was found in *Ae. albopictus* colonies and is due to differences in the number or type of transposable elements. The amounts of repetitive DNA in mosquitoes varies from 20% in *An. quadrimaculatus* to 84% in *Ae. triseriatus* (Besansky and Collins 1992).

Satellite DNA is a type of highly repetitive DNA that differs sufficiently in its base composition from the majority of DNA in a eukaryotic species that it separates out as one or more distinct bands when DNA is isolated by centrifugation with cesium chloride. Satellite DNA is rich in either adenine + thymine (A + T) or in guanine + cytosine (G + C) sequences, and it is found in long tandem arrays within the heterochromatic regions of chromosomes.

### 3.5 Composition of Insect DNA

Insect DNA base ratios are lower than those found in vertebrates, with G + C comprising from 32 to 42% of the DNA, compared with 45% for vertebrates (Berry 1985). If base composition were random, 50% of the DNA would be G + C.

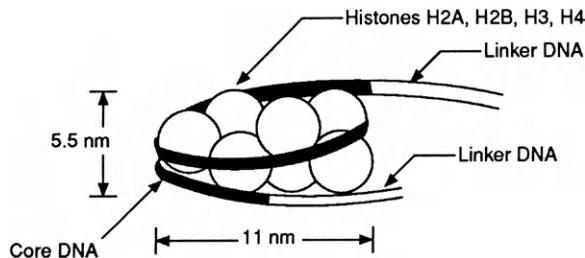
DNA in eukaryotes can occur in different configurations. Most genomic DNA exists in the B-helix form, but other configurations are known, including **triplex DNA**. In triplex DNA, the usual A-T and C-G base pairs of duplex DNA are present, but in addition a pyrimidine strand is bound in the major groove of the helix. DNA sequences that potentially can form triplex DNA structures seem to be common, are dispersed at multiple sites throughout the genome, and comprise up to 1% of the genome. Triplex DNA was identified in polytene chromosomes of *Chironomus tentans* and *D. melanogaster*, where it was found in the euchromatic bands (Burkholder et al. 1991).

### 3.6 Chromosomes are DNA Plus Proteins

Eukaryotes must organize and package their DNA in a sufficiently condensed form that it can fit into a very small space in the nucleus during mitosis and meiosis. Yet, this packaging must be compatible with the ability to separate the DNA strands and unwind the DNA helix during DNA replication and transcription during interphase. Furthermore, the packaging must occur rapidly. Precise and rapid replication of DNA is required in many tissues (Koshland and Strunnikov 1996, Nicklas 1997).

Eukaryotic genes are located on linear DNA molecules, with each chromosome containing a single long DNA molecule. In addition, each chromosome contains an approximately equal amount of proteins with different functions. The proteins include DNA and RNA polymerases and regulatory proteins associated with the DNA. At least five **histones** are associated with the DNA in structures called **nucleosomes** (Figure 3.1). Histone proteins (called H1, H2A, H2B, H3, and H4) contain  $\approx$ 100–200 amino acids, of which 20–30% are arginine and lysine. As a result, the histones have a positive charge that helps histones bind to DNA. Genes that code for the major histones are very highly conserved among eukaryotic species, indicating that they are nearly unchanged over billions of years. Histones are crucial to maintaining chromosome structure, and they may be crucial to the effective function of DNA as the genetic code (Jenuwein and Allis 2001).

Core DNA (the DNA in a nucleosome) is connected by **linker DNA** to the next nucleosome (Figures 3.1, 3.2B). The **core DNA** is protected from digestion



**Figure 3.1** DNA is condensed in eukaryotes by being packaged in nucleosomes, the first level of condensation. A nucleosome consists of core DNA wound around two molecules each of the histones H2A, H2B, H3, and H4. These eight histone molecules are called an octamer. Nucleosomes are connected to other nucleosomes by linker DNA. In addition, a single molecule of the histone H1 (not shown) binds in the linker and helps to condense the nucleosome. Nucleosomes are organized into structures called 30-nm fibers (see text and [Figure 3.2](#)).

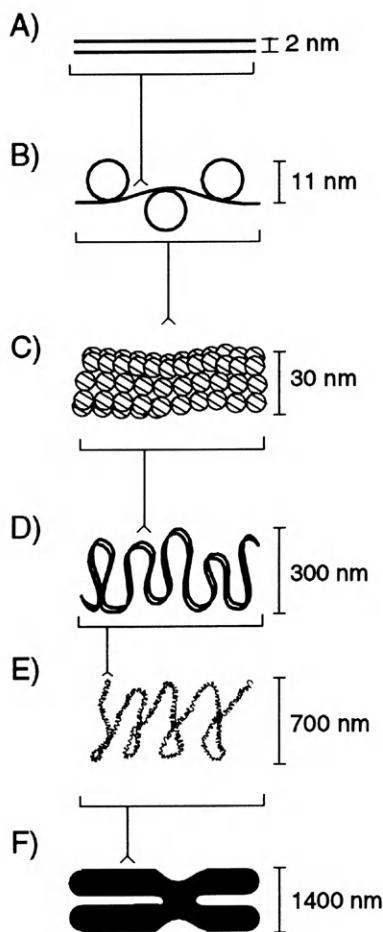
by restriction enzymes, but the linker DNA is vulnerable to these enzymes. The core DNA is wound  $\approx$ 1.75 times around a histone octamer, consisting of two molecules each of H2A, H2B, H3, and H4. A single molecule of histone H1 is associated with the linker DNA and is apparently helpful in compacting the nucleosome. Linker DNA lengths vary from 20 to 100 nucleotide (nt) pairs in different species and in different cell types within the same organism.

Origins of DNA replication in higher eukaryotes are larger and more complex than in prokaryotes (see Chapter 1). However, the underlying molecular processes are conserved ([Benbow et al. 1992](#)).

### 3.7 Packaging Long, Thin DNA Molecules into Tiny Spaces

Eukaryotes have to solve a serious packaging problem in their nuclei. If eukaryotic chromosomes were simply linear DNA molecules, the average length might be  $\approx$ 5 cm. If these long, thin chromosomes were tangled together inside the nucleus, replication would be difficult and separation of the intertwined chromosomes during mitosis could result in breakage of the chromosomes and subsequent loss of essential genetic information. Therefore, DNA needs to be condensed, yet packaged so that both DNA replication and transcription can occur without loss or damage. DNA packaging is achieved by a highly organized and hierarchical condensation scheme ([Figure 3.2](#)).

Eukaryotic DNA is **supercoiled**, i.e., the double helix is twisted around itself, which begins the condensation process. The next level of compaction is achieved by organizing the DNA into nucleosomes. Nucleosomes occur in a regular pattern, with linker or intervening DNA between each nucleosome ([Figure 3.2B](#)).



**Figure 3.2** DNA packing in eukaryotic chromosomes must be efficient to achieve a dramatic reduction in DNA length during mitosis and meiosis. This figure illustrates the method by which DNA is thought to be packed, although the organization of elements in D–F are controversial. Naked DNA (A), DNA in nucleosomes (B), 30-nm chromatin fiber (C), the 300-nm fiber made up of looped 30-nm fibers (D), the 700-nm supercoiled structure that comprises the arms of a metaphase chromosome (E), and the metaphase chromosome (F).

Nucleosomes reduce the length of DNA by a factor of  $\approx 6$  to a flexible beaded fiber.

Additional condensation of the DNA occurs when nucleosomes are condensed into a shorter thicker fiber, called the **30-nm fiber** (Figure 3.2C). The structural organization of this 30-nm fiber seems to be an irregular left-handed superhelix with six nucleosomes per turn. Another level of condensation is thought to occur in metaphase chromosomes (Figure 3.2D–F), although the details of this

compaction remain conjectural. [Figure 3.2D–F](#) suggests one model for packing the 30-nm fiber into the highly condensed form found in metaphase chromosomes. Somehow, the length of the chromosomal DNA is reduced by a factor of  $\approx 10,000$  in metaphase chromosomes.

### 3.8 Structure of the Nucleus

The nucleus contains chromosomes, RNAs, and nuclear proteins in an aqueous solution. It also seems to have an internal structure that organizes the chromosomes and localizes some nuclear functions to specific sites ([Kalhor et al. 2012](#), [Misteli 2012](#), [Sexton et al. 2012](#)). The most obvious organized region is the nucleolus, the site at which the ribosomal RNA (rRNA) genes are transcribed and ribosomal subunits are assembled.

Nucleoli are RNA-rich spherical bodies, not surrounded by a membrane, associated with specific chromosomal segments called the nucleolus organizer. The nucleolus contains multiple copies of tandem arrays of rRNA genes. There are four types of rRNA: 5S, 5.8S, 18S, and 28S. The 5.8S, 18S, and 28S rRNAs are transcribed as a single unit by RNA polymerase I, yielding a 45S precursor rRNA. The 45S pre-rRNA is processed into the 18S rRNA of the 40S (small ribosomal subunit) and into the 5.8S and 28S rRNAs of the 60S (large) ribosomal subunit. Transcription of the 5S rRNA, which is found in the 60S subunit, takes place outside the nucleolus and is catalyzed by RNA polymerase III. The nucleolus is particularly important during development, when a large number of rRNA genes are transcribed so that large numbers of ribosomes can be produced.

Much of the **heterochromatin** (chromosome regions that remain condensed during most of the interphase of the cell cycle and seem to contain mostly inactive genes or noncoding DNA) is localized at the edge of the nucleus, apparently because the heterochromatin binds to a protein of the inner nuclear membrane. Much of the heterochromatin consists of transposable elements. Heterochromatin is maintained by DNA methylation, histone modifications, and a cellular defense mechanism called RNAi ([Biemont 2009](#)). The presence of heterochromatin in centromeric and pericentromeric regions is important in cell division because it enables pairing of homologous chromosomes.

Active euchromatic DNA is arranged into discrete functional domains that are important in regulating gene expression. Thus, functional euchromatin is non-randomly distributed within the interphase nucleus. Apparently, each chromosome occupies a discrete region of the nucleus. The chromosomes are closely associated with the nuclear envelope at many sites, with their centromeres and telomeres clustered at opposite poles ([Cooper 2000](#)).

### 3.9 Euchromatin and Heterochromatin

Insect nuclear genomes have two types of chromatin during somatic interphase: **euchromatin** and **heterochromatin**. **Euchromatin** is uncoiled during interphase, presumably to allow for gene transcription. Euchromatin is condensed during mitosis, with a maximal condensation at metaphase. Euchromatin contains most of the protein-coding genes. In polytene salivary gland chromosomes of *Drosophila*, the darkly staining segments are heterochromatic (DNA that is condensed because it is not actively being transcribed) and the intervening less well-stained regions are euchromatic (regions that are genetically active and do not dye as well).

The term *heterochromatin* was coined originally to define the chromosome regions that remain condensed during most of the cell cycle and have a coiling cycle out of phase with the rest of the genome. Unlike euchromatin, heterochromatin exhibits maximal condensation in nuclei during interphase. Heterochromatin replicates late in the cell cycle, compared with euchromatin, and contains a considerable amount of middle- and highly repetitive DNA ([Weiler and Wakimoto 1995](#)).

In many organisms, large regions of the chromosome near the centromeres and the telomeres are heterochromatic, and these regions contain primarily middle- and highly repetitive DNA that plays an essential role in centromere and telomere function ([Henikoff 2000](#)). It is now thought that heterochromatin is not a type of DNA sequence but rather a “chromatin state” ([Jenuwein and Allis 2001](#)). Potentially, all parts of the genome could enter this state. The repression of transcription in heterochromatin seems to involve a set of proteins and RNA molecules, although the details of how they function remain limited ([Hennig 1999](#), [Leach et al. 2000](#), [Redi et al. 2001](#)).

Heterochromatin serves an important role in chromosome mechanics ([Wallrath 1998](#)). Without sufficient heterochromatin, chromosomes segregate inappropriately to daughter cells during mitosis. Strangely, a few genes normally located in heterochromatic regions are active, but become silenced or inactive if moved into euchromatic regions ([Eissenberg and Hilliker 2000](#)).

Heterochromatic regions in *D. melanogaster* can cause **position effect variegation** by inactivating (silencing) euchromatic genes that have been moved to regions adjacent to heterochromatin by chromosomal rearrangements ([Wallrath 1998](#)). A change in location of a gene within the nucleus significantly modifies the amount of “gene silencing,” perhaps due to its location within the nuclear compartment. Furthermore, foreign genes (transgenes) experimentally inserted into an insect’s genome can be silenced because they become

heterochromatinized. The original assumption was that the transgenes became inactive because they had been inserted into a heterochromatic site (Henikoff 2000). Thus, understanding the mechanism(s) by which heterochromatin forms is essential in improving the function of transgenes inserted into genetically modified insects developed for pest-management programs.

Heterochromatin is hypothesized to serve as a defense mechanism after parasitic DNA invades genomes (Henikoff 2000). In *Drosophila miranda*, heterochromatin forms at clusters of retrotransposons (a type of transposable element) that have recently invaded the genome. The transformation of chromosome regions into heterochromatin might prevent these invasive elements from functioning and causing damage to the genome.

Many functional genes do occur within heterochromatic regions in *Drosophila*. For example, the Y chromosome of *D. melanogaster* is heterochromatic, yet carries some genes that are required for male fertility (Gatti and Pimpinelli 1992).

### 3.10 Centromeres

Most chromosomes possess a centromere. The centromere is important in the organization of the developing spindle before mitosis or meiosis and the separation of the daughter chromosomes at anaphase. Chromosome fragments lacking centromeres, **acentric fragments**, do not get transmitted to daughter cells in an orderly manner, and the genetic information contained on them is eventually lost, which can be lethal.

Some species do not have localized centromeres; rather, the whole chromosome seems to have centromeric properties (**holocentric chromosomes**). If holocentric chromosomes are fragmented, each portion can attach to the spindle, and these fragments are not lost at mitosis. Holocentric chromosomes are found in the orders Hemiptera, Mallophaga, Anoplura, and Lepidoptera (White 1973). Centromeres are difficult to study (Tyler-Smith and Floridia 2000, Henikoff et al. 2001).

Analysis of a centromere in a *Drosophila* minichromosome indicated that the essential core of the centromere is a 220-kb region containing complex DNA. In addition, another 200 kb of DNA on either side is essential to centromeric function and contains highly repeated sequences (Murphy and Karpen 1995).

Analysis of a centromere from a standard chromosome confirmed that the *Drosophila* centromere spans 420 kb, more than 85% of which consists of two highly repeated satellite DNAs with the sequences AATAT and AAGAG. The remainder of the centromere consists of interspersed transposable elements, as

well as a nonrepetitive segment of AT-rich DNA (Sun et al. 1997). Both the repetitive and nonrepetitive sequences contribute to the centromere function.

The chromosomal region adjacent to the *Drosophila* centromere contains very long blocks of highly repetitive DNA in which simple sequences are repeated thousands of times (satellite DNA). There may be several different satellite DNA types in a given species; for example, three satellites are found near centromeric DNA of *Drosophila virilis*. One satellite type has a 5'-ACAAACT-3' repeat, the second satellite type is 5'-ATAAACT-3', and the third satellite type is 5'-ACAAATT-3'. These satellite DNAs apparently are not transcribed and they may bind proteins essential for centromere function. Likewise, for the Australian sheep blowfly *Lucilia cuprina*, several subfamilies of satellite DNA are present in the centromeric regions of the chromosomes, as well as in the sex chromosomes (Perkins et al. 1992).

Although each insect species normally has several types of satellite DNA in the centromere, two parasitic wasps, *Diadromus pulchellus* and *Eupelmus vuilleti*, have only one type (Bigot et al. 1990). In these two species, satellite DNA constitutes 15 and 25% of the genome, respectively. Likewise,  $\approx$ 50% of the genome of the tenebrionid *Tenebrio molitor*, consists of only one type of satellite DNA, and it is distributed evenly over the centromeric regions (Plohl et al. 1992). In the tenebrionid *Palorus ratzeburgii*,  $\approx$ 31% of the genome is a single type of satellite DNA (Ugarkovic et al. 1992).

### 3.11 Telomeres

The ends of the chromosomes have distinct structures called **telomeres**. Telomeres have two important functions: 1) maintain the length of chromosomes despite the inability of DNA polymerase to replicate linear DNA ends completely, and 2) distinguish natural chromosome ends from double-stranded breaks in DNA. The latter function, known as "capping," is important because DNA damaged by double-stranded breaks is attacked by repair and degradative enzymes. Telomeres also associate with one another, with the nuclear matrix, and with the nuclear envelope, an association that could be important in maintaining organization within the nucleus and for meiotic chromosome pairing (Mason and Biessmann 1995).

Molecular analyses indicate telomeres consist of a series of repeated nucleotides and proteins (Blackburn 1991, Wagner et al. 1993, Zakian 1989, de Lange 2001). As described in Chapter 1, DNA replication conventionally occurs only in the 5' to 3' direction and cannot be initiated without a primer, which is usually RNA. After primer removal, gaps would remain at the 5' ends of new DNA

strands in eukaryotes if it were not for telomeres. Telomeres prevent the loss of genetic information from the ends of chromosomes.

Telomeric DNA sequences and structure seem to be similar among many arthropods, including Hymenoptera, Lepidoptera, Trichoptera, Mecoptera, Coleoptera, Orthoptera, Isoptera, Blattodea, and Crustacea (Okazaki et al. 1993, Sahara et al. 1999). The telomeric sequence TTAGG was isolated from the silk moth, *B. mori* (Okazaki et al. 1993), where it is repeated over a 6- to 8-kb segment. TTAGG has been found only in telomeres from arthropods and could be ancestral in this group.

Not all insects have the TTAGG sequences in their telomeres, including several dipterans. For example, *Drosophila* species have non-long terminal repeat (LTR) retrotransposons (called HeT-A, TART, and TAHRE) that function as telomeres (Mason and Biessmann 1995, Pardue and DeBaryshe 2011). In *Drosophila*, chromosome length is maintained, despite incomplete DNA replication, by the addition of retropon sequences through transposition, but the retrotransposons do not transpose into the gene-rich euchromatic regions. It is not known whether *Drosophila* recruited existing retrotransposon elements to replace the "standard" telomeres or whether the retrotransposons represent a more ancient arrangement used by the earliest eukaryotes (Pardue and DeBaryshe 1999). Eickbush (1997) suggests that non-LTR retrotransposons gave rise to telomerases and that a parasite was recruited in early eukaryotes to supply this important function. Perhaps the *D. melanogaster* case can be viewed as a recent example of a similar recruitment event.

Two dipteran genera, *Chironomus* and *Anopheles*, and some dermopterans, hemipterans, and coleopterans also lack the TTAGG repeat, suggesting that the loss of the TTAGG sequences has occurred independently several times during insect evolution (Sasaki and Fujiwara 2000). In chironomids, a third type of chromosome termination occurs (Kamnert et al. 1997, Rosen and Edstrom 2000). In *Chironomus pallidivittatus* and *C. tentans*, the size of the repeats is unusually large, consisting of 340 bp and 350 bp, respectively. In *C. pallidivittatus*, the telomeric repeat units are present in blocks up to 200 kb.

**Telomere terminal transferase, or telomerase,** was discovered first in the protozoan *Tetrahymena*. Telomerase is a ribonucleoprotein, whose RNA component is essential to its function. Telomerase recognizes single-stranded oligonucleotides ending in -GGG3' in a variety of eukaryotes and adds GGGG sequences to the ends of chromosomes. Telomerase functions as a kind of reverse transcriptase, because its own RNA codes for a DNA sequence.

Adjacent to the telomeres are subtelomeric repetitive sequences. In most insects, recombination is relatively common in subtelomeric DNA. Subtelomeric

DNA is thought to act as a buffer zone to protect nearby genes from DNA loss during DNA replication, to promote the spread of short repeats between telomeres through recombination, to give rise to telomere-to-telomere associations that affect the three-dimensional arrangement of chromosomes in the interphase nucleus, and to promote meiotic pairing of chromosomal homologs (Kamnert et al. 1997).

Telomeres are crucial for the viability of the cell. If telomeres are damaged, progressive loss of DNA occurs, and the sticky ends of damaged chromosomes will bind to other chromosomes with sticky ends, resulting in chromosome abnormalities such as dicentric chromosomes (chromosomes with two centromeres), which can lead to chromosome breakage and loss. Our understanding of telomeres continues to increase and could continue to provide some surprises (Greider 1999).

### 3.12 Chromosomes during Mitosis and Meiosis

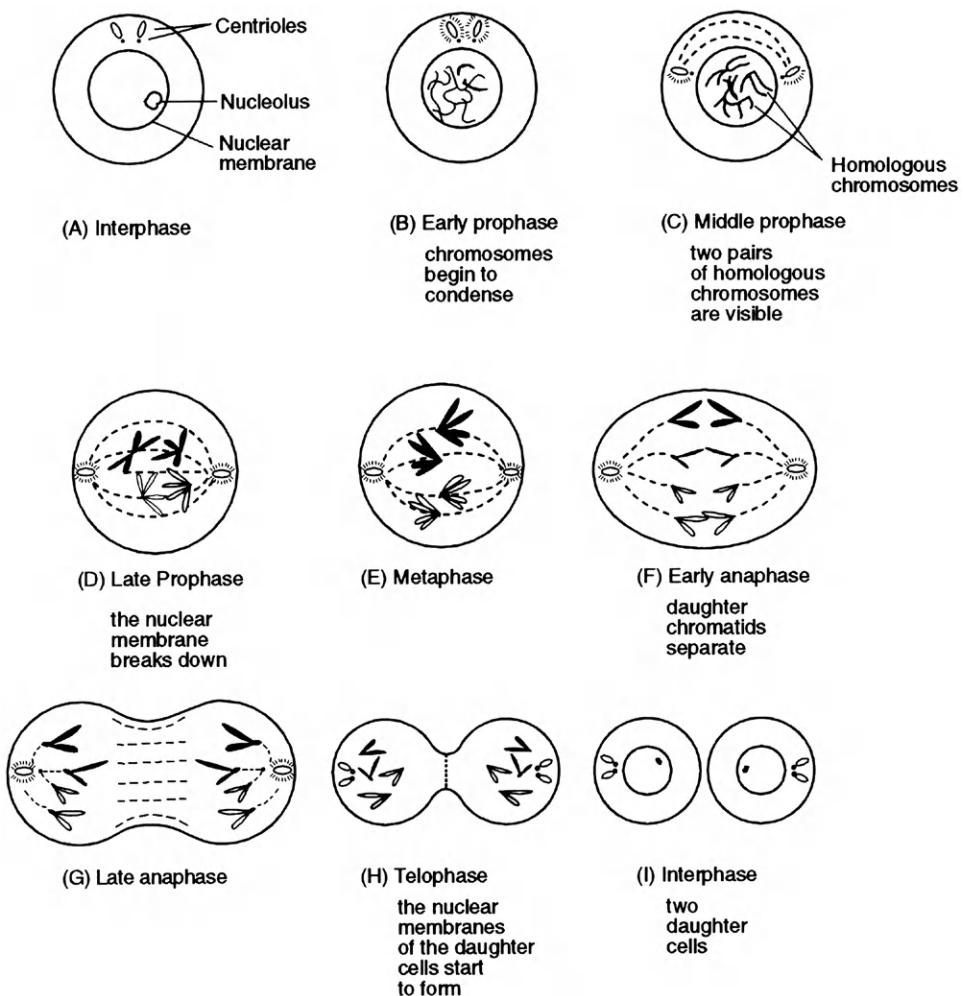
Chromosomes are visible by light microscopy during the cell divisions called mitosis and meiosis. The following reviews the basic aspects of these two types of cell division, both of which are essential for growth, development, and reproduction. Mitosis occurs in somatic cells, whereas meiosis occurs only in germ-line cells in eukaryotic organisms.

#### 3.12.1 *Mitosis*

Somatic cells divide by **mitosis**. Mitosis produces two nearly identical daughter cells, each containing the same number of chromosomes as the original cell (Figure 3.3). Before the onset of mitosis, the chromosomes within the nuclear membrane are not visible by light microscopy, because they are not condensed. Cells not actively undergoing mitosis are in the **interphase** state.

The **cell cycle** consists of a coordinated set of processes by which a cell replicates all its components and is divided into two nearly identical daughter cells. The coordination of cell growth and periodic chromosome replication and division during the cell cycle has important implications for understanding the evolution of cells and development (Edgar and Lehner 1996, Novak et al. 1997, Dobie et al. 1999, Zachariae 1999, Zhang 1999).

The cell cycle consists of four phases:  $G_1 \rightarrow S \rightarrow G_2 \rightarrow M$ . DNA synthesis and chromosome duplication take place during the portion of the cell cycle called the **S phase** (for synthesis), but they do not occur during the  $G_1$  and  $G_2$  phases of the cell cycle. The **M phase** represents **mitosis**, in which the duplicated chromosomes and the cytoplasm are divided into two daughter cells.  $G_1$  is the gap between



**Figure 3.3** Mitosis of a diploid cell involves duplication of each homologous chromosome and its distribution to the daughter cells. DNA replication occurs during interphase when the DNA is not condensed. During prophase, the two daughter chromatids are attached to each other at the centromere. During metaphase, the highly condensed chromosomes line up on the metaphase plate, and during anaphase the daughter chromatids separate and begin moving to opposite poles. During telophase, the nuclear membranes reform, and two identical daughter cells with a complete complement of chromosomes have been produced.

mitosis and DNA synthesis (S), whereas G<sub>2</sub> is the gap between S and mitosis (M). The length of a cell cycle varies by cell type, but it typically lasts ≈18–24 hours, with mitosis requiring 0.5–2 hours (see Figure 1.14).

Mitosis is divided into four main stages: prophase, metaphase, anaphase, and telophase (Figure 3.3). In prophase, the nuclear envelope is still intact, and each

chromosome condenses to form two visible, thin threads (chromatids) within the nucleus. Because chromosome duplication occurred in the S phase, each chromosome consists of two chromatids connected at the **centromere**. The centromere is the attachment point for the spindle fibers that will draw each of the chromosomes into their respective nuclei later in mitosis. In late prophase, the nuclear membrane disappears and a mitotic spindle begins to form.

During prometaphase, the spindle develops. The spindle is a complex structure consisting of centrosomes (two centrioles oriented at right angles to one another) and microtubules (hollow protein cylinders consisting of tubulin). The two bundles of fibers extend between the opposite poles of the cell and attach to the centromere of each chromosome (Wolf 1995, Gonzalez et al. 1998). Then, the chromosomes move toward the center of the cell in a plane equidistant from the spindle poles. By the end of **metaphase**, the duplicated chromosomes are lined up on the metaphase plate and are at their most condensed stage, making it easy to examine them for differences in morphology.

During the next stage, **anaphase**, the centromeres divide and the two sister chromatids now have their own centromeres, so they have become independent chromosomes. These newly separated chromosomes move toward the opposite poles. At the end of anaphase, a complete set of chromosomes lies near each opposite pole.

During **telophase**, the chromosomes have reached the spindle poles, and the cleavage furrow within the cytoplasm has become visible. The nuclear membrane reforms around each group of chromosomes, the chromosomes decondense, cleavage progresses, and the spindle disappears. The mitochondria often align parallel to the spindle, which may guarantee that they are distributed to both daughter cells. A gradually deepening furrow divides the cytoplasm, and a new cell membrane forms. If all has gone well, the result should be the formation of two nearly identical cells with perfectly duplicated genetic information in the nucleus and in the mitochondria within the cytoplasm.

**Check points** occur during the cell cycle to ensure that genetic information is duplicated perfectly. During the checkpoints, the genetic material is monitored for integrity and status of replication before the cells commit either to replicate the DNA during S phase, or to segregate it during mitosis (Elledge 1996). If the cell cycle were not well regulated, the cell would subject to genetic instability or death.

Cyclins and cyclin-dependent protein kinases regulate the cell cycle (King et al. 1996, Stillman 1996, Piwnica-Worms 1999). The checkpoints involve

signal-transduction pathways whose effectors interact with the cyclin/cyclin-dependent protein kinases to block the cell cycle. Blocking the cell cycle allows time for repair of damage at G<sub>1</sub> (before DNA replication) or just before mitosis at the G<sub>2</sub> DNA-damage checkpoint.

Chromosome replication takes place during the S phase, and the duplicated chromosomes remain physically connected (as sister chromatids) until anaphase of mitosis. The cohesion of the sister chromatids is what permits chromosome segregation to take place long after chromosome duplication, and this cohesion is due to a multisubunit complex called cohesin. Cleavage of one of cohesin's subunits seems to trigger separation of the sister chromatids at the onset of anaphase ([Nasmyth et al. 2000](#)). The ability of eukaryotic cells to delay segregation of the replicated chromosomes until long after they have been duplicated distinguishes the eukaryotic cell cycle from that of bacteria, in which chromosome segregation starts immediately after DNA replication is initiated. The separation of chromosome duplication and segregation has played a central role in the evolution of eukaryotic organisms ([Nasmyth et al. 2000](#)). Mitotic chromosome condensation, without which large genomes cannot be partitioned between daughter cells at cell division, would not be possible if chromosome segregation coincided with DNA replication. The gap (G<sub>2</sub>) between S and M phases in the cell cycle thus makes possible the evolution of large genomes that can be transmitted safely to daughter cells.

### 3.12.2 Meiosis

Meiosis probably evolved from a mitosis-like process ([van Heemst and Heyting 2000](#)). Meiosis is responsible for two essential aspects of the sexual life cycle in eukaryotes: the transition from the diploid to the haploid state and the generation of new combinations of alleles. Meiosis occurs only in the germ line (ovaries or testes).

During **meiosis**, cells are produced that have a reduced number of chromosomes (the haploid or n number). This reduction means that when the germ cells (eggs and sperm) fuse, the diploid (2n) number of chromosomes is restored ([Figure 3.4](#)). If meiosis did not reduce the number of chromosomes to n, the number of chromosomes in a sexually reproducing organism would double each generation. Both divisions in meiosis have prophase, metaphase, anaphase, and telophase stages, but their details are different ([Figure 3.4](#)). Meiosis may require days or weeks (or longer) to complete. The essence of meiosis is that only one duplication of the chromosomes occurs, but two cell divisions occur, producing four haploid gametes from the original diploid cell. Meiosis requires

## MEIOSIS I



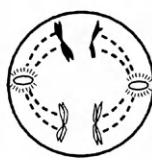
(A) Early prophase  
crossing over may  
occur between  
homologous  
chromosomes



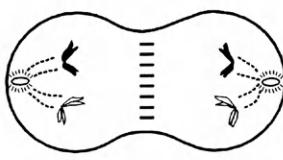
(B) Middle prophase



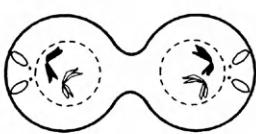
(C) Late prophase



(D) Metaphase



(E) Anaphase  
segregation of  
homologous  
chromosomes



(F) Telophase

## MEIOSIS II



(G) Interphase



(H) Prophase



(I) Metaphase



(J) Anaphase

segregation of  
daughter chromatids



(K) Telophase

two cell divisions (I and II) to produce daughter cells with the haploid set of chromosomes.

Meiosis I is the *reductional division*, in which the number of chromosomes is reduced from  $2n$  to  $n$ . Prophase of meiosis I can take a long time, and it has been divided into substages (Figure 3.4). During prophase I, the chromosomes condense and become visible. Homologous chromosomes pair and become closely associated along their length. Each homologous chromosome consists of two sister chromatids joined at the centromere; thus, the pairing of homologous chromosomes produces a four-stranded structure. During prophase I, the paired chromosomes are able to exchange genetic information by **crossing over**; this crossing over results in a shuffling of the genetic information in the gametes. The number of locations where genetic information was exchanged by crossing over often is indicated by the formation of **chiasmata**. These chiasmata are visible under the microscope during prophase I. Chiasmata result from the physical exchange of nucleotides between chromatids of the homologous chromosomes.

During metaphase I, the two homologous chromosomes are located on opposite sides of the metaphase plate (Figure 3.4). The orientation of each chromosome relative to the two poles is random and thus which member of each pair of chromosomes (one set was originally derived from the mother and the other set was originally derived from the father) will move to a particular pole is random. This **random alignment of chromosomes** on the metaphase plate is the basis of Mendel's **Law of Independent Assortment**. Thus, chromosomes originally derived from the individual's mother and father will end up assigned to daughter cells in a random manner.

During anaphase I, the homologous chromosomes separate from each other and move to opposite poles. This physical separation of homologous chromosomes during anaphase I is the physical basis of Mendel's **Law of Segregation**. After anaphase I, a haploid set of chromosomes consisting of one homolog from each pair is located near each pole of the spindle. During telophase I, the spindle breaks down (Figure 3.4). Chromosomes may pass directly from telophase I to prophase II of meiosis II. Alternatively, there may be a pause between the two meiotic divisions. Chromosome duplication does not occur between meiosis I and II, however.

Meiosis II is similar to a mitotic division, with each daughter cell from meiosis I being replicated, resulting in the production of four haploid cells (Figure 3.4). Meiosis II is different from mitosis, however, because the chromatids of a chromosome usually are not identical along their entire length. This is because crossing over could have occurred during prophase of meiosis I and resulted in an exchange of genetic information between the chromatids.

Meiosis has two unusual aspects: one of the most extraordinary aspects of meiosis I is that the two homologous chromosomes that are destined to pair and undergo recombination (crossing over) are able to find each other in a vast set of nonhomologous sequences. How this is achieved is a matter of considerable interest (Roeder 1997, Haber 1998). In *Drosophila*, pairing of homologous chromosomes may be facilitated through specialized pairing sites on the chromosomes. Heterochromatin, especially in the centromeres and telomeres, has been implicated as a mechanism that facilitates chromosomal pairing (Walker and Hawley 2000). During the pairing of homologous chromosomes, an elaborate ladder of protein called the synaptonemal complex is formed that helps to hold them together (Haber 1998).

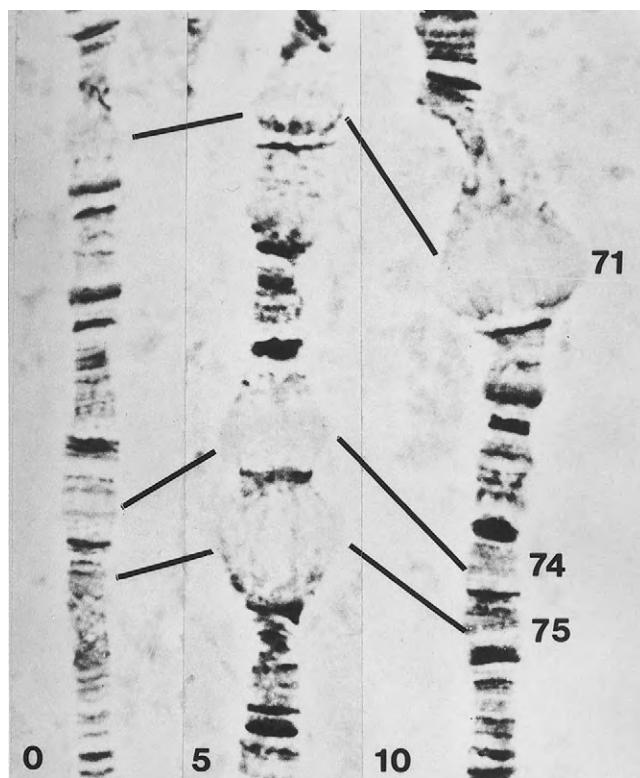
A second extraordinary aspect of meiosis is the pairing of sister chromatids until their disjunction. Protein complexes facilitate this cohesion (van Heemst and Heyting 2000). During pairing, recombination by crossing over occurs at a 100- to 1000-fold higher frequency in meiosis than in mitosis. Recombination tends to occur at certain chromosomal loci called “hotspots” and occurs more often between homologous chromosomes rather than between the sister chromatids.

Crossing over occurs about twice per paired set of chromosomes and serves two roles: the resulting recombination yields new combinations of alleles and plays a mechanical role in separation (disjunction) of the homologous chromosomes at meiosis I (van Heemst and Heyting 2000). Appropriate separation of homologous chromosomes in meiotic anaphase I requires that paired homologous chromosomes, rather than individual chromosomes, line up on the metaphase I spindle. At anaphase I, the homologous chromosomes move to opposite poles, resulting in meiosis I being the reductional division.

Each metaphase chromosome has a distinct morphology that is identifiable by staining with lactic-acetic orcein or other stains. The location of the centromere allows cytogeneticists to distinguish particular chromosomes. The arms of the chromosome take up stains in a banding pattern that is characteristic, especially in salivary gland chromosomes of dipterans; these chromosomes are polytene, i.e., they contain multiple copies of the DNA (Figure 3.5).

### 3.13 Chromosome Damage

Chromosome damage probably occurs continuously in all cells. Types of damage range from single-base changes that result from mistakes made by DNA polymerases during replication, to chromosome breakage. DNA and chromosomes are damaged by many factors, including the production of metabolic mutagens



**Figure 3.5** Polytene chromosomes from *Drosophila* salivary glands with puffing of different regions. Polyteny occurs when the DNA replicates but does not separate. Each “chromosome” may consist of hundreds of DNA molecules, making them easy to see under the microscope. Polyteny is one mechanism by which an insect can produce abundant mRNAs in a short time; each DNA strand can be transcribed at the same time.

within the cells. Water and oxygen can cause mutations, as well as ionizing radiation and UV light.

Cells have active repair processes to repair such damage. Repairs occur by direct reversal of damage and by excision of a damaged segment of DNA followed by its replacement. Insects have many genes involved in DNA repair, with some encoding products that recognize DNA damage, some that can excise the damaged region, and others that repair the damage.

Chromosome breaks can occur at any stage of the cell cycle and generally are repaired by rejoining the broken ends, so that the repaired chromosome seems intact. Unfortunately, not all chromosomal damage is repaired, and chromosomal breaks can lead to large-scale rearrangements of chromatin within chromosomes or to exchanges of chromatin between nonhomologous chromosomes.

If the rearranged chromosome lacks a centromere, it is **acentric**. A cell containing the acentric fragment will be unlikely to transmit this fragment to its daughter cells during meiosis or mitosis, resulting in loss of significant amounts of genetic information, which is usually lethal. Chromosomes that end up with two centromeres (**dicentric**) also are unstable, leading to breaks in the chromosomes if the centromeres are distributed to opposite poles during meiosis or mitosis. This instability results in breakage and loss of genetic information, which is often lethal.

In some cases, two different chromosomes are broken, and the repair results in chromosomes that contain fragments from two different chromosomes. These changes are called **translocations**. If the broken chromosomes are repaired but the broken piece is inserted in the wrong orientation, it is called an **inversion**. Sometimes, small portions of a chromosome are deleted (deletion), and sometimes a small portion of the chromosome receives a novel piece (insertion).

### 3.14 Polyteny

In a normal chromosome replication cycle (mitosis), chromosomes condense, replicate, separate, and segregate to daughter cells. In **polytene** cells,  $\geq 10$  DNA replication cycles may occur but the daughter chromosomes remain in an extended state and *do not separate*; such cells become larger and do not divide. The daughter DNA strands stay paired, with homologous regions aligned, and this gives rise to a banding pattern along the length of the chromosome under the light microscope even during interphase. In some cases, the maternal and paternal homologous chromosomes may synapse, resulting in an apparently haploid ( $n$ ) number of giant chromosomes.

Polyteny is particularly common in larval salivary glands of Diptera, especially in flies from the Drosophilidae, Chironomidae, Cecidomyiidae, and Sciaridae, but it also occurs in the midgut and fat body in these insects. Polyteny also occurs in Collembola. The number of rounds of DNA strand replication varies from tissue to tissue, with the largest number found in the salivary glands, where there may be as many as 1000–2000 chromatids per giant chromosome.

In *Drosophila* salivary gland chromosomes, the euchromatin regions contain genes, whereas the heterochromatic regions primarily contain repetitive DNA sequences including centromeres and telomeres (Leach et al. 2000). The banding patterns make it easy to identify specific sites on *Drosophila* salivary chromosomes (Figure 3.5). Approximately 5000 chromosome bands have been identified in *D. melanogaster*, providing a detailed cytological map. *Drosophila*

*melanogaster* has four pairs of chromosomes; chromosomes 2 and 3 are large with central centromeres, and chromosome 4 is the shortest. Females have two X chromosomes, whereas males have an X and a Y chromosome. The Y chromosome is largely heterochromatic, containing only a few genes.

Because polytene salivary gland chromosomes of *Drosophila* are large and have a well-defined morphology, specific genes can be localized by a procedure called *in situ* hybridization. Radiolabeled DNA or RNA probes can be added to salivary gland cells that have been squashed on glass slides. (A probe is a molecule labeled with radioactive isotopes, or another tag, that is used to identify or isolate a specific gene, gene product, or protein.) The labeled probes will anneal to the homologous DNA by base pairing after the chromosomal DNA is denatured (the DNA strands are separated). After washing off any excess probe, the position of the specific gene can be localized to a specific band or interband region of a specific chromosome by the presence of radioactive grains on an X-ray film. Genes, identified by a particular phenotype, also can be localized to specific sites in polytene salivary gland chromosomes if the mutation is associated with duplications, deletions, inversions, translocations, or other chromosomal abnormalities that can be detected by abnormal banding patterns under the light microscope.

Polytene chromosomes are thought to represent a special case of the more general phenomenon of endopolyploidy (White 1973). In **polyploidy**, an increase in chromosome *number* occurs within the nuclei of certain tissues without a breakdown in the nuclear membrane. Thus, chromosome duplication takes place and the chromosomes separate after replication (unlike the situation in polyteny), but no cell division occurs. Many insect cells have 4n, 8n, 16n, and so on, numbers of chromosome sets.

### 3.15 Chromosomal Puffing

At particular stages in development of many Diptera, some of the genes in salivary gland polytene chromosomes undergo swelling or **puffing**. This puffing is correlated with gene activity. Puffing involves an unraveling of the DNA in a region of the chromosome  $\approx$ 1 to 10 bands in length. The patterns of puffing differ in different instars in *D. melanogaster*, indicating different genes are active in different instars. Puffing is controlled by the hormone ecdysterone, heat shock, and other environmental conditions. The largest puffs contain genes coding for proteins that are produced in very large amounts in the salivary gland, such as the salivary gland secretions and silk. Puffs are associated with extensive transcription of DNA (Figure 3.5).

### 3.16 B Chromosomes

**B chromosomes** are a heterogeneous class of chromosomes found in the nucleus and also are called accessory or supernumerary chromosomes. B chromosomes are found in many insects, and they probably originate by several mechanisms, including being derived from autosomes and sex chromosomes in intra- and interspecies crosses (Camacho et al. 2000).

B chromosomes may only be present in some individuals from some populations in a species. B chromosomes have irregular mitotic and meiotic behavior, thereby allowing them to accumulate in the germ line, so that they are transmitted at rates higher than those of normal chromosomes.

Over evolutionary time, genes on B chromosomes may be silenced, undergo heterochromatinization, and accumulate repetitive DNA and transposons. B-chromosome frequencies in populations result from a balance between their transmission rates and their effects on host fitness. The long-term survival of B chromosomes depends on their ability to survive efforts by their host to eliminate or suppress them because they are often considered to be parasites. Because B chromosomes can interact with standard chromosomes, they could play a positive role in genome evolution if they contribute useful genetic information.

An example of a very interesting B chromosome is that of the paternal sex-ratio (PSR) chromosome of the parasitoid *Nasonia*. PSR increases in frequency because it is able to destroy paternal chromosomes (described in Chapter 10). The effects of other B chromosomes on their hosts are often unknown.

### 3.17 Sex Chromosomes

In eukaryotes with identifiable sexes, there generally is a pair of chromosomes called **sex chromosomes**. The sex chromosomes are often morphologically different from the rest of the chromosomes (**autosomes**). In most species, the male is the **heterogametic sex**, meaning the sex that has heteromorphic, only partially homologous sex chromosomes that are usually called X and Y. The X and Y chromosomes pair in the first prophase of spermatogenesis and, as a result of segregation, two types of gametes are produced: one type containing the X chromosome and one type containing the Y chromosome. Sperm containing the Y chromosome will result in male progeny, whereas sperm containing the X chromosome will produce female progeny. Some heterogametic species have males that are XO, lacking a Y chromosome.

Typically, the Y chromosome is smaller than the X chromosome and has very few of the genes that are on the X chromosome. The Y chromosome is often

composed primarily of heterochromatin. The X chromosome is usually more like an autosome in function and appearance. However, because it exists in one copy in the heterogametic sex (XY or XO), some form of dosage compensation is required to equalize the amount of gene product in the two sexes (for a discussion of dosage compensation, see Chapter 10).

The female is usually the **homogametic sex**, with two X chromosomes, and thus produces only eggs containing an X chromosome. In some insects, such as the Lepidoptera, females are the heterogametic sex. In this case, the sex chromosomes often are designated as W and Z, with the W chromosome analogous to the Y chromosome of the male (White 1973, Wagner et al. 1993).

### 3.18 Extranuclear Inheritance in Mitochondrial Genes

Genes located in the nucleus show Mendelian inheritance because they segregate in a regular manner during meiosis. However, not all genes in eukaryotic organisms are located in the nucleus. **Mitochondria** are inherited cytoplasmically, and they primarily are transmitted through the maternal gametes.

Mitochondria are self-replicating organelles that occur in the cytoplasm of all eukaryotes. Mitochondria are considered to be the descendants of an aerobic eubacterium that became an endosymbiont within an early anaerobic cell that may or may not have contained a nucleus (Kobayashi 1998). Mitochondria are thought to have originated once (Gray et al. 1999). Based on studies of DNA sequences, members of the rickettsial subdivision of the  $\alpha$ -Proteobacteria, a group of obligate intracellular parasites that include the *Rickettsia*, *Anaplasma*, and *Ehrlichia*, are the closest known relatives of mitochondria (Gray et al. 1999).

Mitochondria have evolved and the mitochondrial genomes of insects have departed radically from the ancestral pattern (Gray et al. 1999). Mitochondrial genomes range in size from 6kb to >2Mb (Sogin 1997). Genome size has been reduced after endosymbiosis because some mitochondrial genes became expendable in the internal environment of the host cell (Blanchard and Lynch 2000). Some nuclear genes have replaced the function of the mitochondrial (mt) genes, but much of the reduction in mitochondrial genome size occurred through the **transfer** of mitochondrial protein-coding genes into the nuclear genome. As a result, the mitochondrion has an incomplete set of genes for its own function (Blanchard and Lynch 2000).

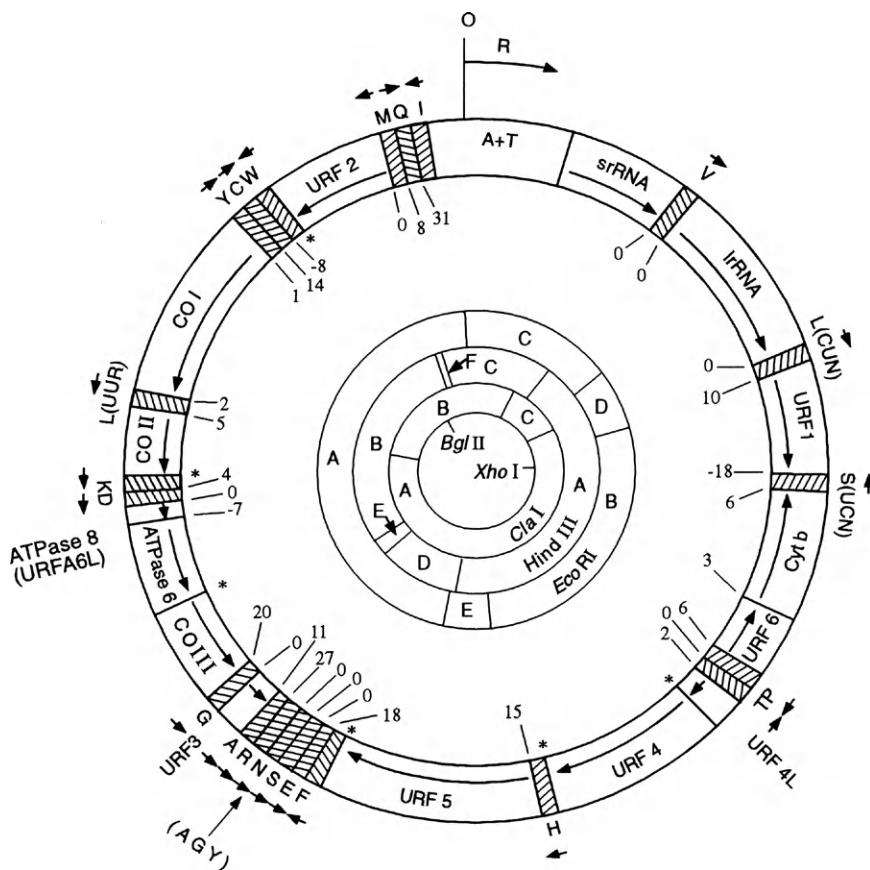
The movement of mitochondrial genes into the nuclear genome is an example of **horizontal gene transfer**. The result of such horizontal transfer means that genes originally located in the mitochondria, but now in the nucleus,

must be transcribed, translated, and acquire a sequence that targets the protein produced in the cytoplasm back into the mitochondrion (Ryan et al. 1997). Furthermore, the new nuclear gene must be properly regulated. On reaching the mitochondrion, the protein must be properly folded, modified, and assembled into a larger protein complex. Proteins destined for mitochondria are maintained in their proper form by **molecular chaperones**, proteins that bind to and assist in the folding of proteins into their functional states. Chaperones do not form part of the final protein structure nor do they contain information specifying a particular folding or assembly pathway. Examples of molecular chaperones include proteins produced by the *Hsp70* and *chaperonin* gene families (Ryan et al. 1997).

A double membrane surrounds each mitochondrion. The inner membrane is highly invaginated, with projections called cristae that are tubular or lamellar. These cristae are the sites of oxidative phosphorylation that result in the formation of **adenosine triphosphate** (ATP), the primary molecule for storing chemical energy in a cell (Saraste 1999). Mitochondrial DNA is a significant component of the total DNA in insect cells. Approximately half of the DNA in an unfertilized *D. melanogaster* egg is mtDNA.

The coexistence of more than one type of mtDNA within a cell or individual (**heteroplasmy**) is thought to be rare in natural populations. Paternal mtDNA either is not transmitted at fertilization, or it contributes only a small fraction of the mtDNA in the developing embryo, and the paternally derived mitochondria typically are lost during development. Because insect mitochondria are transmitted from mother to progeny, they are inherited asexually (Birky 1995). Until recently, it was assumed that inheritance of mitochondria was a passive process, a consequence of their random diffusion throughout the cytoplasm. Now, it is known that mitochondria are associated with the cytoskeleton and move in coordinated ways during cell division and differentiation (Yaffe 1999).

Mitochondria contain distinctive ribosomes, tRNAs, and aminoacyl-tRNA synthetases (Gray 1989, Sogin 1997, Kobayashi 1998). Mitochondria have their own genetic code that differs slightly from the universal genetic code in the nucleus. The mitochondrion of *Drosophila yakuba* codes for 37 genes: two genes are rRNA, 22 are tRNA, and 13 are protein genes that code for subunits of enzymes functioning in electron transport or ATP synthesis (Clary and Wolstenholm 1985, Figure 3.6). The organization and evolution of insect mitochondrial genomes is being derived from analysis of the complete sequences of mitochondria isolated from an increasing number of species. Sequences of mitochondria have been obtained from many insects and are deposited in GenBank and other databases.



**Figure 3.6** Diagram of the circular DNA molecule from the mitochondria of *Drosophila yakuba*. The outside circle shows the open reading frames (URF1–URF6 and URF 4L) that code for subunits of the respiratory chain NADH dehydrogenase and of the genes coding for cytochrome *b*; cytochrome *c* oxidase subunits I, II, and III; and ATPase subunits 5 and 6. The origin of and direction of replication are indicated by O and R. The variable A + T region is shaded. The arrows indicate the direction of gene transcription. The tRNA genes are crosshatched and indicated by their single-letter amino-acid codes. lrRNA and srRNA are the large and small rRNA genes, respectively. The numbers on the inside of the outer circle are the numbers of apparently noncoding nucleotides that occur between the genes. The innermost circles indicate restriction fragments produced with the enzymes indicated (from Clary and Wolstenholme 1985).

More detailed investigation of insect mitochondrial genomes may provide some contradictions to the above-mentioned generalizations. For example, mitochondrial genomes >20 kb have been found in three species of curculionid beetles (*Pissodes strobi*, *Pissodes nemorensis*, and *Pissodes terminalis*) (Boyce et al. 1989). The large size (30–36 kb) in these three *Pissodes* species is due to

an enlarged A + T-enriched region (9–13 kb) and a series of 0.8–2.0 kb tandemly repeated sequences adjacent to the A + T region. Every weevil sampled in all three species had two to five distinct size classes of mtDNA (exhibited heteroplasmy). The magnitude of the size differences, the number of size classes found within individual weevils, and the abundant mtDNA heteroplasmy is unusual (Boyce et al. 1989).

The dogma that mtDNA is exclusively inherited in a maternal manner has been questioned in *Drosophila* and marine mussels. Incomplete maternal inheritance of mtDNA occurs in *Drosophila simulans* (Satta et al. 1988, Matsuura et al. 1991), and the high level of heteroplasmy found in the three *Pissodes* species could be due to paternal transmission of mtDNA.

Mitochondrial chromosomes are circular, supercoiled, double-stranded DNA molecules. The mitochondrial chromosome of *D. yakuba* contains  $\approx$ 18.5 kb of DNA, and each mitochondrion contains multiple copies of the chromosome (Figure 3.6). Mitochondrial genes in insects lack introns and intergenic regions usually are small or absent. The ribosomes found in the mitochondria are smaller than the ribosomes in the cytoplasm.

Most eggs and somatic cells contain hundreds or thousands of mtDNA molecules, so a new mutation can result in a situation in which two or more mtDNA genotypes coexist within an individual (heteroplasmy). Heteroplasmy, however, is apparently a transitory state in germ cells. Thus, the majority of individuals are effectively haploid with regard to the number of types of mtDNA transmitted to the next generation.

Mitochondrial DNA evolves faster than single-copy nuclear DNA because mitochondria are relatively inefficient in repairing errors during DNA replication or after DNA damage. For example, in Hawaiian *Drosophila*, mtDNA seems to evolve  $\approx$ 3 times faster than the genes of nuclear DNA (Moritz et al. 1987). Because mtDNA does not code for proteins involved directly in its own replication, transcription, or translation, mtDNA has a large number of length mutations and transitions.

Mitochondrial DNA can be amplified easily from mitochondria by the polymerase chain reaction (PCR) (see Chapter 8 for a description of the PCR) because there are multiple copies in each cell. Mitochondrial DNA is easier to purify from cells than a specific segment of nuclear DNA. This is due to its buoyant density, high copy number within cells, and its location within an organelle, making mtDNA a useful subject for systematics or population genetics studies, as is described in Chapters 12 and 13.

### 3.19 Transposable Elements are Ubiquitous Agents that Alter Genomes

Every insect genome probably contains multiple types of **transposable elements** (TEs) (Berg and Howe 1989, Craig et al. 2001). An organism may contain active and inactive TEs, and the inactive TEs are subject to mutation and eventual loss from the genome. During the process, the sequence of the TE may become unrecognizable as a TE and be considered junk DNA.

Transposable elements are genetic elements that can move from one chromosomal site to another, and they are usually present in multiple copies within a genome. The ubiquity of TEs in a diverse array of organisms has raised several interesting questions about their evolutionary impact. New TEs are still invading and spreading within insect populations, and the role of TEs in insect evolution and genetic manipulation are discussed further in Chapters 4, 7, 9, and 14.

The diversity of arthropods and their genetic systems has only been hinted at. In Chapter 4, we will explore additional details of genome organization, developmental processes, and diversity in insects.

### References Cited

- Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D., and Amanatides, P.G., et al. (2000). The genome sequence of *Drosophila melanogaster*. *Science* **287**: 2185–2195.
- Ault, J.G., (1996). Contributions of insect systems to the chromosome theory of inheritance and to our understanding of chromosome behavior on the spindle. *Int. J. Insect Morphol. Embryol.* **25**: 93–114.
- Benbow, R.M., Zhao, J., and Larson, D.D., (1992). On the nature of origins of DNA replication in eukaryotes. *BioEssays* **14**: 661–669.
- Berg, D.E., Howe, M.M., Eds., (1989). *Mobile DNA*, ASM Press, Washington, DC.
- Berry, S.J., (1985). Insect nucleic acids In: Kerkut, G.A. Gilbert, L.I., Eds., *Insect Physiology Biochemistry and Pharmacology*, Vol. 10 Pergamon, Oxford, UK, pp. 219–253.
- Besansky, N.J., and Collins, F.H., (1992). The mosquito genome: organization, evolution and manipulation. *Parasitol. Today* **8**: 186–192.
- Biemont, C., (2009). Are transposable elements simply silenced or are they under house arrest? *Trends Genet.* **25**: 333–334.
- Bigot, Y., Hamelin, M., and Periquet, G., (1990). Heterochromatin condensation and evolution of unique satellite-DNA families in two parasitic wasp species: *Diadromus pulchellus* and *Eupelmus vuilleti* (Hymenoptera). *Mol. Biol. Evol.* **7**: 351–364.
- Birky Jr., C.W., (1995). Uniparental inheritance of mitochondrial and chloroplast genes: mechanisms and evolution. *Proc. Natl. Acad. Sci. USA* **92**: 11331–11338.
- Black, W.C., and Rai, K.S., (1988). Genome evolution in mosquitoes: intraspecific and interspecific variation in repetitive DNA amounts and organization. *Genet. Res.* **51**: 185–196.
- Blackburn, E.H., (1991). Structure and function of telomeres. *Nature* **350**: 569–573.
- Blanchard, J.L., and Lynch, M., (2000). Organellar genes: why do they end up in the nucleus? *Trends Genet.* **16**: 315–320.
- Boyce, T.M., Zwick, M.E., and Aquadro, C.F., (1989). Mitochondrial DNA in the bark weevils: size, structure and heteroplasmy. *Genetics* **123**: 825–836.

- Brown, S.J., Henry, J.K., Black, W.C., and Denell, R.E., (1990). Molecular genetic manipulation of the red flour beetle: genome organization and the cloning of a ribosomal protein gene. *Insect Biochem.* **20**: 185–193.
- Burkholder, G.D., Latimer, L.J.P., and Lee, J.S., (1991). Immunofluorescent localization of triplex DNA in polytene chromosomes of *Chironomus* and *Drosophila*. *Chromosoma* **101**: 11–18.
- Camacho, J.P.M., Sharbel, T.F., and Beukeboom, L.W., (2000). B-chromosome evolution. *Philos. Trans. R. Soc. Lond. B.* **355**: 163–178.
- Celniker, S.E., (2000). The *Drosophila* genome. *Curr. Opin. Gen. Dev.* **10**: 612–616.
- Clary, D.O., and Wolstenholm, D.R., (1985). The mitochondrial DNA molecule of *Drosophila yakuba*: nucleotide sequence, gene organization, and genetic code. *J. Mol. Evol.* **22**: 252–271.
- Cooper, G.M., (2000). *The Cell: A Molecular Approach*. ASM Press, Washington, DC.
- Craig, N.L., Craigie, R., Gellert, M., and Lambowitz, A.M., (2001). *Mobile DNA II*. ASM Press, Washington, DC.
- Crain, W.R., Davidson, E.H., and Briteen, R.J., (1976). Contrasting patterns of DNA sequence arrangement in *Apis mellifera* (honeybee) and *Musca domestica* (housefly). *Chromosoma* **59**: 1–12.
- de Lange, T., (2001). Telomere capping—one strand fits all. *Science* **292**: 1075–1076.
- Dobie, K.W., Hari, K.L., Maggert, K.A., and Karpen, G.H., (1999). Centromere proteins and chromosome inheritance: a complex affair. *Curr. Opin. Genet. Dev.* **9**: 206–217.
- Edgar, B.A., and Lehner, C.F., (1996). Developmental control of cell cycle regulators: a fly's perspective. *Science* **274**: 1646–1652.
- Eickbush, T.H., (1997). Telomerase and retrotransposons: which came first? *Science* **277**: 911–912.
- Eissenberg, J.C., and Hilliker, A.J., (2000). Versatility of conviction: heterochromatin as both a repressor and an activator of transcription. *Genetica* **1009**: 19–24.
- Elledge, S.J., (1996). Cell cycle checkpoints: preventing an identity crisis. *Science* **274**: 1664–1672.
- Gatti, M., and Pimpinelli, S., (1992). Functional elements in *Drosophila melanogaster* heterochromatin. *Annu. Rev. Genet.* **26**: 239–275.
- Gonzalez, C., Tavasanis, G., and Mollinari, C., (1998). Centrosomes and microtubule organization during *Drosophila* development. *J. Cell Sci.* **111**: 2697–2706.
- Gray, M.W., (1989). Origin and evolution of mitochondrial DNA. *Annu. Rev. Cell Biol.* **5**: 25–50.
- Gray, M.W., Burger, G., and Lang, B.F., (1999). Mitochondrial evolution. *Science* **283**: 1476–1481.
- Greider, C.W., (1999). Telomeres do D-loop-T-loop. *Cell* **97**: 419–422.
- Haber, J.E., (1998). Searching for a partner. *Science* **279**: 823.
- Henikoff, S., (2000). Heterochromatin function in complex genomes. *Biochim. Biophys. Acta* **1470**: 1–8.
- Henikoff, S., Ahmed, K., and Malik, H., (2001). The centromere paradox: stable inheritance with rapidly evolving DNA. *Science* **293**: 1098–1102.
- Hennig, W., (1999). Heterochromatin. *Chromosoma* **108**: 1–9.
- Jenuwein, T., and Allis, C.D., (2001). Translating the histone code. *Science* **293**: 1074–1080.
- Juan, C., and Petitpierre, E., (1991). Evolution of genome size in darkling beetles (Tenebrionidae, Coleoptera). *Genome* **34**: 169–173.
- Kalhor, R., Tjong, H., Jayathilaka, N., Alber, F., and Chen, L., (2012). Genome architectures revealed by tethered chromosome conformation capture and population-based modeling. *Nat. Biotechnol.* **30**: 90–100.
- Kamnert, I., Lopez, C.C., Rosen, M., and Edstrom, J.E., (1997). Telomeres terminating with long complex tandem repeats. *Hereditas* **127**: 175–180.
- King, R.W., Deshaies, R.J., Peters, J.M., and Kirschner, M.W., (1996). How proteolysis drives the cell cycle. *Science* **274**: 1652–1658.
- Kobayashi, I., (1998). Selfishness and death: raison d'être of restriction, recombination and mitochondria. *Trends Genet.* **14**: 368–373.

- Koshland, D., and Strunnikov, A., (1996). Mitotic chromosome condensation. *Annu. Rev. Cell Dev. Biol.* **12**: 305–333.
- Kumar, A., and Rai, K.S., (1990). Intraspecific variation in nuclear DNA content among world populations of a mosquito, *Aedes albopictus* (Skuse). *Theor. Appl. Genet.* **79**: 748–752.
- Leach, T.J., Chotkowski, H.L., Wotring, M.G., Dilwith, R.L., and Glaser, R.L., (2000). Replication of heterochromatin and structure of polytene chromosomes. *Mol. Cell. Biol.* **20**: 6308–6316.
- Ma, R.Z., Black, W.C., and Reese, J.C., (1992). Genome size and organization in an aphid (*Schizaphis graminum*). *J. Insect Physiol.* **38**: 161–165.
- Manuelidis, L., (1990). A view of interphase chromosomes. *Science* **250**: 1533–1540.
- Manuelidis, L., and Cher, T.L., (1990). A unified model of eukaryotic chromosomes. *Cytometry* **11**: 8–25.
- Mason, J.M., and Biessmann, H., (1995). The unusual telomeres of *Drosophila*. *Trends Genet.* **11**: 58–62.
- Matsuura, E.T., Fukuda, H., and Chigusa, S.I., (1991). Mitochondrial DNA heteroplasmy maintained in natural populations of *Drosophila simulans* in Reunion. *Genet. Res.* **57**: 123–126.
- Misteli, T., (2012). Parallel genome universes. *Nat. Biotechnol.* **30**: 55.
- Moritz, C., Dowling, T.E., and Brown, W.M., (1987). Evolution of animal mitochondrial DNA: relevance for population biology and systematics. *Annu. Rev. Ecol. Syst.* **18**: 269–292.
- Murphy, T.D., and Karpen, G.H., (1995). Localization of centromere function in a *Drosophila* minichromosome. *Cell* **82**: 599–609.
- Nasmyth, K., Peters, J.M., and Uhlmann, F., (2000). Splitting the chromosome: cutting the ties that bind sister chromatids. *Science* **288**: 1379–1384.
- Nicklas, R.B., (1997). How cells get the right chromosomes. *Science* **275**: 632–637.
- Novak, B., Csikasz-Nagy, A., Gyorffy, B., Nasmyth, K., and Tyson, J.J., (1997). Model scenarios for evolution of the eukaryotic cell cycle. *Philos. Trans. R. Soc. Lond. B* **353**: 2063–2076.
- Okazaki, S., Tsuchida, K., Maekawa, H., Ishikawa, H., and Fujiwara, H., (1993). Identification of a pentanucleotide telomeric sequence, (TTAGG)<sub>n</sub>, in the silkworm *Bombyx mori* and in other insects. *Mol. Cell. Biol.* **13**: 1424–1432.
- Otto, S.P., and Whitton, J., (2000). Polyploid incidence and evolution. *Annu. Rev. Genet.* **34**: 401–437.
- Pardue, M.L., and DeBaryshe, P.G., (1999). Telomeres and telomerase: more than the end of the line. *Chromosoma* **108**: 73–82.
- Pardue, M.L., and DeBaryshe, P.G., (2011). Retrotransposons that maintain chromosome ends. *Proc. Natl. Acad. Sci. USA* **108**: 20317–20324.
- Perkins, H.D., Bedo, D.G., and Howells, A.J., (1992). Characterization and chromosomal distribution of a tandemly repeated DNA sequence from the Australian sheep blowfly, *Lucilia cuprina*. *Chromosoma* **101**: 358–364.
- Petrov, D.A., Lozovskaya, E.R., and Hartl, D.L., (1996). High intrinsic rate of DNA loss in *Drosophila*. *Nature* **384**: 346–349.
- Petrov, D.A., Sangster, T.A., Johnston, J.S., Hartl, D.L., and Shaw, K.L., (2000). Evidence for DNA loss as a determinant of genome size. *Science* **287**: 1060–1062.
- Piwnica-Worms, H., (1999). Fools rush. *Nature* **401**: 535–537.
- Plohl, M., Borstnik, B., Lucijanić-Justić, V., and Ugarković, D., (1992). Evidence for random distribution of sequence variants in *Tenebrio molitor* satellite DNA. *Genet. Res.* **60**: 7–13.
- Redi, C.A., Garagna, S., Zacharias, H., Zuccotti, M., and Capanna, E., (2001). The other chromatin. *Chromosoma* **110**: 136–147.
- Roeder, G.S., (1997). Meiotic chromosomes: it takes two to tango. *Genes Dev.* **11**: 2600–2621.
- Rosen, M., and Edstrom, J.E., (2000). DNA structures common for chironomid telomeres terminating with complex repeats. *Insect Mol. Biol.* **9**: 341–347.
- Ryan, M.T., Naylor, D.J., Hoj, P.B., Clark, M.S., and Hoogenraad, N.J., (1997). The role of molecular chaperones in mitochondrial protein import and folding. *Int. Rev. Cytol.* **174**: 127–193.

- Sahara, K., Marec, F., and Traut, W., (1999). TTAGG telomeric repeats in chromosomes of some insects and other arthropods. *Chromosome Res.* **7**: 449–460.
- Samols, D., and Swift, H., (1979). Genomic organization in the flesh fly *Sarcophaga bullata*. *Chromosoma* **75**: 129–153.
- Saraste, M., (1999). Oxidative phosphorylation as the fin de siecle. *Science* **283**: 1488–1493.
- Sasaki, T., and Fujiwara, H., (2000). Detection and distribution patterns of telomerase activity in insects. *Eur. J. Biochem.* **267**: 3025–3031.
- Satta, Y., Toyohara, N., Ohtaka, C., Tasuno, Y., Watanabe, T., and Matsuura, E.T., et al. (1988). Dubious maternal inheritance of mitochondrial DNA in *D. simulans* and evolution of *D. mauritiana*. *Genet. Res.* **52**: 1–6.
- Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., and Hoichman, M., et al. (2012). Three-dimensional folding and functional organization principles of the *Drosophila* genome. *Cell* **148**: 458–472.
- Sogin, M.L., (1997). History assignment: when was the mitochondrion founded? *Curr. Opin. Genet. Dev.* **7**: 792–799.
- Stillman, B., (1996). Cell cycle control of DNA replication. *Science* **274**: 1659–1664.
- Sun, X., Wahlstrom, J., and Karpen, G., (1997). Molecular structure of a functional *Drosophila* centromere. *Cell* **91**: 1007–1019.
- Tyler-Smith, C., and Floridia, G., (2000). Many paths to the top of the mountain: diverse evolutionary solutions to centromere structure. *Cell* **102**: 5–8.
- Ugarkovic, D., Plohl, M., Lucijanic-Justic, V., and Borstnik, B., (1992). Detection of satellite DNA in *Palorus ratzeburgii*: analysis of curvature profiles and comparison with *Tenebrio molitor* satellite DNA. *Biochimie* **74**: 1075–1082.
- van Heemst, D., and Heyting, C., (2000). Sister chromatid cohesion and recombination during meiosis. *Chromosoma* **109**: 10–26.
- Wagner, R.P., Maguire, M.P., and Stallings, R.L., (1993). *Chromosomes A Synthesis*. Wiley-Liss, New York.
- Walker, M.Y., and Hawley, R.S., (2000). Hanging on to your homolog: the roles of pairing, synapsis and recombination in the maintenance of homolog adhesion. *Chromosoma* **109**: 3–9.
- Wallrath, L.L., (1998). Unfolding the mysteries of heterochromatin. *Curr. Opin. Genet. Dev.* **8**: 147–153.
- Weiler, K.S., and Wakimoto, B.T., (1995). Heterochromatin and gene expression in *Drosophila*. *Annu. Rev. Genet.* **29**: 577–605.
- Wells, R., Royer, H., and Hollenberger, C.P., (1976). Non *Xenopus*-like DNA organization in the *Chironomus tentans* genome. *Mol. Gen. Genet.* **147**: 45–51.
- White, M.J.D., (1973). *Animal Cytology and Evolution*, 3rd ed. Cambridge Univ. Press, Cambridge, MA.
- Wolf, K.W., (1995). Spindle membranes and spindle architecture in invertebrates. *Micron* **26**: 69–98.
- Yaffe, M.P., (1999). The machinery of mitochondrial inheritance and behavior. *Science* **293**: 1493–1497.
- Zachariae, W., (1999). Progression into and out of mitosis. *Curr. Opin. Cell Biol.* **11**: 708–716.
- Zakian, V.A., (1989). Structure and function of telomeres. *Annu. Rev. Genet.* **23**: 579–604.
- Zhang, P., (1999). The cell cycle and development: redundant roles of cell cycle regulators. *Curr. Opin. Cell Biol.* **11**: 655–662.

# Genetic Systems, Genome Evolution, and Genetic Control of Embryonic Development in Insects

## Chapter Outline

- 4.1 Overview 104
- 4.2 Introduction 105
- 4.3 Genetic Systems in Insects 105
- 4.4 Endopolyploidy is Common in Somatic Tissues of Arthropods 107
- 4.5 Genetics of Insects Other than *D. melanogaster* 107
- 4.6 Dynamic Insect Genomes 108
  - 4.6.1 Horizontal Gene Transfer from Microorganisms to Insects 109
- 4.7 B Chromosomes 110
- 4.8 Unique-Sequence DNA in the Nucleus 111
- 4.9 Middle-Repetitive DNA in the Nucleus 111
  - 4.9.1 Heat-Shock Genes 112
  - 4.9.2 Histone Genes 113
  - 4.9.3 Immune-Response Genes 113
  - 4.9.4 Ribosomal Genes 115
  - 4.9.5 Silk Genes 115
  - 4.9.6 Transfer RNA Genes 116
  - 4.9.7 Vitellogenin Genes 116
  - 4.9.8 Transposable Elements 117
- 4.10 Highly Repetitive DNA 122
- 4.11 Producing Large Amounts of Protein in a Short Time: Gene Amplification and Gene Duplication 123
  - 4.11.1 Chorion Genes in *Drosophila* and Moths 124
  - 4.11.2 Insecticide Resistance 127
- 4.12 Multiple Genomes in or on Insects: What is the “Biological Individual”? 127
  - 4.12.1 Multiple Symbionts 128
  - 4.12.2 *Wolbachia* and *Cardinium* 132
  - 4.12.3 Polydnaviruses in Parasitoids 139
  - 4.12.4 Gut Symbionts in Arthropods 141
    - 4.12.4.1 *Termite and Cockroach Symbionts* 141
    - 4.12.4.2 *Rhagoletis Symbionts* 142
  - 4.12.5 Symbiosis in Fungus-Growing Attine Ants 143
  - 4.12.6 Southern Pine Beetle Symbionts 146

4.12.7	Aphid Symbionts	146
4.12.8	Tsetse Fly Symbionts	148
<b>4.13</b>	<b>Insect Development</b>	<b>149</b>
4.13.1	Oocyte Formation in <i>D. melanogaster</i>	149
4.13.2	Embryogenesis in <i>D. melanogaster</i>	151
4.13.3	Postembryonic Development	153
<b>4.14</b>	<b>Dissecting Development with <i>D. melanogaster</i> Mutants</b>	<b>154</b>
4.14.1	Maternal-Effect Genes	155
4.14.2	Zygotic-Segmentation Genes	157
4.14.2.1	<i>Gap Genes</i>	158
4.14.2.2	<i>Pair-Rule Genes</i>	159
4.14.2.3	<i>Segment-Polarity Genes</i>	159
4.14.2.4	<i>Homeotic Genes</i>	159
4.14.3	Insect Metamorphosis	161
<b>4.15</b>	<b>Interactions During Development</b>	<b>162</b>
<b>4.16</b>	<b>Similarities and Differences in Development in Other Insects</b>	<b>162</b>
<b>4.17</b>	<b>Evo-Devo and the Revolution in Developmental Studies</b>	<b>164</b>
<b>References Cited</b>		<b>166</b>

## 4.1 Overview

Molecular genetics has revolutionized our understanding of insect gene structure, organization, regulation, and development. One surprise has been the discovery that genomes are very dynamic over evolutionary time. For example, large portions of the insect genome may consist of multiple families of transposable elements (TEs) that can alter gene structure and function, chromosomal organization, and transfer horizontally between species. The discovery that gene amplification can be involved in resistance to insecticides in aphids and mosquitoes has opened new avenues for understanding this evolutionary and economic problem. Research also suggests that the genetic information contained in microbial symbionts such as *Wolbachia* may play an essential role in speciation and evolution of some insects. Other symbionts provide essential nutrients or other physiological services. Many insects contain three or four (or more) genomes: nuclear, mitochondrial, one or more gut symbionts, and *Wolbachia*, raising questions about the definition of the “biological individual.” Perhaps we should consider insects to be a biome that contains multiple genomes (species) living together. Our knowledge of the diversity of organisms associated with arthropods as symbionts and the role these play in the biology of their arthropod hosts is expanding, in large part due to our ability to use molecular tools, including whole-genome sequencing of symbiont and host.

Insect nuclear genomes can be diverse. Although many species have diploid males and females, some have haploid males and diploid females (arrhenotoky)

and some have only females (thelytoky). In some insect species, diploid males may undergo chromosome heterochromatinization and loss during development to become haploid (parahaploidy). This diversity of genome organization in insects reflects their long evolutionary history.

An understanding of the stages of embryonic development and many of the major genes that influence these stages in *Drosophila melanogaster* is emerging. A field of study called “evo-devo” compares development among organisms and is based on molecular analyses of development. Embryonic development in *D. melanogaster* is determined first by maternal genes and then by zygotic genes. Maternal-effect genes determine the polarity of the embryo. Next, segmentation genes influence the development of major bands or parasegments. Another determination is accomplished by the interaction of homeotic genes that provide a finer definition of segmental structures.

## 4.2 Introduction

Insects are both extremely numerous and diverse. Insects comprise approximately half of all described eukaryotic species and approximately  $\frac{3}{4}$  of all described animals. There are  $\approx 883,475$  described insect species in  $> 762$  families organized in 32 orders (Daly et al. 1998, Whitfield and Purcell 2013). Many insect species may remain undescribed because some estimates suggest there are as many as 2 million insect species. Insects have a long evolutionary history, live in a great variety of habitats, exhibit diverse types of life styles, have an extraordinary range of structural variations, eat an astonishing variety of food, and are among the most abundant animals on earth. The long evolutionary history of insects has provided sufficient time for them to develop a diversity of genetic systems (for an overview of insect evolution, see Chapter 12).

This chapter provides only a small sample of the diversity of insect genetic systems. More extensive reviews of insect cytogenetics can be found in White (1973) and Wagner et al. (1993). This chapter also provides an overview of the diversity of microbial symbionts associated with insects, the diversity of their transposable elements, gene regulation in some insect-specific genes, as well as a brief introduction to the molecular genetics of insect embryonic development, all topics that are important in understanding the mechanism of transforming insects by injection of cloned genes into embryos (see Chapter 9).

## 4.3 Genetic Systems in Insects

Most insects are diploid ( $2n$ ) in their somatic cells and haploid ( $n$ ) in their gametes. Other systems can be found; some insect groups are parthenogenetic and

may be polyploid, including species in the Orthoptera (Blaberidae, Tettigoniidae), Hemiptera (Coccidae, Delphacidae), Embioptera (Oligotomidae), Lepidoptera (Psychidae), Diptera (Chamaemyiidae, Chironomidae, Psychodidae, Simuliidae), Coleoptera (Ptinidae, Chrysomelidae, Curculionidae), and Hymenoptera (Diprionidae, Apidae) (Otto and Whitton 2000). Polyploid insects usually are 3n or 4n, but exceptions include curculionid weevil species that are 5n and 6n (Retnakaran and Percy 1985). Parthenogenesis has not been found in the Diplura, Protura, Odonata, Plecoptera, Dermaptera, Grylloblattodea, Zoraptera, Megaloptera, Mecoptera, and Siphonaptera, although only a few species in these groups have been examined carefully.

Parthenogenesis can be divided into three major types: arrhenotoky, thelytoky, and deuterotoky. **Deuterotoky** is the least common and involves the development of unfertilized eggs into either males or females, and at least one insect, a mayfly, is reported to exhibit facultative deuterotoky (White 1973). In **arrhenotoky**, insects are haplodiploid, with males developing from unfertilized haploid eggs and females developing from fertilized diploid eggs. Typically, arrhenotokous females determine the sex ratio, which is often biased toward females. The entire order Hymenoptera and many species in the Hemiptera, Thysanoptera, and Coleoptera are arrhenotokous (Hartl and Brown 1970, White 1973).

Thelytokous insect species have females only. **Thelytoky** has arisen repeatedly in evolution, consists of several types, and can be induced experimentally in several ways (White 1973; for examples, see Chapter 10). In some cases of thelytoky, eggs only develop after penetration by a sperm (pseudogamy or gynogenesis), but the sperm nucleus degenerates without fusing with the egg nucleus so that it makes no genetic contribution to the embryo. The sperm may be derived from the testis or ovotestis of a hermaphrodite or from a male of a different, but closely related, species. Thelytoky may be the sole mode of reproduction in a species, or it may alternate with sexual reproduction in regular manner (cyclical thelytoky), as happens in some aphids (Hales et al. 1997), gall wasps, and some cecidomyiids. In species that reproduce by cyclical thelytoky, genetic recombination is possible, but in species with complete thelytoky (only females) there is no way in which mutations that have occurred in two unrelated individuals can be combined in a third individual.

Thelytokous reproduction can be induced in the eggs of many species by pricking the egg, exposing it to chemical agents, or heat. In many normally bisexual insects, a few eggs deposited by virgin females can hatch spontaneously, and the incidence of such egg hatch can be increased by artificial selection. White (1973) suggests that the capacity for artificial parthenogenesis,

induced thelytoky, or facultative thelytoky indicates that some capacity for parthenogenesis is probably present in all eggs. Thelytokous species or thelytokous populations of bisexual species have been found in the Diptera, Hymenoptera, Lepidoptera, Orthoptera, and Coleoptera.

In the Hemiptera, both arrhenotoky and thelytoky occur (Retnakaran and Percy 1985), but even more complex genetic systems can be found (White 1973, Haig 1993). For example, in some mealybugs (Pseudococcidae), both males and females develop from fertilized eggs but, in the embryos that develop into males, the paternally derived chromosomes become heterochromatic, genetically inactive, and are not transmitted to the male progeny. This genetic system has been called **parahaploidy** (Brown and Nur 1964, Nur 1990). **Chromosome imprinting** is probably involved to ensure that the paternally derived chromosomes are eliminated and not the maternally derived chromosomes. The mechanisms involved in chromosomal imprinting could be associated with methylation of DNA (Sapienza et al. 1987, Solter 1988, Wagner et al. 1993).

#### 4.4 Endopolyploidy is Common in Somatic Tissues of Arthropods

The discussion of ploidy is confusing because, in most insects, some of the somatic tissues exhibit high levels of endopolyploidy. For example, haploid male honey bees have about the same amount of DNA as females in their somatic tissues because nuclei of the male undergo compensatory endomitosis so that equal amounts of DNA are present. In some cases, haploid males are known to exhibit higher levels of endopolyploidy in some tissues than the diploid females of the same species.

#### 4.5 Genetics of Insects Other than *D. melanogaster*

Much of what we know about the genetics of insects is derived from the study of *Drosophila* species (Ashburner 1989, Brody 1999). Extensive genetic information is available for *D. melanogaster*, including the complete sequences of the genome (Adams et al. 2000, Hawley and Walker 2000, Jabbari and Bernardi 2000, Otto 2000; for additional details on genome projects, see Chapter 7). For updated information on the *Drosophila* genome and other aspects of *Drosophila* biology, search the web at The Interactive Fly (<http://www.sdbonline.org/fly/aimain/1aahome.htm>) and FlyBase, A Database of the *Drosophila* Genome.

Relatively little genetic information has been available for the majority of the 883,475 known insect species. Genetic studies previously were made on insects of economic importance, such as mosquitoes, honey bees, the Mediterranean fruit fly, the parasitoid wasp *Nasonia*, the silk moth, *Tribolium* beetles, tsetse,

and the screwworm (Wright and Pal 1967; King 1975; Robinson 1971; Sokoloff 1966, 1977; Steiner et al. 1982; Heckel 1993; Goldsmith et al. 2005; Gooding and Krafur 2005). Genetic studies of honey bees and silk moths were pursued with the goal of improving these beneficial insects (Rinderer 1986, Robinson 1971, Tazima 1964, Tazima et al. 1975). By 2012, the completed genome sequences of a variety of insects and mites had been sequenced, allowing much more sophisticated analysis of their biology, ecology, behavior, and genetics. Genomes sequenced to date include the honey bee *Apis mellifera*, silk moth *Bombyx mori*, pea aphid *Acyrtosiphon pisum*, the parasitoid *Nasonia*, the mosquitoes *Anopheles gambiae* and *Aedes aegypti*, the flour beetle *Tribolium castaneum*, the twospotted spider mite *Tetranychus urticae*, the tick *Ixodes scapularis*, red imported fire ant *Solenopsis invicta*, the human body louse *Pediculus humanus*, and the monarch butterfly *Danaus plexippus*. Plans are under way to sequence 5000 arthropod genomes now that sequencing has become relatively inexpensive using next-generation sequencing methods (Robinson et al. 2011). Genome sequencing and the results obtained are discussed in Chapter 7.

Analyzing genes, development, and genetic systems from insects other than *Drosophila melanogaster* could help solve both basic and applied problems because *D. melanogaster* may be a highly specialized insect with unique genetic characteristics.

## 4.6 Dynamic Insect Genomes

Until recently, the eukaryote genome was considered to be relatively stable, with every cell having the same DNA sequences in the same amounts and in the same location. Genomes were perceived to respond slowly to evolutionary pressures. It is now apparent that somatic genomes are more diverse than previously imagined, with polyteny, polyploidy, and gene amplification occurring in different tissues at different developmental stages in the organism (Edgar and Orr-Weaver 2001). It is also clear that DNA can move within the nuclear genome via a wide array of transposable elements (TEs).

DNA has been found in interesting structures outside the nuclear chromosomes and mitochondrion, but their significance is unresolved. For example, covalently closed circular DNAs that seem to be derived from chromosomal DNA have been found in cell cultures of *Drosophila* (Gaubatz 1990). Much of this circular DNA is middle-repetitive chromosomal DNA and may be associated with gene amplification during development or with DNA rearrangements during aging (Gaubatz 1990). Some circular DNA molecules in *D. melanogaster* embryos apparently contain 5S ribosomal RNA genes, satellite DNA, or histone genes (DeGroote et al. 1989).

Minichromosomes have been found in *D. melanogaster* that apparently originated from the transposable element TE1 (Block et al. 1990). The minichromosome contains two structural genes, *white* and *roughest*, from the *Drosophila* X chromosome and part of chromosome 2. This minichromosome was relatively stable and inherited by 33–47% of progeny, indicating that it contains a centromere. Centromere-like elements lacking chromosome arms have been found in the phorid *Megaselia scalaris* (Wolf et al. 1991). The function of these elements is unknown, but they could be B chromosomes that have been reduced to a minimal size.

#### 4.6.1 Horizontal Gene Transfer from Microorganisms to Insects

Moran and Jarvik (2010) documented that horizontal or lateral transfer of carotenoid-production genes from fungi to aphids has occurred. The aphid carotenoid genes are related to fungal genes, and the fungal genes were found to have integrated into the aphid genome and duplicated. As a result, aphids are the only insects known to produce carotenoids. How the fungal genes transferred into the aphid genome is unknown. Contamination of the sample with fungal DNA was ruled out, and sequence analysis suggests the genes transferred as a single event, preserving the gene arrangement observed in certain fungi. The genes were found in both *Aphis pisum* and *Myzus persicae*, suggesting the transfer preceded their shared ancestor. In *A. pisum*, the presence of red coloration from the carotenoids is correlated with different susceptibilities to natural enemies.

Horizontal transfer of genes from bacterial endosymbionts to arthropod genomes also has been found. Hotopp (2011) suggested transfer of endosymbiont DNA to animals is probably more common than transfer of genes from free-living bacteria, especially if the endosymbiont is present in the germ line. Several examples are known in which portions of the genome of the symbiont *Wolbachia* have been transferred into the host genome. The bean beetle *Callosobruchus chinensis* contains ≈30% of a *Wolbachia* genome inserted into the X chromosome. Some of these genes are transcribed at a low level, but it is not known whether the genes are functional (Hotopp 2011). *Drosophila ananasae* has nearly the entire 1.4-Mb *Wolbachia* genome in the 2L chromosome. At least 28 *Wolbachia* genes are transcribed at low levels (Hotopp et al. 2007).

The parasitoid wasp *Nasonia vitripennis* genome contains 13 proteins previously only found in poxviruses, but the proteins also are found in multiple strains of *Wolbachia*, and analysis indicated that the wasps acquired these proteins from *Wolbachia* (Werren et al. 2010). Once *Nasonia* had the proteins, the genes were duplicated and diverged over time; the genes are transcribed in

males and females through various life stages. [Hotop et al. \(2007\)](#) concluded that up to 70% of *Wolbachia*-infected hosts might have *Wolbachia* genes inserted into their nuclear genomes and that horizontal gene transfer from microorganisms may have a significant impact on insect evolution.

Analysis of the genome of the pea aphid, *Acrthosiphon pisum*, indicates that only 12 genes or gene fragments are of bacterial origin, with only two from *Buchnera aphidicola*, their obligatory endosymbiont, both of which seem to be inactive ([Nikoh et al. 2010](#)). The other genes seem to be related to genes from relatives of *Wolbachia*. At least eight of the genes seem to be functional, and expression of seven of these genes is highly up-regulated in bacteriocytes. Thus, [Nikoh et al. \(2010\)](#) concluded that reduction of the genome in *Buchnera*, the ancient endosymbiont in aphids with a highly reduced genome (1/7th the size of related bacteria), did not result in transfer of *Buchnera* genes to the aphid nuclear genome. Rather, it seems that aphids use a set of duplicated genes acquired from other bacteria to aid the *Buchnera*–aphid mutualism.

[Acuna et al. \(2012\)](#) found a bacterial gene was transferred into the coffee berry borer, *Hypothenemus hampei*, a beetle pest of coffee. The gene encodes a mannanase that hydrolyzes the major storage polysaccharide in the coffee bean, that is the beetle's presumed food. The gene is found in *H. hampei* but not in the closely related *Hypothenemus obscurus*, indicating the gene has possibly allowed adaptation to a new host plant. This is the first time this gene has been found in an insect. A phylogenetic analysis of diverse mannanase genes from prokaryotes and eukaryotes indicated the beetle gene was most closely related to prokaryotic mannanases, grouping with the *Bacillus* clade. The gene is flanked by two eukaryotic transposons of the hAT and Tc1/mariner superfamilies, further supporting the hypothesis that the gene is integrated into the beetle genome and not present in gut symbionts. Furthermore, the presence of the TEs suggests a possible mechanism for the horizontal gene transfer event.

## 4.7 B Chromosomes

B chromosomes are a heterogeneous class of, often heterochromatic, chromosomes sometimes referred to as accessory or supernumerary chromosomes that occur in plants and animals. B chromosomes may have little effect on the phenotype, differ in number from one cell type to another, and may occur only in some individuals of the species. B chromosomes may not segregate normally in mitosis and meiosis ([Wagner et al. 1993](#)). A B chromosome in the parasitic wasp *Nasonia vitripennis* causes the compaction and loss of paternally derived chromosomes in fertilized eggs, leading to the production of all male progeny

in this arrhenotokous (haplodiploid) species (Eickbush et al. 1992). The notion that insect genomes simply consist of nuclear and mitochondrial chromosomes should be discarded (Pardue 1991).

The notion that eukaryotic genes should contain introns is not always sustained. For example, hemoglobin genes sequenced from the midge *Chironomus thummi* lack introns, even though they show sequence similarity with vertebrate hemoglobin genes that do contain introns (Antoine and Niessing 1984). Because the cloned *Chironomus* genes were expressed *in vivo*, the hypothesis that they are pseudogenes was rejected. (A **pseudogene** is a gene with a close resemblance to a known gene, but it is nonfunctional because mutations prevent normal transcription or translation.) An alternative explanation is that the hemoglobin genes originated by **reverse transcription** of spliced messenger RNA (mRNA) in germ-line cells. Reverse transcription involves synthesis of DNA from an mRNA template, which lacks introns, to produce complementary DNA (cDNA). If the intronless cDNA subsequently integrated into the *C. thummi* genome, then the hemoglobin gene would lack introns.

Several major categories of nuclear DNA are known: unique-sequence, middle-repetitive, and highly repetitive. Within the middle-repetitive class of DNA, examples are presented here of some particularly interesting insect genes.

## 4.8 Unique-Sequence DNA in the Nucleus

Most protein-coding genes are contained in unique-sequence DNA. The proportion of unique sequences varies among species. For example, among four lepidopterans—*Antheraea pernyi*, *Hyalophora cecropia*, *Bombyx mori*, and *Manduca sexta*—the proportions of unique DNA range from 55 to 80% (Berry 1985).

Some unique-sequence DNA is present in multiple copies in specific insect cells or tissues. This occurs by one of two mechanisms: multiple copies of unique sequences occur if the cells are polyploid (polyploidy means that cells contain multiple copies of each chromosome, with  $n > 2$ ), or multiple copies of unique sequences also can occur through **gene amplification** in which a portion of the chromosome is replicated. For example, the chorion genes of *Drosophila* are amplified during specific stages of chorion production (see Section 4.11.1), although this amplification is limited to ovarian follicle cells. Some insects that are resistant to pesticides have amplified genes (see Section 4.11.2).

## 4.9 Middle-Repetitive DNA in the Nucleus

Middle-repetitive DNA is found in more than one copy, but still in modest amounts. Such sequences include genes that code for ribosomal RNAs (rRNAs),

transfer RNAs (tRNAs), histone proteins, transposable elements, and developmentally regulated multigene families such as actins, cuticle genes, heat-shock genes, larval-serum genes, silk genes, and yolk-protein genes. One solution to producing large amounts of gene product in a relatively short time and in a coordinated manner is to duplicate the gene. Duplicated genes may be present in tandem arrays on the same chromosome or be present on separate chromosomes.

#### 4.9.1 Heat-Shock Genes

The heat-shock response originally was discovered in *D. melanogaster* and has since been found in organisms ranging from bacteria to humans. Heat-shock genes are activated in response to environmental stresses such as heat or chemical shock. Heat-shock proteins are present in small amounts in many cells in the absence of stress, but they rapidly increase after stress. Heat-shock genes are an evolutionarily conserved response to stress in all organisms (Morimoto et al. 1992).

If *Drosophila* adults are exposed to a severe heat shock ( $\approx 40^{\circ}\text{C}$ ), most die. If they undergo a mild shock at  $33^{\circ}\text{C}$ , additional heat-shock proteins are synthesized, and many flies then can survive subsequent heat shocks at  $40^{\circ}\text{C}$ . In *D. melanogaster*, nine chromosomal sites puff in response to heat shock, and specific mRNAs are produced that code for seven heat-shock proteins. There are several types (or families), including the *hsp70*, *hsp83*, and the small heat-shock gene family (Pauli et al. 1992).

*hsp70* is virtually inactive in unstressed cells, but hsp proteins become very abundant during and after heat shock, accounting for 1% of the total cellular protein (Feder and Krebs 1997). There are 10 copies of the *hsp70* gene in *Drosophila*. It is the most abundant and highly conserved. At the amino-acid level, the *Drosophila* hsp70 protein shares 73% overall similarity with the human and 50% with that of the bacterium *Escherichia coli*. In addition, seven cognate genes of *hsp70* are constitutively expressed and may be important during *Drosophila* development. The hsp70 proteins are molecular chaperones, minimizing aggregation of peptides in nonnative conformation.

The *hsp83* gene products are general chaperones involved in several developmental pathways in *D. melanogaster* and have both housekeeping and stress-related functions (Rutherford and Lindquist 1998, Mayer and Bukau 1999). When the *D. melanogaster* *hsp83* gene is mutated or impaired, variability in many adult structures is induced, with specific variants depending on the genetic background. Thus, *hsp83* gene products buffer variation, allowing genetic variation to accumulate under neutral conditions. When the organism

is stressed by heat or cold, the hidden variants are expressed and selection then could lead to the continued expression of these traits, providing a mechanism for promoting evolutionary change in an otherwise entrenched developmental process. The protein may act as a “capacitor for morphological evolution” (Rutherford and Lindquist 1998).

The small *hsp* gene family includes genes encoding hsp22, hsp23, hsp26, and hsp27 proteins, all of which are expressed at several developmental stages.

#### 4.9.2 Histone Genes

Five histone-gene families code for the basic proteins contained in eukaryotic chromosomes. The basic unit of chromosomes, the nucleosome, is composed of 146 bp of DNA coiled around a histone octamer, two molecules each of histone H2A, H2B, H3, and H4 (see Figure 3.1). Linking two nucleosomes is a small stretch of DNA to which the fifth histone, H1, is bound. The histone genes share regulatory sequences and are coordinately expressed. In some species, there are tissue- or stage-specific gene sets. In *Drosophila*, the histone genes are tandemly repeated and closely linked. The histone genes of the midge *Chironomus thummi* are different from those found in *D. melanogaster* (Hankeln and Schmidt 1991).

Histone genes typically lack introns. It is thought that introns were eliminated because these genes must be expressed efficiently and rapidly during development. Having histone genes organized in a tandem-repeat structure also ensures that there will be equivalent amounts of the five proteins produced.

There is a 10-fold difference in copy numbers of histone genes in three species of *Drosophila* (*melanogaster*, *hydei*, and *hawaiensis*) (Fitch et al. 1990). *D. melanogaster* has 100 tandemly arranged histone genes, far more genes than would be required for the maximal rate of transcription during development. By contrast, *D. hydei* has five and *D. hawaiensis* has  $\approx$ 20 tandem histone-repeat copies per haploid genome. In *D. melanogaster*, the histone genes are located adjacent to a heterochromatic region of chromosome 2, whereas they are located in euchromatic regions in the other two species. This difference suggests there are more histone genes in *D. melanogaster* to compensate for the fact that the genes are less active because they are located near heterochromatin.

#### 4.9.3 Immune-Response Genes

Insects defend themselves against bacteria, viruses, fungi, and parasitoids with both cellular and humoral immune responses (Gillespie et al. 1997, Khush and Lemaitre 2000, Carton and Nappi 2001, Lazzaro and Rolff 2011). The first line

of defense against microbes is structural and comprises the exoskeleton, the peritrophic membrane that lines the gut, and the tracheal linings. In addition, insects typically maintain a low pH and digestive enzymes and antibacterial lysozymes in their midguts.

Infections induce local immune responses that include the synthesis and secretion of peptides in barrier tissues such as the tracheal and gut epithelium. In addition, systemic responses are activated that result in encapsulation of pathogens by blood cells, melanization of parasites and pathogens, phagocytotic uptake of pathogens by blood cells, and the production by the fat body of anti-fungal and antibacterial peptides that are secreted into the hemolymph where they accumulate to high concentrations (Khush and Lemaitre 2000). Many antibacterial proteins and peptides, such as cecropins, attacins, lysozymes, and defensins, serve as induced immune responses.

As discussed in Section 4.12, insects are hosts to microbial symbionts. One question is, How does an insect discriminate between useful microbial symbionts and pathogenic microbes? Another question is, How does an insect maintain an appropriate level of beneficial microorganisms? Lazzaro and Rolff (2011) point out the immune system is able to maintain homeostasis and regulate microbes that provide useful services to the insect. This service requires the immune system to distinguish between cues that are produced by pathogens (danger signals) while maintaining the beneficial microbes, which are producing molecular signals, as well. Both types of organisms produce molecular cues. Current data suggest that the joint presence of signals from microbes **and** the presence of danger signals, such as the release of collagen or nucleic acids from damaged cells, are required to elicit an immune response to pathogens. Molecules produced by gut bacteria, for example, induce expression of host proteins that degrade these molecules, allowing the insect to regulate the abundance of gut symbionts without eliminating them. Similar responses occur in the specialized structures such as mycetocytes inhabited by symbionts in the insect body. The current concept for how insects can maintain microbial symbionts while responding appropriately to pathogens is as follows: where microbial signals are present, but no damage is done, the physiological cost of clearing the infection is not worthwhile, allowing beneficial microbes to persist. Only if danger signals occur in combination with the molecular cues produced by the microorganisms will the immune system respond.

Some organisms have evolved to evade the insect immune system. For example, an analysis of the immune responses by the mosquito *Culex quinquefasciatus* to West Nile virus, the filarial worm *Wuchereria bancrofti*, and an avian malaria parasite showed that mosquito-borne pathogens have "evolved to

evade innate immune responses in three vector mosquito species of major medical importance" ([Bartholomay et al. 2010](#)).

#### 4.9.4 Ribosomal Genes

The ribosome is the site in the cell where proteins are synthesized ([Frank 2000](#)). The ribosome is a particle made of two subunits, each formed of ribosomal RNAs (rRNAs) and ribosomal proteins. Protein synthesis is a serious business, and it has to be done quickly and accurately. A typical protein takes  $\approx$ 15 seconds to make ([Frank 2000](#)). Protein folding ensures the proper function of the protein, and the folding relies on the location of particular amino acids, which can be jeopardized by even a single point mutation. Typically, ribosomes have an error rate of only 1 in 1000 to 10,000 amino acids. Ribosomes take up much of the cell's mass and much of the cell's metabolism is devoted to making ribosomal proteins and rRNAs. Ribosomes interact with mRNAs, initiation factors, and transfer RNAs during protein synthesis; >120 macromolecular components are needed to produce polypeptides in ribosomes ([Kaulenas 1985](#)).

Different arthropod species have different numbers of ribosomal genes located in the nuclear chromosomes. For example, *Drosophila erecta* has 160 genes, whereas *D. hydei* has >500 ([Berry 1985](#)). The fungus fly *Sciara coprophila* has 65–70 genes, one of the lowest numbers reported ([Kerrebrock et al. 1989](#)). Most insect genomes have between 200 and 500 rRNA genes. The ribosomal genes of *Drosophila* are arranged into two clusters, with one cluster in the nucleolar organizer of each of the sex chromosomes ([Williams and Robbins 1992](#)). The 5.8S, 18S, and 28S rRNAs are transcribed as a single unit that is then processed in the nucleus to provide the separate subunit RNAs. Ribosomal genes comprise 2% of the total genome and  $\approx$ 20% of the middle-repetitive sequences of *D. melanogaster*.

#### 4.9.5 Silk Genes

Silk is important in the life history of many insects. Silk is produced in labial glands, Malpighian tubules, and a variety of dermal glands ([Sutherland et al. 2010](#)). It seems that there are at least 23 different categories of silk-producing insects, possibly representing independent evolution in different lineages of insects. However, silks have high levels of protein crystallinity and similar amino-acid compositions, allowing silk to be produced in fine protein fibers. Silk is used in cocoons by Lepidoptera; to produce an egg stalk and cocoons by Neuroptera; to line tunnels by Ephemeroptera; to line tunnels and coat eggs by Embiididae; for egg coverings and nests by Psocoptera; to produce cocoons to protect against predators, high temperatures and low relative humidity by Thysanoptera; to produce silk rafts to support eggs by water beetles

(Hydrophilidae); to produce cocoons, nests and webs by Hymenoptera; to produce cocoons by Siphonaptera; and to produce underwater prey-capture nets by Trichoptera (Craig 1997, Sutherland et al. 2010). The best-studied are the silks produced by the Lepidoptera, including the silk moth *Bombyx mori*.

Silks are composed of one or more proteins called **fibroins**, proteins composed of several simple amino-acid sequences in reiterated arrays (Craig 1997, Sezutsu and Yukuhiro 2000). The silk gland provides a model system for cell biologists and molecular geneticists to study gene regulation and development. Silk-gland cells of *B. mori* are polyploid (up to 20-fold), which may explain how silk moth larvae produce huge amounts of silk proteins within a short time (5–6 days) before pupating. The cells from the posterior silk gland produce fibroin, and those from the middle part of the silk gland store fibroin and **sericin**, a mixture of four to six hot water-soluble polypeptides. Sericin binds strands of raw silk fibers together. Silk proteins have an unusual amino-acid composition, with a predominance of glycine and alanine in fibroin, and serine in sericin (Prudhomme et al. 1985).

Although *Drosophila melanogaster* does not produce silk, a silk-encoding gene (*P25*) of *Bombyx mori* was expressed in the anterior salivary gland after flies received the moth silk gene by *P* element-mediated transformation (Bello and Couble 1990; for a discussion of *P* element-mediated transformation, see Chapter 9). The *P25* *B. mori* gene was appropriately expressed in the fly larval salivary glands, indicating that *Drosophila* salivary glands can recognize *Bombyx* silk protein-coding sequences and control their expression, despite the evolutionary divergence of flies and moths >250 million years ago.

#### 4.9.6 Transfer RNA Genes

More than 90 transfer RNAs (tRNAs) have been identified during *Drosophila* development that are encoded by at least 670 genes, which can be divided into 60 separate groups. One to 18 tRNA genes are contained in each of 30 chromosomal sites, but there are no tandem repeats.

#### 4.9.7 Vitellogenin Genes

Yolk proteins provide embryos with nutrients essential for growth within the egg. Most are phosphoglycoproteins, and they provide a source of amino acids, phosphate, lipids, and carbohydrates. The major yolk proteins are derived from **vitellogenins** that are produced by the fat body and secreted for uptake by maturing oocytes. Vitellogenin-gene structure and regulation have been studied in *Locusta migratoria*, the tobacco hornworm *Manduca sexta*, *Bombyx mori*, the boll weevil *Anthonomus grandis*, the Mediterranean fruit fly *Ceratitis*

*capitata*, and *Drosophila* (Bownes 1986, Rina and Savakis 1991, Trewitt et al. 1992). The fat body of the mother is the primary producer of yolk proteins, but some yolk proteins are synthesized by the follicular epithelium of the ovary in *D. melanogaster*.

Yolk proteins in *Drosophila* consist of three polypeptides: YP1, YP2, and YP3. YP1 is expressed by the fat body and, after posttranslational processing and glycosylation, the proteins are secreted into the hemolymph and delivered to the oocyte. YP2 is expressed in ovaries. The production and delivery of the three proteins are coordinately regulated and under the control of two hormones, 20-hydroxyecdysone and juvenile hormone (Bownes 1986). These two hormones also regulate molting and metamorphosis during development.

Production of yolk proteins begins during the first day of *Drosophila* adult life. The production rate is high, with yolk proteins representing  $\approx\!1/3$  of the total proteins in the hemolymph. YP1 and YP2 are closely linked genes on the X chromosome, whereas YP3 also is sex-linked but more distant. YP1 and YP2 show much sequence homology and probably resulted from a fairly recent gene duplication event. Only one small intron is found in YP1 and YP2, and two introns in YP3. Extensive yolk-protein synthesis in *Drosophila* is achieved because tissues are polytene and polyploid.

#### 4.9.8 Transposable Elements

Transposable elements (TEs) are middle-repetitive DNA sequences that can move (transpose) to new sites, invert, and undergo deletion or amplification (Berg and Howe 1989, Finnegan 1989). TEs may cause damage in genomes and must establish in the germ line to be maintained in the population, so they are considered "selfish elements." Organisms have several mechanisms of reducing the impact of TEs on their genomes (Blumenstiel 2011, Levin and Moran 2011). For example, host DNA that is methylated inhibits transcription of TEs. Small RNAs (small interfering RNAs [siRNAs] and Piwi-interacting RNAs [piRNAs]) defend against TEs (Blumenstiel 2011). As one might expect, TEs counter these defense mechanisms in their hosts when mobilizing in germ cells or during early development. The evolution of suppression systems in the host can result in an evolutionary arms race that drives a high rate of evolution in the TEs.

TEs are not always bad for a genome. They can generate genetic variability, sometimes they can acquire useful functions in genomes, and they can modify genome structure and functions, resulting in diversity upon which evolution can act (Feschotte and Pritham 2007, Blumenstiel 2011, Werren 2011). In fact, Biemont (2009) proposes that TEs are important in chromatin formation

and centromere function, the assembly of synaptonemal complexes (the protein structure that forms between homologous chromosomes during meiosis and that is involved in pairing and crossing over), and cell division. Thus, although TEs are silenced, they also are maintained in genomes, and, according to Biemont (2009), "The presence of numerous TEs in centromeric and pericentromeric regions can be considered to be a key to success in the evolution of eukaryote organisms owing to their function in cell division."

TEs initially originally were divided into two classes according to their structure and mechanism of transposition (Table 4.1). Class I elements transpose by reverse transcription of an RNA intermediate. Class I elements include elements

**Table 4.1: A Classification of Transposable Elements by their Method of Transposition.**

<b>Class I Transposable Elements Transpose by Means of RNA Intermediates</b>	
A.	Viral superfamily (retrovirus-like retrotransposons) Have long direct-terminal repeats (LTRs), encode reverse transcriptase from open reading frames (ORFs) in DNA between LTRs, able to generate 4- to 6-bp target-site duplications, have no 3' terminal poly(A) tract, are dispersed in genome. Examples: <i>Copia</i> -like elements in <i>Drosophila melanogaster</i> <i>Gypsy</i> -like elements in <i>D. melanogaster</i> <i>Pao</i> in <i>Bombyx mori</i>
B.	Nonviral superfamily (nonviral retroposons) Have no terminal repeats, have ORFs, do not encode enzymes responsible for their transposition, have 3' terminal poly(A) tract, are dispersed in genome. Examples: <i>F</i> family in <i>D. melanogaster</i> R2 retroposons in many insects HeT-A retroposons in telomeres of <i>D. melanogaster</i>
<b>Class II Transposable Elements Transpose Directly from DNA to DNA</b>	
A.	All have a transposase and terminal inverted repeats (IRs) With short inverted repeats (SIRs) Examples: <i>P</i> and <i>hobo</i> in <i>D. melanogaster</i> <i>mariner</i> in many insect species
B.	With long inverted repeats (LIRs) Example: <i>FB</i> ( <i>foldback</i> ) in <i>D. melanogaster</i>
<b>Rolling-Circle Transposons (<i>Helitrons</i>)</b>	
	<i>Helitrons</i> do not have inverted repeats and move by a different method than class I and II elements. <i>Helitrons</i> have been found in diverse organisms, including insects. Transposition is by a semireplicative model; only one strand of the transposed transposon is transferred from one genomic site to another where it serves as a template for DNA synthesis catalyzed by the host repair mechanism.

(From Finnegan 1989, Robertson 1993, Xiong et al. 1993, Kapitonov and Jurka 2007.)

related to retroviruses that have long-terminal repeats (LTRs). They also include elements that lack long-terminal repeats (non-LTR retrotransposons).

Class II elements transpose directly from DNA to DNA. They include elements with short-inverted terminal repeats and have a coding region for a transposase. They also include elements with long-inverted repeats. Many TEs have been discovered in *D. melanogaster* (Bowen and McDonald 2001). A diversity of TEs is known from other insects as well.

Another group of transposable elements, called *Helitrons*, have been found (Kapitonov and Jurka 2001, 2007, 2008). *Helitrons*, a new class of TEs, are found in eukaryotes (Kapitanov and Jurka 2001). These unusual TEs, first found in plants and the nematode *Caenorhabditis elegans* but now known to occur in many organisms, including insects, seem to replicate by a rolling-circle method. Transposition is thought to occur by movement of one strand from one genomic site to another, where it is the template for DNA repair, resulting in a double-stranded DNA insert (Kapitonov and Jurka 2007).

At least half of all spontaneous mutations in *D. melanogaster* are due to insertions of TEs. For example, *P* elements in *D. melanogaster* cause excisions, chromosome rearrangements, and insertions. The *foldback* (*FB*) transposon (the names of most TEs are italicized) is associated with deletions, inversions, reciprocal translocations, and insertional translocations in which normally unique *Drosophila* DNA is flanked by two FB elements. All well-characterized, highly unstable genes in *D. melanogaster* contain either the *P* element or FB elements (Berg and Howe 1989). Members of the HeT-A and TART families of TEs are found at telomeres and in centromeric heterochromatin and never in the euchromatin regions of chromosomes in *D. melanogaster* (Mason et al. 2000). Kapitonov and Jurka (2003) analyzed the newly sequenced genome of *D. melanogaster* and discovered additional, previously unknown TEs called *Transib*. They also found ancient, 5-million-year old ancestors of the *P* element, as well as *Helitrons* and they estimated that TEs are “three times more abundant than reported previously, making up to ~22% of the whole genome.” Biemont and Vieira (2006) reported “...50–80% of mutations are due to such insertions” and that TEs are “...more powerful producers of the raw material of evolution than is the classical nucleotide-base substitution rate, which is around  $10^{-8}$  to  $10^{-9}$  per nucleotide per generation.”

TEs may carry genetic information, regulate genes, or initiate genetic changes (Britten 1997, Miller et al. 1997, Shapiro 1999, Biemont and Vieira 2006). Wilson (1993) suggested that TEs could lead to resistance to pesticides, although he did not provide any direct evidence for this conclusion. Agarwal et al. (1993) found a TE named *Juan* associated with amplification of an esterase gene in

pesticide-resistant *Culex* mosquitoes, but a direct involvement in inducing gene amplification was not demonstrated. [Waters et al. \(1992\)](#) suggested the TE called 17.6 is involved in susceptibility to pesticides in *Drosophila* associated with a P450 gene. However, [Delpuech et al. \(1993\)](#) screened colonies of *D. melanogaster* and *Drosophila simulans* from around the world, and they found no relationship between the presence or absence of 17.6 and resistance. [Daborn et al. \(2002\)](#), however, found a P450 allele associated with resistance to DDT due to overtranscription of the gene and associated with the insertion of an Acord TE into the 5' end of the gene. [Aminetzach et al. \(2005\)](#) found that insertion of long interspersed elements (LINEs) is associated with "increased resistance to organophosphates," by the generation of a new protein.

One example of TEs containing genetic information may be found in *Drosophila hydei* males. TEs and repetitive DNA sequences comprise the majority of the Y chromosome of *D. hydei*. Apparently, the lampbrush loop-forming fertility genes on the Y chromosome consist, at least in part, of retrotransposons of the *micropia* family ([Huijser et al. 1988](#)).

R1 and R2 are class I TEs originally found in some of the 28S rRNA genes of the silk moth *Bombyx mori*, and several dipterans. A survey suggested that R1 and R2 elements occur within the rRNA genes of many insects ([Jakubczak et al. 1991](#)). Forty-three of 47 species surveyed, including species in Odonata, Orthoptera, Dermaptera, Hemiptera, Coleoptera, Hymenoptera, Lepidoptera and Diptera, contained the insertions in 5–50% of their 28S genes ([Burke et al. 1993](#)). The broad distribution of these elements raises the question of whether they could have been present in insects before their radiation. The R2 elements block the production of functional 28S rRNA by the inserted gene, but this is not usually deleterious to the organism because insects contain many more rRNA genes than are needed ([Eickbush and Eickbush 2011](#)). [Ye et al. \(2005\)](#) found that R1 and R2 non-LTR retrotransposons compete; R1 elements insert into a site that is 74bp downstream of the R2 insertion site, so the presence of an R2 element may inhibit the expression of R1. As a result, the R1 element of *D. melanogaster* rarely inserts into 28S genes already containing an R2 element.

Very little is known about the origin and evolutionary history of TEs. A TE family might originate in a species, or TEs might be acquired by **horizontal transmission** from another species. Normally, DNA or RNA sequences are transmitted *vertically* from parent to progeny, but in horizontal transfer, DNA sequences are transferred laterally across species, taxonomic borders that were once thought to be inviolable ([Daniels et al. 1990](#), [Kidwell 1992](#), [Plasterk et al. 1999](#)). For example, the *hobo* transposable element of *D. melanogaster* has a

similar sequence to TEs from plants (*Activator* from corn and *Tam3* from snap-dragon) (Calvi et al. 1991). Another element, *jockey*, identified from *D. melanogaster* has been found in the distantly related *Drosophila funebris*, but not in species closely related to *funebris*. This occurrence again suggests that *jockey* moved horizontally from *D. melanogaster* into the genome of *D. funebris* (Mizrokhi and Mazo 1990).

A possible superstar at horizontal transfer may be *mariner*. Originally, *mariner* was found in *Drosophila mauritiana* and several other species of *Drosophila*, as well as in the moth *Hyalophora cecropia*. Subsequently, Robertson (1993) found that several types (subfamilies) of *mariner* are widespread in insects. It is found in other organisms as well. For example, *mariner* was found in the predatory mite *Metaseiulus occidentalis* (Acari: Phytoseiidae) (Jeyaprakash and Hoy 1995). It is likely that *mariner* has moved horizontally among diverse insect and mite species, although the frequency of horizontal transfer is infrequent on a human time scale (Robertson and Lampe 1995). Many *mariner* elements have degenerated and become inactive in the genomes of their hosts.

Relatively little information is available as to how TEs invade populations and the mechanisms involved in the first step of the invasion (Biemont et al. 1999). One of the best-known examples is the invasion of *D. melanogaster* by the *P* element (for a review, see Chapter 9). Another evaluation of TE invasion was carried out by Biemont et al. (1999) in natural populations of *D. simulans*: populations around the world are in the process of being invaded by a variety of TEs with whimsical names. Populations differ in the number and type of TEs, with some TEs absent in most populations, except for one or two populations that have high copy numbers. Biemont et al. (1999) suggest that the “initially selfish genes will surely appear more and more as ‘symbionts’ that have played a major role in evolution and that may still provide genomic flexibility and variability for population adaptation.”

If horizontal transmission of TEs occurs with some regularity, the implications are dramatic for evolutionary theory, likely serving as a “major force propelling genomic variation and biological innovation” (Schaack et al. 2010). Horizontal-transfer risks also could influence regulations regarding the risks associated with releases of transgenic arthropods into the environment (Brosius 1991, Plasterk 1999, Hoy 2000). As discussed in Chapter 9, studies of the transfer of *P* elements by a mite vector provide an intriguing glimpse at one possible mechanism by which TEs are able to move between species (Houck et al. 1991).

Other possible vectors of TEs are insect viruses. Insect viruses may carry DNA from their hosts; a proportion of foreign DNA within insect viruses consists of

TEs from the viruses' insect hosts (Fraser 1985). Four families of TEs were found in the triatomine bug *Rhodnius prolixus* that feeds on the blood of opossums, squirrel monkeys, and other vertebrates (including humans, to which it can transmit Chagas' disease) (Gilbert et al. 2010). The sequences of the TEs in the insect and in the vertebrates were very similar, suggesting that host-parasite interactions are important in horizontal transfer among animals.

New TEs are being found on a consistent basis, and a database of repetitive elements, including TEs, can be found at the website for the Genetic Information Research Institute (GIRI) (Jurka et al. 2005).

Thomas et al. (2010) found that *Helitrons* (rolling-circle transposons) have moved horizontally among mammals, reptiles, fish, invertebrates, and insect viruses. "The presence of horizontally transferred *Helitrons* in insect viruses [Bracoviruses], in particular, suggests that this may represent a potential mechanism of transfer in some taxa" (Thomas et al. 2010). Another Bracovirus was found to have a large DNA transposon (*Maverick*) in it, suggesting that TEs may transfer from parasitoid insects to their lepidopteran hosts by these viruses (Dupuy et al. 2011).

Yuan and Wessler (2011) analyzed the sequences of the catalytic portion of all superfamilies of cut-and-paste (class II) TEs in eukaryotes, and they found that all superfamilies have a common molecular signature and, thus, evolutionary origin. It seems that most superfamilies of cut-and-past TEs originated before the divergence of eukaryotic supergroups such as green algae, fungi, vertebrates, invertebrates, ciliates, and diatoms (Feschotte and Pritham 2007). All class II transposons probably rely heavily on horizontal transfer for their propagation and maintenance over evolutionary timescales.

## 4.10 Highly Repetitive DNA

Highly repeated DNA sequences with a uniform nucleotide composition can, upon fractionalization of the genomic DNA and separation by density-gradient centrifugation, form one or more bands that are clearly different from the main band of DNA and from the smear created by other fragments of a more heterogeneous composition. These sequences are called **satellite DNA**.

Satellite DNA is sometimes described as minisatellite or microsatellite DNA, depending on the length of the repeated sequences. **Microsatellites** consist of tandem repeats of between 1 and 6 bp, often in long arrays; like other classes of repetitive DNA, microsatellites have high mutation rates (Bachtrog et al. 1999). Satellite DNA can comprise a large fraction of the arthropod genome; for

example, in *Tribolium madens*, two satellite DNAs were characterized: one satellite DNA is a 225-bp monomer comprising 30% of the genome and the other satellite DNA is a monomer of 711 bp, constituting 4% of the genome (Durajlija Zinic et al. 2000).

The role of highly repetitive sequences in genome evolution is not well understood (Pardue and Hennig 1990, Ohno and Yomo 1991), but such sequences are associated with heterochromatin in the centromeres and telomeres, and they are important in chromosome pairing. The sequence (TTAGG)<sub>n</sub> is found at the extreme terminal region of all *Bombyx mori* chromosomes, as well as being associated with the ends of chromosomes in the Isoptera, Orthoptera, Hymenoptera, Trichoptera, Mecoptera, some Coleoptera, Hemiptera, and Lepidoptera (Okazaki et al. 1993). The sequence (TTAGG)<sub>n</sub> seems to be found only in arthropod telomeres, although not all arthropods have this telomeric sequence. *Drosophila melanogaster* and some coleopterans have different telomeric structures. The telomeres of *Drosophila* have TEs called HeT-A and TART in the subtelomeric region (Biessmann et al. 1993, Mason et al. 2000).

#### 4.11 Producing Large Amounts of Protein in a Short Time: Gene Amplification and Gene Duplication

When it is desirable to produce large amounts of gene product (protein) in a short time, several mechanisms could be used, including duplication of chromatids resulting in polyteny, polyploidy (the duplication of whole chromosomes), hypertranscription of the gene, gene amplification, and gene duplication. **Hypertranscription** involves producing large amounts of gene product from a single copy of genes on a chromosome, and it is the mechanism by which *D. melanogaster* males, which have only one X chromosome, produce as much gene product as females with two X chromosomes (for details, see Chapter 10). Definitions of gene amplification and gene duplication can be confusing (Edgar and Orr-Weaver 2001).

**Gene amplification** usually is defined as the replication of a gene (at a single locus) so that multiple copies can be transcribed at once. One way to visualize gene amplification is to imagine that gene amplification occurs by an “onion-skin model” in which a segment of the chromosome is replicated and multiple copies of that segment are transcribed, but the rest of the chromosome retains its normal structure. Gene amplification originally was coined to describe the production in mammalian cell cultures of multiple copies of cancer-drug resistance genes; gene amplification is associated with an initial low drug concentration of the drug, with the surviving cells subjected to multiple rounds of

selection with increasing concentrations of toxin. Gene amplification results in the production of more protein per unit time.

Gene amplification may be important to economic entomologists because some aphids and mosquitoes that are resistant to insecticides have amplified resistance genes, as described in Section 4.11.2, although the definition of gene amplification in these examples differ from that described for mammalian cells.

**Gene duplication** involves copying a gene multiple times; the copies may be maintained on the same chromosome in tandem array or be transferred to other chromosomes over evolutionary time.

#### 4.11.1 Chorion Genes in *Drosophila* and Moths

Both gene amplification and gene duplication occur in the chorion genes of *Drosophila* and the moths *Bombyx mori* and *Antheraea polyphemus* (Kafatos 1981, Eickbush and Burke 1985, Kafatos et al. 1986, Orr-Weaver 1991, Carminati et al. 1992, Lecanidou and Papantonis 2010). Analyses of these chorion genes resulted in significant advances in knowledge of the mechanisms of gene regulation and development.

In both *Drosophila* and silk moths, the egg is produced in the ovary and consists of follicles composed of three cell types: 1) the oocyte, 2) a small number of nutritive nurse cells connected to the oocyte, and 3) follicular epithelial cells that surround the oocyte and nurse cells. There are  $\approx$ 1000 follicular epithelial cells per follicle in *Drosophila* and up to 10,000/oocyte in silk moths. These cells synthesize a complex mixture of proteins and secrete them onto the surface of the oocyte to form the outer covering, or **chorion**. The chorion protects the embryo after fertilization and oviposition, preventing desiccation, yet enabling respiration to occur.

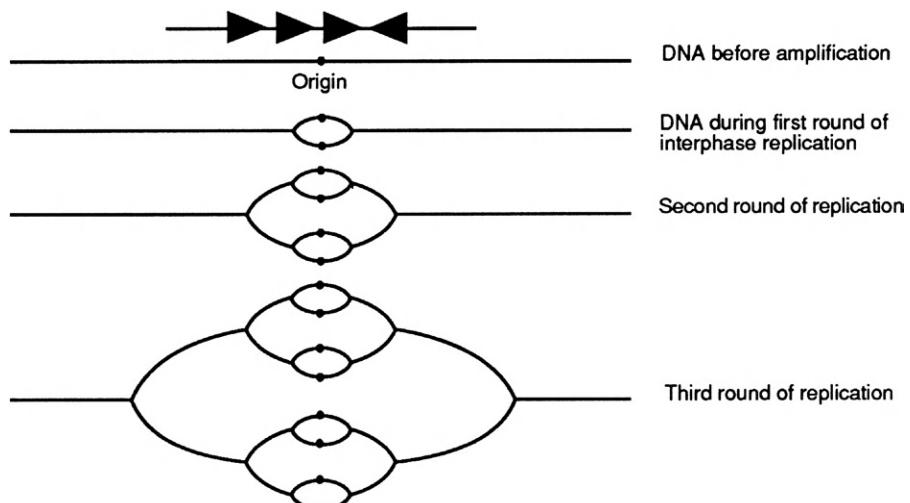
*Drosophila* and *B. mori* chorions are quite different. The *Drosophila* chorion is comparatively simple, with an endochorion and exochorion composed of six major and 14 minor proteins that are produced over  $\approx$ 5 hours. In silk moths, the number of genes and the time devoted to producing the chorion is much greater. There are three gene families in the wild silk moth *Antheraea polyphemus*, and the same three families, plus two others, in the domesticated silk moth *B. mori*. Approximately 100 chorion proteins are produced during a period of  $\approx$ 51 hours in silk moths (Kafatos 1981).

Moth and fly chorion genes are organized differently. In *Drosophila*, large amounts of the chorion proteins are produced in a relatively short time by *amplification* of the chorion genes (Lu et al. 2001). In *Drosophila*, two chorion-gene clusters, each 5–10kb, encode tandemly oriented chorion genes. One gene

cluster is found on the X chromosome, and one gene cluster is on the third chromosome. Chorion proteins could not be synthesized quickly and in sufficiently large quantities unless gene amplification occurred because each chorion–gene cluster is represented only once in the haploid genome. A 20-fold amplification of the chorion genes on the X chromosome and an 80-fold amplification of the genes on chromosome 3 is found in follicle cells. Amplification is achieved by replicating the DNA segments at multiple replication origins ([Heck and Spradling 1990](#)). DNA amplification extends bidirectionally for a distance of up to 40–50 kb to produce a multiforked “onion-skin” structure that contains multiple copies of DNA containing the chorion genes ([Figure 4.1](#)).

Gene amplification also occurs in the chorion genes of the Mediterranean fruit fly *Ceratitis capitata*. The overall organization of the cluster is similar to that of *Drosophila*, with the same four genes maintained in tandem, in the same order, and with similar spacing ([Konsolaki et al. 1990](#)). Despite the divergence of *Drosophila* and *Ceratitis* family lineages, ≈120 million years ago, there is high conservation in coding sequences and regulatory properties of their chorion genes.

Silk moth chorion proteins are produced over a longer time and involve larger numbers of genes that have probably arisen by **gene duplication**. More than 100



**Figure 4.1** Amplification of the *Drosophila* chorion genes in follicle cells. The first three rounds of DNA replication at the 66D locus on chromosome 3 are shown. The three small arrows represent three well-characterized chorion genes in this cluster. The polarity of a fourth chorion gene and the precise location of the origin are unknown. The boundaries of the amplified DNA are much larger than the chorion protein transcription units within it.

structural proteins are in the chorion of the silk moth *A. polyphemus*, which has an elaborate organization. In *B. mori*, the chorion genes are on chromosome 2 and consist of two segments that total >1000 kb of DNA. Subsets of the genes are expressed at different periods of choriogenesis (early, middle, late, very late), with the early proteins associated with framework formation, middle proteins with framework expansion, late proteins with densification, and very-late proteins with surface sculpturing of the chorion.

Silk moths have solved the problem of producing large amounts of protein quickly by gene duplication. Silk moth chorion genes are found in multiple copies of divergently transcribed, coordinately expressed pairs ([Kafatos et al. 1986](#)). For example, all members from each of two late-gene families are arranged in 15 pairs on a 140-kb segment. The members of each family have a high degree of sequence homology, although they are not identical. **Concerted evolution** could be the mechanism for maintaining this homology ([Eickbush and Burke 1985](#)). Concerted evolution often occurs in multigene families and could be maintained by two mechanisms: either unequal crossing over or gene conversion. Unequal crossing over may occur between the two sister chromatids of a chromosome during mitosis of a germ-line cell or between two homologous chromosomes at meiosis. It is a reciprocal recombination that results in sequence duplication in one chromatid or chromosome and a corresponding deletion in the other chromatid or chromosome. As a result of unequal exchange, daughter chromosomes become more homogeneous than the parental chromosomes. If the process is repeated, the numbers of each variant repeat on a chromosome will fluctuate with time, and eventually variant repeat will become dominant in the family. **Gene conversion** involves recombination in which a DNA fragment in one region is transferred to the corresponding place in another region so the transferred region becomes identical ([Osada and Innan 2008](#)).

Despite the very different organization of chorion genes in *Drosophila* and silk moths, silk moth chorion genes can function in *D. melanogaster*. Moth chorion genes were cloned into *P*-element vectors and inserted into the *D. melanogaster* germ line ([Mitsialis and Kafatos 1985](#)). (For a description of the methods used in inserting moth genes into *Drosophila* using a *P* element that has been modified to carry exogenous genes, see Chapter 9). Analysis of RNAs from transformed flies indicated that moth genes are expressed in an appropriate manner in the correct sex, tissue, and time in *D. melanogaster*. Fly and moth lineages diverged >250 million years ago, yet regulatory elements conferring sex, tissue, and temporal specificity of gene expression must have been conserved. Chorion gene promoter sequences from *Antheraea pernyi* and *A. polyphemus*

also functioned in *D. melanogaster* after *P* element-mediated transformation, although some regulatory interactions had diversified (Mitsialis et al. 1989).

#### 4.11.2 Insecticide Resistance

Geneticists have identified a new mechanism by which insects become resistant to pesticides (Mouches et al. 1990, Devonshire and Field 1991, Pasteur and Raymond 1996, Hemingway et al. 1998, Field 2000, Paton et al. 2000, Puinean et al. 2010). Amplification of esterase genes in the aphid *Myzus persicae* and the mosquito *Culex pipiens quinquefasciatus* results in identical gene copies present in tandem arrays in each cell.

Whether exposure to pesticides can induce resistance in insects by gene amplification is an interesting question. It has long been assumed that pesticide resistance in insects is due to the presence of rare alleles in populations that are selected for by pesticide applications (preadaptive mutations). However, amplification of genes in mammalian cells, plants, yeast, and microorganisms has been shown to occur in response to exposure to toxins (Stark and Wahl 1984). For example, amplification of the dihydrofolate reductase gene in mammalian cells in tissue culture occurs in response to exposure to the cancer drug methotrexate. A 100-fold amplification in a cholinesterase gene in two generations of a human family subjected to prolonged exposure to parathion was demonstrated and could be due to genetic changes induced by prolonged exposure to this insecticide (Prody et al. 1989). Mosquito cells selected with methotrexate also developed amplified genes (Fallon 1984). Thus, insecticide resistances due to gene amplification in insects could, at least in some cases, be induced by exposure to insecticides.

Resistance to neonicotinoids in the aphid *Myzus persicae* was investigated using microarrays containing all known detoxification genes. Overexpression (22-fold) of a single P450 gene was due to gene amplification (for a description of microarrays, see Chapter 6) (Puinean et al. 2010). The microarray analysis also showed overexpression of cuticular protein genes (2- to 16-fold), and penetration assays using radiolabeled insecticide indicated reduced cuticular penetration also contributed to the resistance.

### 4.12 Multiple Genomes in or on Insects: What is the “Biological Individual”?

Eukaryotes have a nuclear genome and a mitochondrial genome. Mitochondria are now generally accepted as microbial **symbionts** that were modified after a long process of evolution within eukaryotic cells (Gray 1989, Martin 1999). Mitochondria retain a distinctive genome that is replicated and expressed, but

mitochondria are incapable of independent existence. In the course of evolution, some mitochondrial genes were transferred to the nuclear genome of its host. Mitochondria are examples of symbionts that have evolved into organelles. The study of symbiosis in arthropods has revealed diverse relationships between arthropods and their symbionts. Molecular tools and sequencing of genomes likely will reveal more intriguing details of these relationships. The following examples may convince you of the critical importance of symbiosis in arthropod biology, ecology, behavior, and evolution that we are just beginning to understand.

#### 4.12.1 Multiple Symbionts

In addition to mitochondria, insects have intimate intra- and extracellular relationships with a diverse array of organisms, including viruses, Eubacteria, Archaea, yeasts, fungi, and rickettsia (Schwemmler and Gassner 1989; O'Neill et al. 1997; Bourtzis and Miller 2003, 2006, 2009; Douglas 1992, 2010; Zchori-Fein and Bourtzis 2011). The details of the relationships between the host and these microorganisms often are unknown, but we are learning rapidly with the aid of molecular tools. Insect symbionts may be beneficial to their hosts by providing nutrition, improving tolerance to environmental perturbations, enhancing fitness and host plant suitability, as well as maintaining or enhancing the insect immune system (Oliver et al. 2010, Weiss and Aksoy 2011). Caldera et al. (2009) recently noted the following:

*"Considering the diversity of insects and microbes on the planet, their mutual abundance and co-occurrence in virtually every terrestrial and fresh water habitat and their shared ancient evolutionary histories, it is likely that the biology of every insect species on the planet is influenced by microbial symbionts."*

For example, the rice weevil *Sitophilus oryzae* (Rhynchophoridae) has four intracellular genomes. These genomes are nuclear, mitochondrial, principal endosymbiont, and *Wolbachia* (Heddi et al. 1999, 2001). The principal endosymbiont is found ( $3 \times 10^3$  bacteria/cell) in specialized bacteriocytes. In total,  $3 \times 10^6$  bacteria are found in each weevil, 10-fold more cells than there are beetle cells (Heddi et al. 2001). These symbionts induce the specific differentiation of the bacteriocytes and increase mitochondrial oxidative phosphorylation through the supply of pantothenic acid and riboflavin. Their elimination impairs many important physiological traits, including flight ability. This weevil supports the "serial endosymbiotic theory"; according to this view, endosymbiosis did not occur just once in eukaryotic evolution with the origin of a nucleus, or even twice, when an anaerobic protist acquired a respiration bacterium to give rise to the mitochondrion. The acquisition of genomes by eukaryotic cells "continues today in the multicellular organism" (Heddi et al. 1999). The rice weevil gut symbiont allowed the weevil to colonize cereal plants because it supplied

vitamins. [Heddi et al. \(1999\)](#) “consider symbiosis in the rice weevil a sophisticated mechanism for acquiring new sets of genes.”

Symbionts may be intracellular or extracellular within the insect, and they may inhabit specialized structures. If located within the insect, the symbiont may inhabit the gut, the reproductive tract, the fat body and the salivary glands ([Box 4.1](#)). They also may occur on the outside of the insect, sometimes in specialized structures of the exoskeleton. Symbionts may possess metabolic capabilities that the insect host lacks, and the insect uses these capabilities to survive on poor or unbalanced diets while the symbiont obtains nutrition and shelter ([Douglas 1998](#)). Under such circumstances, the insect–microbe relationship is required (obligate symbiosis). Many insects freed of their symbionts grow slowly and produce few or no progeny; many symbionts cannot grow outside their insect host, indicating the relationship is long and intimate. Intracellular symbionts are found in the Anoplura, Mallophaga, Isoptera, Orthoptera, Hemiptera, Coleoptera, Diptera, and Hymenoptera. The amazing diversity of relationships and organisms involved in these relationships with insects has raised many questions, but provided relatively few clear-cut answers because many endosymbionts cannot be cultured outside their hosts. Many microorganisms are contained in special structures and transmitted by a highly specific method, including transovarial transmission, to progeny. Transmission also can occur when larvae feed on contaminated eggshells or feces.

#### Box 4.1 Symbiotic Relationships between an Insect and Microorganisms Can Be Complex

- Symbiosis is a broad term that includes parasites, pathogens, and mutualistic or beneficial interactions, although most associate symbiosis as having a positive or mutualistic relationship between the host and the symbiont.
- Symbionts in arthropods may be Eubacteria, fungi, yeasts, viruses, protozoa, or Archaea.
- Insects are in contact with many types of microorganisms in their environment. Some will be found externally, some internally. That does not make them a symbiont if they are only occasionally found associated with the insect.
- Symbionts may be found in the digestive tract or associated with it in specialized cells, in the fat body, in the reproductive system, free in the hemolymph, and in the salivary glands.
- Symbionts may be obligatory, or facultative in relation to their arthropod host. Many insects harbor multiple symbionts, including obligatory and one or more facultative organisms.
- Obligatory symbionts often are intracellular and may exist in specialized structures (mycetocytes, pouches, cavities) and are transmitted in specialized ways, including transovarial transmission. Sometimes transmission is by feeding on contaminated egg shells, or by contact in the environment.
- Some symbionts are extracellular but may be found internally. They may not inhabit specialized structures, although some do; often, these organisms are facultative or secondary symbionts, meaning that the insect can survive without them.
- Symbionts may provide nutrients, affect host range, temperature tolerance, longevity, fecundity, sex ratio, behavior, responses to natural enemies, or other aspects of their biology.

Some insect species contain several different types of symbionts in different tissues, including the gut, Malpighian tubules, fat body, or gonads. Bacteroids, spiroplasmas, rickettsia, mycoplasmas, or virus-like symbionts are found in dipteran testes, ovaries, pole cells, nurse cells, and gut wall cells. Symbionts in scale insects are particularly diverse, with almost 20 different types of associations described so far. In the leafhopper *Euscelidium variegatus*, specific bacteria are essential for normal growth and development, breaking down uric acid in the host cells and synthesizing amino acids and vitamins. Symbionts are involved in normal egg development of *E. variegatus*; embryos artificially lacking symbionts fail to develop normal abdomens. It is hypothesized that some genes from this microorganism have been transferred to the nuclear genome of *E. variegatus* in a manner parallel to that of mitochondria.

Some insects lacking their symbionts are apparently completely normal. For example, in the beetle family Cerambycidae, all of which live in wood, some species have symbionts, whereas others lack them. The hypothesis that symbionts supply a nutrient deficiency in the insect's diet thus seems to be simplistic; some insects feeding on a well-balanced diet have symbionts.

In some cases, symbionts increase the likelihood that an insect vector can transmit (vector) a disease. For example, rickettsia-like organisms in the tsetse fly *Glossina morsitans morsitans* affect infection by the sleeping sickness trypanosomes ([Welburn et al. 1993](#)). The rickettsia-like organisms produce endochitinases in the tsetse gut that inhibit lectins in newly emerged adults. Tsetse flies lacking the rickettsia-like organisms are less susceptible to trypanosomes (refractory).

In some cases, symbionts may prevent vector insects from transmitting pathogens ([Teixeira et al. 2008](#), [Hughes et al. 2011](#), [Pan et al. 2012](#)). For example, a strain of *Wolbachia* induces cytoplasmic incompatibility and blocks transmission of dengue in the mosquito *Aedes albopictus* ([Blagrove et al. 2012](#)).

Microbial symbionts are common in insects, but a full understanding of their genetic and evolutionary role is an exciting area of research that is expanding at a rapid rate ([Schwemmler and Gassner 1989](#), [Moran and Baumann 2000](#), [Zchori-Fein and Bourtzis 2011](#)). In the cases that have been well studied, a genetic interplay between insect host and primary (obligate) symbiont occurs; each supplies factors to the other, and the microorganism has specific means of movement and relocation within the insect. Insects must recognize a symbiont as "self" rather than as foreign or the microorganism would be subject to the insect's immune system. Our understanding of how microorganisms have become incorporated into insect organ tissues and cells is advancing with the use of molecular tools.

von Dohlen et al. (2001) found that there can be “bugs within bugs within mealybugs.” Mealybugs (Pseudococcidae) have endosymbionts that live within the cytoplasm of large, polyploid host cells within a specialized structure (bacteriome). These symbionts provide nutrients to their hosts. The relationship between the mealybugs and these primary endosymbionts is ancient, perhaps dating to the origins of the families or superfamilies 100–250 million years ago (von Dohlen et al. 2001). The mealybug host *Planococcus citri* packages its intracellular endosymbionts into mucus-filled spheres that surround the mealybug-cell nucleus and occupy most of the cytoplasm. These spheres are structurally unlike eukaryotic cell vesicles and actually harbor two types of Proteobacteria. The two bacteria are not coinhabitants of the spheres. Rather, the spheres themselves involve the internalization of the one bacterium by the second bacterium, which may facilitate the exchange of genes and gene products that could slow or reverse the genetic degradation that is common to organelles or long-term intracellular symbionts over evolutionary time.

Broderick et al. (2006) found that midgut bacteria are required for *Bacillus thuringiensis* to be effective as an entomopathogen. Kikuchi et al. (2012) found that the bean bug *Riptortus pedestris*, a pest of legumes in Japan (as well as related bugs in the family Alydidae), has a gut symbiont in the genus *Burkholderia* in a posterior region of the midgut. The adult bug has up to  $10^8$  *Burkholderia* cells. The symbiont enhances growth and size of the bug and is transmitted when the bug acquires the symbiont as a second instar from the soil. If the bug acquires a strain of *Burkholderia* from the soil that is resistant to the pesticide fenitrothion, the bugs become resistant. Experimental applications of the pesticide increased the resistant bacterial strain to >80%. More than 90% of the bean bugs reared in the enriched soil established symbiosis with the pesticide-resistant *Burkholderia*, conferring a useful level of pesticide resistance to the bug. The bugs amplify the bacteria in their symbiotic organ and could spread the bacteria when they fly to new sites.

An unusual symbiosis involves the presence of the bacterium *Candidatus Midichloria mitochondrii* in the mitochondria of ticks (Pistone et al. 2012). (When a new bacterial species name is proposed, but the international committee on bacterial nomenclature has not yet accepted it, the proposed species is called *Candidatus*, which is italicized, but the proposed genus and species names are not italicized.) This bacterium is a member of the Rickettsiaceae and is found inside mitochondria in hard ticks. It has been best studied in *Ixodes ricinus* (Ixodidae), where it seems to inhabit 100% of the wild-caught females, although it is present in lower frequencies in laboratory populations. The bacteria replicate in the mitochondria, reducing the numbers of mitochondria. It

seems that the tick and the symbiont have not coevolved, but the bacteria are vertically transmitted from females to their progeny and are transmitted transstadially. [Pistone et al. \(2012\)](#) suggest, "The high prevalence of *M. mitochondrii* in wild-collected adult females of *I. ricinus* seems to suggest a beneficial role.... However, the decrease in the number of bacteria in laboratory-reared ticks, without visible effects on the biology of the hosts, suggests the presence of the symbionts, if beneficial, is facultative, at least in laboratory conditions."

Antlions (Myrmeleontidae) suck out the body fluid after first paralyzing their prey with a toxin produced by salivary gland secretions produced by bacteria located in the salivary glands. The paralyzing toxin produced by these bacterial endosymbionts is a homolog of GroEL, a heat-shock protein that functions as a molecular chaperone in *E. coli* ([Yoshida et al. 2001](#)). In the antlion, the GroEL protein may act on receptors in prey insects to induce paralysis. The antlion symbionts perhaps evolved this nonchaperone function to establish a mutually beneficial antlion–bacterium relationship. [Yoshida et al. \(2001\)](#) speculated that insecticidal proteins may be produced by other endosymbionts to help additional fluid-feeding predatory insects.

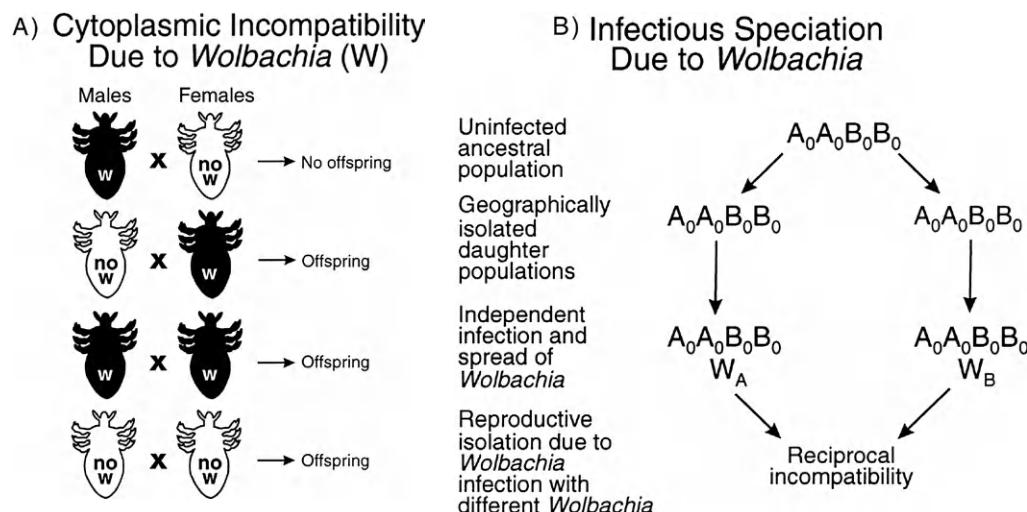
#### 4.12.2 *Wolbachia* and *Cardinium*

*Wolbachia*, a genus of  $\alpha$ -proteobacteria, commonly are found in arthropods ([Werren et al. 1995](#), [Rigaud and Rousset 1996](#), [O'Neill et al. 1997](#), [Jeyaprakash and Hoy 2000](#)). *Wolbachia* are intracellular gram-negative rods that cannot be cultured easily outside their hosts. *Wolbachia* infection rates range from  $\approx 17$  to 76% of all arthropod species ([Werren et al. 1995](#), [Jeyaprakash and Hoy 2000](#)). *Wolbachia* also have been found in crustaceans ([Rigaud 1999](#)) and nematodes ([Bandi et al. 1999](#), [Bazzocchi et al. 2000](#)). An understanding of their physiological and phenotypic effects on their hosts is still being developed. The phylogeny of *Wolbachia* in nematodes is congruent with the phylogeny of their hosts, suggesting they share a long coevolutionary history. Treatment with the antibiotic tetracycline inhibits normal reproduction and development of filarial nematodes that harbor *Wolbachia*, suggesting that *Wolbachia* are necessary to the nematode ([Langworthy et al. 2000](#)). That is not true for arthropods, where it seems a great deal of horizontal transfer has occurred.

*Wolbachia* have been implicated as both the cause of alterations in sex ratio (resulting in thelytoky and male killing), to be discussed further in the Chapter 10 on sex determination, and **cytoplasmic incompatibility** in arthropods. Some *Wolbachia* improve fertility or vigor, whereas others seem to decrease these traits in their hosts. The molecular mechanism(s) by which reproductive incompatibility are induced by *Wolbachia* are hypothesized to be due to *Wolbachia*'s

ability to modify sperm. This hypothesis suggests that paternal chromosomes are modified during spermatogenesis by *Wolbachia* and this modification is “rescued” in eggs of females infected with the same strain of *Wolbachia* during fertilization. If, however, the female is *not* infected with *Wolbachia* and mates with an infected male or male infected with a different strain of *Wolbachia*, then the embryos die (Figure 4.2). Some *Wolbachia* strains have been identified that fail to modify sperm but can rescue the modification in eggs of other *Wolbachia* strains (Bourtzis et al. 1998).

Cytoplasmic incompatibility caused by *Wolbachia* may be partial or complete. Sometimes incompatibility is found in both reciprocal crosses (A × B and B × A, **bidirectional incompatibility**), perhaps due to the presence of different strains of *Wolbachia* in each population. Incompatibility is more often found in one reciprocal cross (AxB or BxA, **unidirectional incompatibility**). Cytoplasmic incompatibility typically is incomplete (<100%), perhaps due to inefficient transfer of *Wolbachia* to all progeny or to differences in the titer of *Wolbachia*. Such differences in titer could occur naturally if the infected insects encounter antibiotics in their environment or if they experience high temperatures (typically >30°C) (Snook et al. 2000). Tram and Sullivan (2002) found that the cytoplasmic incompatibility in *Nasonia vitripennis* was due to a delay in the nuclear envelope breakdown in the male pronucleus; this delay resulted in an asynchrony between the male and female



**Figure 4.2** A) Cytoplasmic incompatibility due to *Wolbachia* between different individuals or populations can result in a failure to produce progeny when a *Wolbachia*-infected male mates with an uninfected female. B) If isolated populations of a species become infected with different types of *Wolbachia* ( $W_A$  and  $W_B$ ), then these populations could become reproductively isolated if they later come into contact.

pronuclei and loss of paternal chromosomes at the first mitosis, resulting in the production of haploid males derived from the maternal chromosome set.

Some insects seem to have *Wolbachia* only in their germ-line tissues (ovaries and testes), whereas others have *Wolbachia* in somatic tissues as well (Dobson et al. 1999). Large numbers of *Wolbachia* have been found in ovaries and testes of populations with cytoplasmic incompatibilities. Incompatible strains have been converted to compatible by treating the colonies with heat or antibiotics, which eliminates or greatly reduces the *Wolbachia* population.

*Wolbachia* can be transferred to new populations experimentally by micro-injecting infected egg cytoplasm into uninfected eggs. Transinfected strains of *D. simulans* and *D. melanogaster* with high titers of *Wolbachia* exhibited cytoplasmic incompatibilities at high levels, but those with low titers exhibited low levels of incompatibility, suggesting that a threshold level of infection is required and that host factors may determine the density of the *Wolbachia* in the host (Boyle et al. 1993).

*Wolbachia* have been identified in many species of parasitic Hymenoptera, including species in the Aphelinidae, Encyrtidae, Eulophidae, Pteromalidae, Torymidae, Trichogrammatidae, Cynipidae, Eucoilidae, Braconidae, Ichneumonidae, and Proctotrupidae and in three dipteran parasitoids (Tachinidae) (Cook and Butcher 1999). *Wolbachia* may cause both cytoplasmic incompatibility and induction of parthenogenesis in these parasitoids. Many hymenopteran parasitoids have both bisexual (arrhenotokous) and unisexual strains consisting only of females (thelytoky), probably due to the presence of *Wolbachia*.

Phylogenetic analysis suggests that the *Wolbachia* common ancestor evolved between 80 and 100 million years ago (O'Neill et al. 1992), whereas the arthropod common ancestor occurred at least 200 million years earlier. Thus, *Wolbachia* probably have invaded arthropods through **horizontal transmission** (O'Neill et al. 1992, Heath et al. 1999, Jeyaprakash and Hoy 2000). Some arthropods have been found to have double, triple, or even greater numbers of infections with different strains of *Wolbachia*. The effects of these multiple infections usually are unknown.

Several methods have been proposed as mechanisms for horizontal transfer, including the movement of *Wolbachia* from host arthropods to their parasitoids. Heath et al. (1999) experimentally transferred *Wolbachia* from *Drosophila simulans* to a novel host, its endoparasitoid *Leptopilina boulardi*, 0.711% of the time. The *Wolbachia* infection rate diminished during subsequent vertical transmission to the F<sub>2</sub> and F<sub>3</sub> generations, perhaps due to poor maternal transmission (unstable vertical transmission).

Experimental microinjection (artificial horizontal transfer) of *Wolbachia* from the parasitoid *Muscidifurax uniraptor* into its host *D. simulans* resulted in a temporary infection, but no specific phenotypic effects were observed (van Meer and Stouthamer 1999). These results suggest that host–symbiont interactions are important for successful establishment of a *Wolbachia* infection in a new host, although it is clear that *Wolbachia* has successfully bridged large phylogenetic distances in its horizontal movements over evolutionary time.

The availability of PCR primers for *Wolbachia* genes revolutionized the study of the distribution and evolution of *Wolbachia*. The *Wolbachia* genome project further revolutionized such studies. Based on a phylogeny developed using the *ftsZ* gene, *Wolbachia* infecting arthropods have been divided into groups A and B, and these groups are estimated to have diverged from each other 58 to 67 million years ago (Werren et al. 1995). Phylogenies based on *wsp* gene sequences have yielded more groups, indicating considerable genetic variation exists (Zhou et al. 1998, van Meer and Stouthamer 1999, Jeyaprakash and Hoy 2000). It is unclear whether these *Wolbachia* groups are strains or species because the definition of a bacterial species is complex.

*Wolbachia* may have a role in speciation of arthropods by generating reproductive isolation (Rokas 2000), although some argue that *Wolbachia*'s role(s) remain unproved (Hurst and Schilthuizen 1998). Typically, *Wolbachia* cause unidirectional cytoplasmic incompatibility when a *Wolbachia*-infected male mates with an uninfected female (Figure 4.2A). The eggs or embryos of such matings die, resulting in a fitness cost to uninfected females that, over time, results in the infected cytotype becoming fixed in the population. A problem with the “speciation hypothesis” is that *Wolbachia* are not transmitted 100% of the time from a female to her progeny, so progeny will be produced that are compatible. Also, incompatibility is not completely expressed (incomplete penetrance of the trait) when infected males and uninfected females mate in natural populations (perhaps due to differences in the titer of the *Wolbachia* within individuals). Furthermore, selection on both the host and *Wolbachia* may favor reduced penetrance of the incompatibility phenotype or loss of *Wolbachia*, leading to a situation in which there is no gene flow to some gene flow (Hurst and Schilthuizen 1998). Thus, unidirectional incompatibility caused by *Wolbachia* may be insufficient to cause the reproductive barriers that could lead to speciation. Additional factors, such as **hybrid sterility** (sterility of the hybrid when crossed with either of the parental species) and **hybrid breakdown** (inviability or sterility of progeny resulting from a backcross of hybrid progeny with either of the parental species) may be necessary (Shoemaker et al. 1999). Thus, *Wolbachia* may enhance the speciation rate by acting in conjunction with behavioral isolation.

A second speciation mechanism associated with *Wolbachia* may be by the induction of thelytoky (reproduction by females only), as has been found in many hymenopteran parasitoids, such as *Encarsia formosa*. Populations of *Encarsia* no longer have males, so that populations essentially become clonal and over time could differentiate genetically (Cook and Butcher 1999).

A third potential *Wolbachia* speciation mechanism is by bidirectional incompatibility; if a population is infected with two different strains of *Wolbachia* that are incompatible with each other, then the incompatibility could act as a postzygotic reproductive barrier, as has been suggested for the species complex of *Nasonia* (Hymenoptera) (Figure 4.2B). Thus, how *Wolbachia* are maintained in populations has considerable theoretical and practical importance. *Wolbachia* have been proposed as mechanisms for driving genes into populations in genetic manipulation projects for improved pest control (for additional discussion of this topic, see Chapter 14). Gazla and Carracedo (2011) reported that females of *D. melanogaster* and *D. simulans* could discriminate between males with and without *Wolbachia* and preferred to mate with males that had the same infection status; this preference would influence sexual isolation and also could play a role in speciation.

The interest in the biology and evolution of *Wolbachia*, with its fascinating effects on reproductive isolation (thus potentially having effects on speciation), sex ratio, feminization, and male killing, led to the development of a *Wolbachia* genome project (Bandi et al. 1999). Genome sizes for six different *Wolbachia* strains were determined by pulsed-field gel electrophoresis (Sun et al. 2001). The *Wolbachia* genomes are circular and range from 0.95 to 1.66 Mb, which is considerably smaller than the genomes of free-living bacteria such as *E. coli* (4.7 Mb).

Despite the wealth of information obtained about *Wolbachia* within the past few years, our understanding of the role of *Wolbachia* in arthropod biology and evolution probably remains fragmentary. For example, some *Wolbachia* in arthropods were shown to contain bacteriophages named WO (Masui et al. 1999, 2000). A phylogenetic analysis of different WOs from several *Wolbachia* strains yielded a tree that was not congruent with the phylogeny of the *Wolbachia*, suggesting that the phages were active and horizontally transmitted among the various *Wolbachia*. Masui et al. (2000) speculated that, because all *Wolbachia* strains they examined contain WO, the phage might have been associated with *Wolbachia* for a very long time, conferring some benefit to its microbial hosts.

*Wolbachia* strains seem to have a very wide range of effects on their hosts, ranging from causing male killing, cytoplasmic incompatibility, modification of immune responses, parthenogenesis, providing nutritional benefits, and tolerance to heat shock (Box 4.2). *Wolbachia* strains even may block transmission of disease-causing agents in mosquitoes and *D. melanogaster* (Box 4.2).

**Box 4.2 Selected Examples of the Different Roles *Wolbachia* Can Play in the Biology of Arthropods**

Effects	References
<b>Block transmission of disease-causing agents</b>	
<i>Wolbachia</i> strain wMel injected into <i>Aedes albopictus</i> causes cytoplasmic incompatibility and blocks dengue transmission in mosquitoes naturally infected with two different strains of <i>Wolbachia</i> (wAlbA and wAlbB).	Blagrove et al. 2012
<i>Wolbachia</i> strain wMel can delay accumulation of RNA (but not DNA) viruses in <i>Drosophila melanogaster</i> .	Teixeira et al. 2008
Unstable (somatic) infections with wMelPop and wAlbB strains are virulent and inhibit the malaria parasite <i>Plasmodium falciparum</i> in <i>Anopheles gambiae</i> .	Hughes et al. 2011
<b>Cytoplasmic incompatibility (CI)</b>	
Tsetse fly <i>Glossina morsitans</i> males infected with <i>Wolbachia</i> that mate with females lacking <i>Wolbachia</i> exhibit CI.	Alam et al. 2011
Males of <i>Epeorus kuehniella</i> infected with <i>Wolbachia</i> transmit fewer fertile sperm than uninfected males.	Lewis et al. 2011
<i>D. melanogaster</i> and <i>D. simulans</i> ; <i>Wolbachia</i> also causes assortative mating by females.	Gazla and Carracedo 2011
<b>Male killing</b>	
<i>Wolbachia</i> strain wSca infecting the moth <i>Ostrinia scapulalis</i> causes male killing. Females died when cured of the infection. wSca carries a genetic factor that feminizes degraded. A mismatch between the genetic and phenotypic the male and the sex-determining system of the host is sex causes sex-specific death.	Sugimoto and Ishikawa 2012
<b>Modification of immune responses</b>	
<i>Wolbachia</i> in <i>Aedes aegypti</i> reduced transmission of dengue virus and increased longevity in the mosquitoes.	Bian et al. 2010
<i>Wolbachia</i> in <i>Anopheles gambiae</i> inhibit the immune system.	Hughes et al. 2011
<i>Wolbachia</i> (wMelPop strain) injected into <i>Anopheles gambiae</i> stimulated immune gene expression and inhibited <i>Plasmodium</i> development.	Kambris et al. 2010
<i>Wolbachia</i> strain wAlbB injected into <i>Aedes aegypti</i> (Waco strain) had a strong resistance to dengue virus, <i>Plasmodium</i> , and filarial nematodes. Resistance is due to activation of the Toll pathway that is responsible for activation of antimicrobial peptides (defensins and cecropins).	Pan et al. 2012
<b>Parthenogenesis</b>	
<i>Wolbachia</i> in the mite genus <i>Bryobia</i> causes all female production due to apomixis (development without fertilization of the egg).	Weeks and Breeuwer 2001
<b>Nutritional mutualism</b>	
Sequencing of different <i>Wolbachia</i> strains indicated they may provide nutritional benefits. <i>Drosophila melanogaster</i> reared on iron-restricted or -overloaded diets had improved fecundity if infected with <i>Wolbachia</i> strain wMel.	Brownlie et al. 2009
<b>Temperature effects</b>	
Males of <i>Drosophila simulans</i> infected with <i>Wolbachia</i> have higher levels of heat-shock proteins, which result in more progeny than uninfected males can produce if treated to a heat shock.	Feder et al. 1999

**Box 4.3 Selected Examples of the Effects *Cardinium* Can Exert on the Biology of Arthropods**

Effects	References
<b>Cytoplasmic incompatibility (CI)</b>	
The spider mites <i>Eotetranychus suginamensis</i> and <i>Bryobia sarothamni</i> and the parasitoid <i>Encarsia pergandiella</i> exhibit incompatibility.	Gotoh et al. 2007 Ros and Breeuwer 2009 Hunter et al. 2003 Wu and Hoy 2012
The predatory mite <i>Metaseiulus occidentalis</i> exhibits CI when uninfected females mate with infected males.	Hunter et al. 2003
<i>Encarsia pergandiella</i> exhibits CI due to <i>Cardinium</i> infection.	
<b>Induces thelytoky</b>	
<i>Encarsia</i> parasitoid wasps are thelytokous.	Zchori-Fein et al. 2004
Mites in the genus <i>Brevipalpus</i> are thelytokous and haploid due to infection with <i>Cardinium</i> ; haploid eggs are feminized.	Weeks et al. 2001 Groot and Breeuwer 2006 Chigira and Miura 2005
<b>Host-selection behavior modified</b>	
<i>Encarsia</i> parasitoid wasps are thelytokous due to <i>Cardinium</i> . Curing of females changed oviposition behavior compared with infected females.	Zchori-Fein et al. 2001 Kenyon and Hunter 2007

*Cardinium* bacteria in the phylum Bacteroidetes were relatively recently discovered to infect arthropods (Zchori-Fein et al. 2001, Hunter et al. 2003, Zchori-Fein et al. 2004). Their effects on their hosts seem to be similar to that of *Wolbachia* in the few cases in which this has been determined (Box 4.3), causing cytoplasmic incompatibility, thelytoky, and feminization of males. *Cardinium* has been found in species of Coleoptera, Collembola, Anoplura, Hymenoptera, Hemiptera, and Diptera, and in spiders (Araneae), ticks (Ixodidae), spider mites (Acari: Tetranychidae), and predatory mites (Acari: Phytoseiidae) (Zchori-Fein and Perlman 2004, Enigl and Schausberger 2007, Nakamura et al. 2009, Chang et al. 2010, Wu and Hoy 2012). Surveys of arthropods for *Cardinium* suggest between 4 and 7% of arthropods contain *Cardinium*, although the incidence in arachnids is greater (22–33%) (Duron et al. 2008, Martin and Goodacre 2009).

Kenyon and Hunter (2007) found that *Cardinium* infection of *Encarsia pergandiella* affected the choice thelytokous females made as to which host in which to oviposit. Sexual forms of this species are autoparasitoids, i.e., they deposit female eggs in whitefly nymphs and males are deposited on conspecific parasitoids developing within the whitefly. Thelytokous populations of *E. pergandiella* were cured of their *Cardinium* with antibiotics, and their behavior was compared with that of the sexual form. Progeny of antibiotic-treated thelytokous wasps

reverted to the behavior of unmated sexual wasps, laying eggs in hosts appropriate for male eggs. *Cardinium*-infected thelytokous wasps distributed their eggs almost equally between host types, similarly to mated sexual females.

Nakamura et al. (2011) compared the effects on the immune system of *Bombyx mori* by using *Wolbachia*- and *Cardinium*-infected cells using a silk moth microarray. *Wolbachia* infection did not change gene expression or induce or suppress immune responses but *Cardinium* infection induced many immune-related genes. The differences were attributed to the fact that the two micro-organisms have different cell wall structure, with *Wolbachia* lacking genes encoding lipopolysaccharide components and other cell wall components.

Much remains to be learned about the role(s) *Cardinium* plays in the biology of their arthropod hosts.

#### 4.12.3 Polydnaviruses in Parasitoids

A particularly interesting example of an intimate relationship between insects and symbionts is illustrated by the relationship between polydnaviruses and parasitoids. The **polydnaviruses** are found only in the Braconidae and Ichneumonidae among the parasitic Hymenoptera (Krell 1991, Fleming 1992, Stoltz and Whitfield 1992). Polydnaviruses are symbiotic proviruses that have double-stranded circular DNA genomes; they are literally “poly DNA-viruses,” having segmented genomes composed of several circular DNA molecules. For example, the viral genome within *Campoletis sonorensis* consists of 28 DNA molecules ranging from ≈5.5 to 21 kb, with the total genome size ≈150 kb.

Polydnaviruses ensure that some species of braconids and ichneumonids (=parasitoids) are able to successfully parasitize their lepidopteran hosts. At least 50 species of parasitoids are known to contain polydnaviruses (Stoltz and Whitfield 1992) and >30,000 species are thought to carry them (Shelby and Webb 1999). Genera of parasitoids containing polydnaviruses seem to be more speciose and have broader host ranges than sibling groups lacking them, suggesting that the viruses contribute to the evolutionary success of their parasitoid hosts (Shelby and Webb 1999). The two polydnaviral groups, Ichnoviridae and Brachoviridae, are phylogenetically and morphologically distinct and use different mechanisms to inhibit the immunity and development of the lepidopteran hosts (Webb 1998). The association between braconid parasitoids and their viruses seems to have lasted at least 60 million years (Whitfield 1997).

Polydnaviruses replicate only in braconid or ichneumonid wasp ovaries and are secreted into the oviducts from where, during oviposition, they are injected

into host lepidopteran larvae. The viruses seem to be vertically transmitted and integrated into the chromosome of the wasp (Fleming and Summers 1991). Each wasp species seems to carry a polydnavirus characteristic of that species. If one species within a genus carries a polydnavirus, all are likely to do so (Stoltz and Whitfield 1992).

Insects possess immune mechanisms that protect them from microorganisms, other invertebrates, and abiotic material (Hultmark 1993, Gillespie et al. 1997). Protection occurs through constitutive factors or by inducible humoral and cellular responses. Many behavioral, morphological, nutritional, and endocrine factors determine whether the interactions between a host and a parasitoid will lead to development of the parasitoid or to its destruction (Fleming 1992). Polydnaviruses alter the host insect's neuroendocrine and immune responses, preventing encapsulation of parasitoid eggs and larvae by host hemocytes, and influence development of the host to benefit the parasitoid (Webb and Cui 1998, Shelby and Webb 1999). The virus replicates asymptotically in the parasitoid, but it causes a pathogenic infection in the lepidopteran host (Webb and Cui 1998). The virus alone can induce altered immune responses in some hosts, but in other hosts the venom injected by the wasp also must be present for the full effect of the virus to occur. Parasitoid wasps seem to benefit significantly from the polydnaviruses. The virus also clearly benefits if the parasitoid is able to reproduce, because polydnaviruses are known to replicate only within their host parasitoids.

Sequencing of genomes of Braco- and Ichnoviruses indicate that the relationship between viruses and host parasitoids has existed for at least 70 million years (Dupuy et al. 2006). One hypothesis is that an ancestral Ichneumonoidea genome contained genes that were transmitted to ancestral Braconidae and Ichneumonidae. Dupuy et al. (2006) suggested the two wasp families each captured mobile elements or viruses whose replication and encapsidation machinery were used to form the basis of the ancestral provirus. The integrated viruses then captured wasp genes that were beneficial for a parasitic life cycle. Convergent evolution led to the capture of related genes and the loss of viral structural genes, with gene flow taking place between the wasp genome and the virus leading to evolved proviruses that eventually developed in the bracoonid and ichneumonid wasp lineages that contain the polydnaviruses.

The polydnavirus-parasitoid-lepidopteran system provides an unusual example of an obligate mutualistic association between a virus and a parasitoid that functions to the detriment of the lepidopteran. The origin of polydnaviruses is unknown, as is how they became established in the parasitoid genome.

Beckage (1998) speculated that polydnaviruses have potential value in agricultural pest-management programs if genetically engineered pathogens (viruses, bacteria, fungi) containing polydnavirus genes could produce products that suppressed the target pest's immune system. Alternatively, genetically engineered parasitoids could be developed that exhibit a modified host range, making them more effective in controlling pests.

#### 4.12.4 Gut Symbionts in Arthropods

Some gut symbionts are obligatory, whereas others are facultative. Gut symbionts often provide key nutritional benefits to their hosts and, in return, receive shelter and nutrition. Insects may contain complex and diverse societies of microbes in their guts; yet, much remains to be learned about how these microbes shape the physiology of their hosts (Cazemier et al. 1997, Kaufman et al. 2000, Konig and Varma 2006). The primary habitat for microorganisms associated with insect digestive tracts is the hindgut.

##### 4.12.4.1 Termite and Cockroach Symbionts

In termites, the symbionts in the hindgut provide nitrogen (N) and vitamins and degrade cellulose and hemicellulose. The microbial components of the termite gut are extremely complex, but molecular tools are improving our ability to resolve the taxonomy and the complex relationships among termite gut symbionts.

The hindguts of termites can be compared to small bioreactors where wood and litter is degraded, with the help of symbiotic microorganisms, to provide nutrients. The hindgut of termites is a structured environment with distinct microhabitats (Brune and Friedrich 2000). The dense gut microbiota includes organisms from the Bacteria, Archaea, Eukaryotes, Protozoa, and yeasts. These diverse organisms do not occur randomly within the gut but have specific locations: they may be suspended in the gut contents, located within or on the surface of flagellates, or attached to the gut wall. The identity, exact number, and location of most are inadequately known because they cannot be cultured. Molecular tools are providing significant new information. For example, the microbiota of termites includes spirochaetes, which account for as many as 50% of the organisms present in some termite species. Spirochaetes are a distinct phylum within the bacterial domain. One molecular analysis of spirochaetes in the termite *Reticulitermes flavipes* suggested there are at least 21 previously unknown species of *Treponema* (Lilburn et al. 1999). The authors concluded that the long-recognized and striking morphological diversity of termite gut spirochaetes is paralleled by their genetic diversity and could reflect substantial physiological diversity (Lilburn et al. 1999).

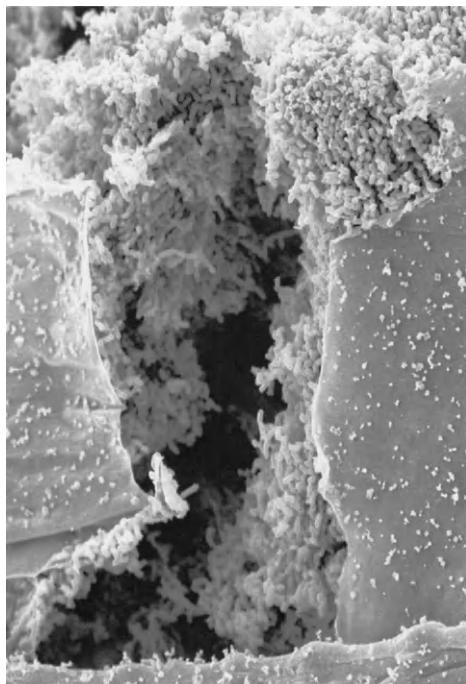
Husseeneder (2010) provided a review of the roles symbionts play in termite nutrition in subterranean termites, especially *Reticulitermes* and *Coptotermes*. Within the subterranean termites, the protozoa are involved in cellulose or hemicellulose digestion and acetogenesis, while some bacteria produce vitamins, some bacteria are nitrogen-fixing and uric-acid degraders, others are cellulolytic, some lactic acid bacteria are fermentative, others are acetogenic CO<sub>2</sub>-reducing, whereas the Archaea are methanogenic. In addition, yeasts and fungi are thought to be important, perhaps as a direct source of food and by modifying wood. Husseeneder (2010) pointed out that the mutualistic relationship of termites with their symbionts may be one of the important factors in allowing termites to have a social life style because each worker termite must acquire an inoculum of symbionts from parents or nestmates after hatching and after each molt so that parental care, group living, and overlapping generations are needed, setting the stage for termites to become social.

Warnecke et al. (2007) conducted a metagenomic and functional analysis of hindgut microorganisms in a wood-feeding higher termite (which does not have hindgut flagellate protozoa, which produce cellulases and hemicellulases in the lower termites). DNA was obtained from the largest hindgut compartment that contained bacteria and then analyzed by cloning and sequencing all DNA. Because the bacterial community was large, only genes could be assembled (rather than whole microbial genomes). The genes analyzed are of interest to biotechnology because some may be used to produce biofuels from wood. Termite symbionts also are of interest because termites are important in carbon turnover in the environment.

Omnivorous cockroaches also have gut microbial communities, but the associations are less interdependent than those of termites. As in termites, the gut microbial communities in cockroaches anaerobically degrade plant polymers and include hydrogen-consuming bacteria, especially methanogens. The densities of these microorganisms can be enormous; for example  $5 \times 10^{12}$  bacteria/ml were found in the hindgut of the cockroach *Periplaneta americana* (Cazemier et al. 1997).

#### 4.12.4.2 Rhagoletis Symbionts

A less intimate relationship between microbial genomes and insects is represented by the relationship between *Enterobacter agglomerans*, found in the gut of the apple maggot *Rhagoletis pomonella* (Lauzon et al. 2000). Enterobacteriaceae are the most common microorganisms associated with the apple maggot in the gut and female reproductive organs, and it seems the flies use the bacteria for some vital function(s) (Lauzon et al. 1998). In addition to *E. agglomerans*, *Klebsiella oxytoca* is found in the gut of *R. pomonella*, and both



**Figure 4.3** This scanning electron micrograph shows a biofilm of *Enterobacter agglomerans* and *Klebsiella* species in the midgut of the apple maggot *Rhagoletis pomonella*. (Photo provided by C. R. Lauzon.)

are most abundant in the esophageal bulb, crop, and midgut. These bacteria are found on host plants and other substrates in the environment. It seems that the bacteria provide usable nitrogen for *R. pomonella* and other tephritids by degrading purines and purine derivatives, making them facultative symbionts. The relationship between the *Enterobacter* and *Klebsiella* species is probably complex. Figure 4.3 illustrates the “biofilm” of *E. agglomerans* and *Klebsiella* in an adult *R. pomonella* midgut. A **biofilm** is a complex, structured community of microbes attached to surfaces. Microbial biofilms function as a cooperative consortium in a complex and coordinated manner (Davey and O’Toole 2000). The role of this biofilm in *R. pomonella* is under study (Lauzon et al. 1998).

#### 4.12.5 Symbiosis in Fungus-Growing Attine Ants

New World attine leaf-cutter ants live in the tropics, where they carry small leaf fragments back to their nests. The leaves are used as fertilizer to cultivate a fungus, which is their food. The fungi produce specialized structures that are consumed by the ants. The worker ants carry out complex behaviors to maintain the fungus gardens, as well as caring for the brood. Ant queens initiate new

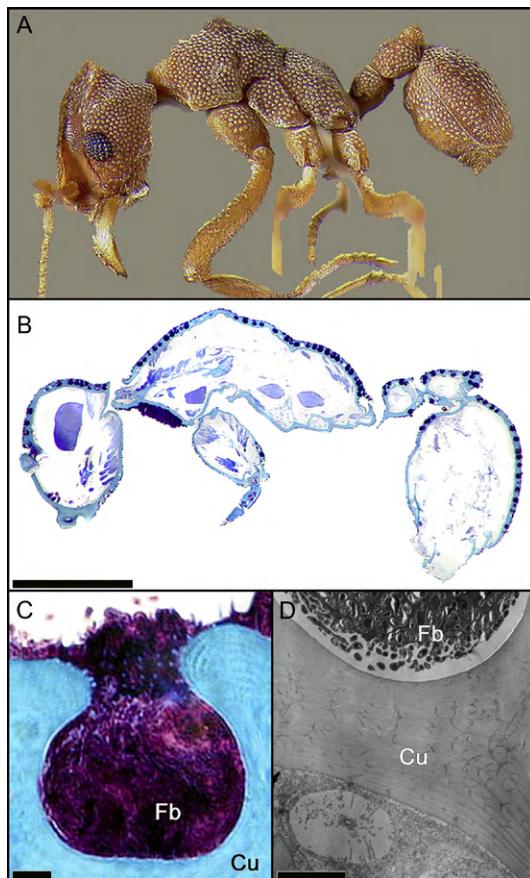
colonies by carrying a small amount of the fungus within a pouch in her oral cavity to a new nest site. She uses the fungus as the inoculum for a new garden (Caldera et al. 2009).

The ant–fungus symbiosis is complex: the ants have associations with at least four microorganisms: 1) the fungi they grow as food in gardens, 2) a parasitic microfungus (*Escovopsis*) that attacks the food fungus, 3) an actinobacterium (*Pseudonocardia*) that produces antibiotics that inhibit the parasitic *Escovopsis*, and 4) a black yeast that parasitizes the ant–actinobacteria mutualism (Caldera et al. 2009).

The symbiosis between the ants, their fungal food and the *Escovopsis* parasite originated ≈50–65 million years ago (Currie 2001, Currie et al. 2006, Caldera et al. 2009, Poulsen and Currie 2010). Ant workers culture the fungus garden by grooming alien fungal spores from the garden or by removing infected leaf material (grooming and weeding). The ants have paired metapleural glands that produce antimicrobial compounds that help protect ants from insect-pathogenic fungi. The ants pass their forelegs across the surface of the gland and then pass their legs through their mouthparts during garden maintenance. It is thought that metapleural gland grooming behavior allows application of chemicals directly to fungus gardens that are infested with parasitic fungi. In addition, the actinobacterium produces antibiotics that are effective against *Escovopsis*. For a long time, it was thought that the ants had a waxy exudate on its body, but a closer examination revealed these “waxes” are aggregations of the actinobacteria (*Pseudonocardia*) growing in elaborate cuticular crypts in the exoskeleton and underneath these are specialized glands that produce secretions used by the actinobacteria (Figure 4.4, Currie et al. 2006).

The fourth symbiont, black yeast (*Phialophora*), grows on the same locations on the ant cuticle as the actinobacteria and parasitizes the system by acquiring nutrients from the bacteria, which could reduce the ability of the bacteria to suppress *Escovopsis* growth.

Are there any more microorganisms associated with attine leaf-cutter ants? Future studies may identify other symbioses, according to Caldera et al. (2009). Pinto-Tomas et al. (2009) demonstrated that fungus gardens of leaf-cutter ants have a higher level of nitrogen than is present in the fresh leaves used to fertilize them. Because ant workers had a higher level of N than was present in the leaf material, it seems that N enrichment occurs as the ants process the plant material. No nitrogen-fixing genes are present in the fungi or ants, but they were found in *Klebsiella* bacteria isolated from the fungus gardens of leaf-cutter colonies. *Klebsiella* are known to fix N<sub>2</sub>, as are bacteria in the genus



**Figure 4.4** A) and B) Photographs of the fungus-growing ant *Cyphomyrmex longiscapus* showing the foveae openings in the exoskeleton covering most of the body. These foveae contain bacteria (*Pseudonocardia*) that produce antibiotics. C) This light micrograph illustrates a single fovea showing many bacteria are present. D) This transmission electron micrograph shows a single glandular cell with the *Pseudonocardia* bacteria within the crypt. (Photos provided by Cameron Currie.)

*Pantoea* (although at a lower level than *Klebsiella*), which also were isolated consistently. [Pinto-Tomas et al. \(2009\)](#) concluded the following:

*"The ecological success of leaf-cutter ants is derived, in large part, from the combined ability of the ants to break down antifungal barriers and of the fungus garden to neutralize plant anti-insect toxins. Consequently, the leaf-cutters are able to use a high diversity of plant families, in contrast to most herbivorous insects....the reliance on leaf material means that leaf-cutter ant colonies are potentially N-limited. Our work shows that these ants can potentially overcome such limitation through symbiotic associations with N<sub>2</sub>-fixing bacteria."*

Microbial symbionts may be important in the biology and behavior of these ants in other ways. For example, cuticular hydrocarbon profiles, used by social insects to recognize nest mates, are determined by both environmental and genetic factors. Richard et al. (2007) compared the cuticular hydrocarbon profiles of 18 colonies of *Acromyrmex* and compared the variability of the 47 compounds. Colony-specific profiles were distinct and significantly different between two *Acromyrmex* species. Interestingly, Richard et al. (2007) showed that workers previously exposed to and fed with the fungal symbiont of another colony “are met with less aggression when they are later introduced into that colony. It appears, therefore, that fungus gardens are an independent and significant source of chemical compounds, potentially contributing a richer and more abundant blend of recognition cues to the colony “gestalt” than the innate chemical profile of the ants alone.” If adjacent colonies contain different strains of fungus, it would allow workers to discriminate readily between nest-mates and nonnestmates.

#### **4.12.6 Southern Pine Beetle Symbionts**

Scott et al. (2008) found that southern pine beetles, *Dendroctonus frontalis*, have a mutualism somewhat similar to attine ants that involves two fungi and an actinomycete bacterium. The adult beetles carry a beneficial fungus (*Entomocorticium*) in a specialized storage compartment called a mycangium. Females excavate galleries within the inner bark and phloem of pine trees to oviposit. At the same time, they inoculate the galleries with the beneficial fungus that provides food for their larvae. However, a fungus that can out-compete *Entomocorticium* can affect the mutualistic relationship between *Entomocorticium* and the beetle larvae. Scott et al. (2008) discovered an actinomycete related to a *Streptomyces* bacterium that produces a diffusible product that inhibits the antagonistic fungus but that only slightly affects the beneficial fungus. The antifungal molecule was named mycangimycin because the bacteria are found in the mycangia as well as in the galleries. Scott et al. (2008) suggested the use of antibiotic-producing actinomycetes could be “a common method for maintaining beneficial microbes.”

#### **4.12.7 Aphid Symbionts**

Aphids not only have a primary endosymbiont that is obligatory, but also can harbor up to seven secondary (facultative) symbionts that can affect several aspects of the biology of their host (Oliver et al. 2010, Simon et al. 2011). Perhaps the best studied is *Buchnera aphidicola*, located in 60–80 very large cells called bacteriocytes (Baumann et al. 1997, Douglas 1998, Moran and Baumann

2000, Hansen and Moran 2011). The complete genome of *Buchnera* has been sequenced (Shigenobu et al. 2000). *Buchnera* is found in huge cells (bacterioocytes) in most of the 4400 aphid species, supplying the aphids with essential amino acids. In return, *Buchnera* is given a stable and nutrient-rich environment. Aphids become sterile or die if *Buchnera* are eliminated. The aphid–*Buchnera* relationship has been stable for up to 250 million years, and about 9% of the *Buchnera* genome is devoted to producing essential amino acids for the aphid. Genes for nonessential amino acids are absent in *Buchnera*, and this symbiont depends on its aphid host for these amino acids, making *Buchnera* and the aphid codependent (Hansen and Moran 2011).

Analyses of different aphid species and their *Buchnera* indicate that vertical transmission of the symbionts has occurred from the time of the common ancestor of aphids, ≈150–250 million years ago (Moran and Baumann 2000). There is “phylogenetic congruence with hosts, implying co-speciation” and there is no evidence of horizontal transfer, even within a single aphid species (Moran and Baumann 2000). In many *Buchnera* lineages, genes involved in tryptophan and leucine biosynthesis are on plasmids rather than in the *Buchnera* genome. The location of these genes on plasmids allows increased gene expression and thus increased benefit to their aphid hosts. The copy number of plasmids in *Buchnera* seems to vary in different aphid lineages, perhaps reflecting coordinated, adaptive adjustment to the nutritional needs of the different aphids. The genome of *Buchnera* is unusual when compared to the free-living bacterium *E. coli*, with a very AT-biased (≈28% GC) genome. Second, DNA sequences evolve faster in *Buchnera* than in free-living relatives. Third, the genome of *Buchnera* (from *A. pisum*) is reduced to ≈650 kb, about one-seventh of the genome size of *E. coli*. *Buchnera* seems to contain only a subset of ≈600 of the 4500 genes present in an *E. coli*-like ancestor.

Each *Buchnera* contains 50–200 chromosomes, with the number of copies varying with the life-cycle stage of the host. This chromosome amplification may be used to vary the contribution of the symbiont to its host’s nutrition (Komaki and Ishikawa 1999, 2000). The amplification of chromosome copy number to 200 copies/cell is very unusual in the microbial world; *E. coli* typically has one or two chromosomes per cell. The dramatic reduction in genome size of *Buchnera* and the extraordinary increase in chromosome copy number make this intracellular symbiont resemble eukaryotic cell organelles such as mitochondria and chloroplasts, which are evolutionary descendants of symbiotic bacteria (Komaki and Ishikawa 2000). *Buchnera* resemble these organelles also in that they are transmitted maternally. However, *Buchnera* is unlike mitochondria in that no genes have been transferred from *Buchnera* to its host (Nikoh et al. 2010).

Studies of the pea aphid, *Acyrtosiphon pisum*, show that one or more species of facultative symbionts are present in different populations. These may protect the wasps from entomopathogenic fungi and parasitoid wasps, modify their tolerance to heat, modify their behavioral responses to parasitoids, and affect host plant suitability. One facultative symbiont, *Hamiltonella defensa*, is maternally transmitted and protects against parasitoids. Because the symbiont provides protection, aphids infected with *H. defensa* "exhibited reduced aggressiveness and escape reactions compared with uninfected aphids" (Dion et al. 2011). Defensive behaviors may be costly to the aphid by reducing feeding opportunities or reducing survival, so a reduction in such behavior could confer a fitness advantage (Dion et al. 2011).

Tsuchida et al. (2010) report that a facultative symbiont (*Rickettsiella*) found in some populations changes the aphid's body color from red to green by increasing the amounts of blue-green polycyclic quinones, which could reduce predation by lady beetles. *Serratia symbiotica* is another secondary symbiont of pea aphids, providing defense against environmental heat stress and is related to free-living bacteria that can be pathogens to animals (Burke and Moran 2011). Although *S. symbiotica* is considered to be a relatively recently evolved symbiont, its genome has characteristics of older symbionts, including elevated rates of evolution and reduction in genome size. It seems to be dependent upon its host for some essential nutrients (Burke and Moran 2011). Simon et al. (2011) found that a *Spiroplasma* infection of some populations of the pea aphid affected the time at which females first reproduced, their total fecundity, longevity, and the proportions of asexual females and males. Male killing caused by *Spiroplasma* occurred in some pea aphid populations, caused by mortality at early nymphal stages of the sexual phase. Another endosymbiont, *Regiella insecticola*, is found in the pea aphid (and other aphids) (Oliver et al. 2010). Tsuchida et al. (2011) showed that *R. insecticola* allows the pea aphid to attack clover and reproduce at a higher rate on clover, broadening the host-plant range of the pea aphid in Japan. Experimental transfer of *Regiella* from the pea aphid to the vetch aphid (*Megoura criassicauda*) allowed the vetch aphid to grow and reproduce on clover, a host that was not used previously.

#### 4.12.8 Tsetse Fly Symbionts

Tsetse flies (Glossinidae) are vectors of African sleeping sickness disease in humans and animals. Microorganisms associated with these flies, which are blood feeders, are responsible for nutrients not found in their restricted diet. Different microorganisms have been found in the midgut, hemolymph, fat body, and ovaries. Until molecular techniques were used, their taxonomic status was

unresolved (Aksoy 2000). Now, we know that at least three different microorganisms are present: the primary (P) symbiont *Wigglesworthia glossinidia* is an intracellular symbiont residing in specialized epithelial cells that form a special U-shaped organ (bacteriome) in the anterior gut. The secondary gut symbiont *Sodalis glossinidius* is present in midgut cells. The third, *Wolbachia*, is found in reproductive tissues. Tsetse females are viviparous, retaining each egg within her uterus where it hatches. The larva matures there and is born as a fully developed third-instar larva. During its intrauterine life, the larva receives nutrients and both of the gut symbionts from its mother via milk-gland secretions; the *Wolbachia* are transmitted transovarially. Efforts to eliminate tsetse symbionts with antibiotics result in retarded growth and a decrease in egg production. Because it is impossible to eliminate only one at a time, it is difficult to decipher the role each plays. However, the gut symbionts supply B-complex vitamins, and *Sodalis* also produces a chitinase that seems responsible for increasing the susceptibility of its host to the sleeping sickness trypanosome (Aksoy 2000). Analysis of the *Wigglesworthia* and *Sodalis* genomes indicate that they each form a distinct lineage in the Proteobacteria. Molecular analyses suggest that a tsetse ancestor was infected with a *Wigglesworthia* and from this ancestral pair evolved the tsetse species and *Wigglesworthia* strains existing today. No evidence was found for horizontal transfer of *Wigglesworthia* symbionts between tsetse species. *Sodalis* infections might represent recent independent acquisitions by each tsetse species or multiple horizontal transfers between tsetse species.

## 4.13 Insect Development

Studies of *Drosophila melanogaster* have provided much of what we know about the genetics of development in insects (Lawrence 1992, Wilkins 1993, Klingler 1994, Powell 1997, Gilbert 2000, Otto 2000), although that is beginning to change. Extensive analyses of insect development became feasible with the tools of molecular genetics and thousands of papers have been published on the molecular genetics of development in *D. melanogaster*. Review articles and books have been published on this rapidly advancing field (Lawrence 1992, Wilkins 1993). A complete discussion of insect development is beyond the scope of this chapter. However, the following provides a brief outline of *D. melanogaster* embryonic development that will be useful in understanding sex determination, behavior, and *P*-element-mediated transformation (Chapters 9–11).

### 4.13.1 Oocyte Formation in *D. melanogaster*

A substantial amount of development of the insect embryo is determined in the oocyte, before oocyte (n) and sperm (n) pronuclei fuse to form an embryo

(2n). Oocyte formation in *D. melanogaster* is complex, involving both somatic and germ-line cells. The ovaries contain oocytes that are formed from the “pole cells,” but the cells that surround each egg chamber and make up the walls of the egg chambers are derived from mesoderm (somatic tissues). The pro-oocyte arises in a set of cell divisions within the ovary from an oogonial stem cell. Each oogonial stem cell divides to give a daughter stem cell and a cystoblast cell. The cystoblast cell gives rise to a set of 16 sister cells after four mitotic divisions, which provides a cyst. One of these 16 cells becomes the pro-oocyte, and eventually the oocyte, whereas its 15 sister cells become nurse cells whose function is to synthesize materials to supply the growing oocyte. The 16-cell cyst, surrounded by a layer of somatic cells is termed the egg chamber. The final stages of egg chamber development involve covering the cyst with a monolayer of pre-follicle cells; these cells are somatic in origin. These 80 somatic cells divide an additional four times to give 1200 follicle cells that cover each cyst.

Initially *Drosophila* oocytes and nurse cells are roughly the same size, but they increase in volume by  $\approx$ 40-fold when vitellogenin begins to accumulate approximately halfway through development of the oocyte. Some vitellogenin is derived from the follicle cells, but most is produced in the fat body and transported to the ovary (Raikhel and Dhadialla 1992). The later stages of oocyte development involve very rapid growth, with the oocyte increasing in volume 1500-fold. While the oocyte is increasing in size, the nurse cells are decreasing because their contents are being deposited in the oocyte. Nurse cells, derived from the germ line, are polyploid, containing 512 and 1024 times the haploid DNA content. These polyploid nurse cells synthesize proteins, ribosomes, and mRNAs. These products, and mitochondria, are transferred to the oocyte by intercellular channels. Thus, the oocyte contains products produced by the mother, which means that initial development in the oocyte is highly dependent upon the genome of the mother (=maternal effects). Finally, the vitelline membrane and the chorion are secreted around the oocyte by follicle cells and the oocyte enters metaphase of meiosis I. The oocyte remains arrested at metaphase of meiosis I until after fertilization.

The oocyte increases in total volume during its development by  $\times$ 90,000-fold. Oogenesis is a complex developmental pathway that is estimated to require the function of 70–80% of all genes in the *Drosophila* genome, although the majority is expressed during other stages of development as well. Only  $\approx$ 75 genes are expressed exclusively during oogenesis (Perrimon et al. 1986). The egg of *D. melanogaster* is rich in stored RNA, including rRNA and mRNA. The bulk of the maternally produced, stored mRNA is derived from transcription of nurse-cell nuclei during egg-chamber growth, but some mRNA may be derived from the oocyte

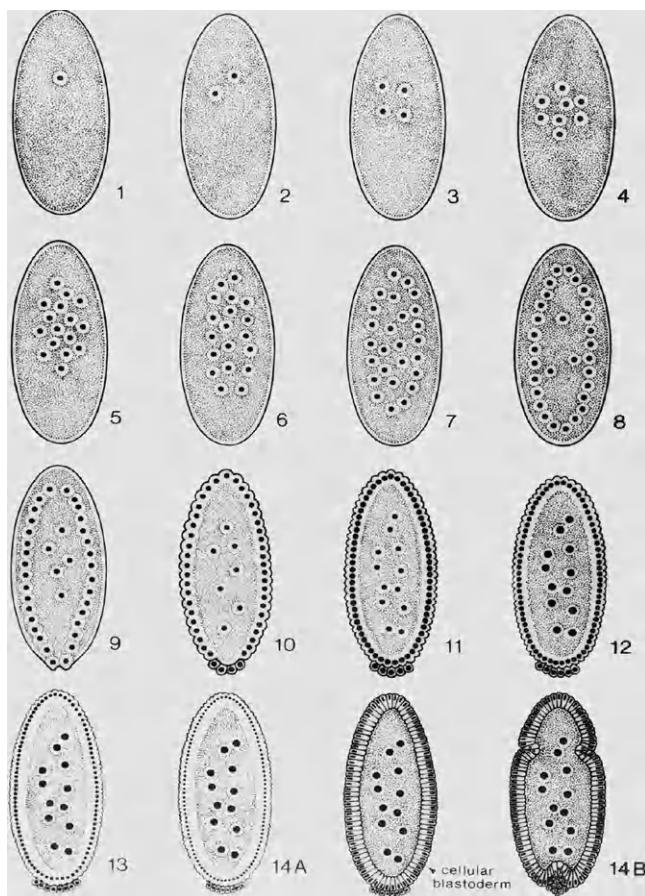
nucleus itself, which is active briefly approximately halfway through development. The total amount of mRNA in the oocyte is equal to  $\approx 10\%$  of the single-copy DNA of the *Drosophila* genome and corresponds to  $\approx 8000$  distinct protein-coding sequences. Most of the mRNA codes for proteins that required early in embryogenesis, including tubulins and histones. Products from a few maternal genes continue to affect development in *D. melanogaster* during the larval stage.

#### 4.13.2 Embryogenesis in *D. melanogaster*

Fertilization occurs when the mature oocytes are released into the oviducts. A single sperm enters the egg cytoplasm through a special channel in the anterior region of the oocyte called the micropyle. Fertilization initiates the completion of meiosis I and II. Two polar body nuclei and the female pronucleus are produced. After the haploid male and female pronuclei unite (**syngamy**), early embryogenesis takes place so rapidly there is no time for cell growth (Figure 4.5). Initial mitoses are atypical because the first nine divisions result in a **syncytium** containing  $\times 512$  nuclei that lack cellular membranes.

After seven nuclear divisions, and when there are 128 nuclei in the central region of the egg, most of the nuclei and their surrounding cytoplasm migrate outward as they continue to divide. A few nuclei are left behind which divide once to become yolk nuclei that do not become incorporated in the embryo (Figure 4.5). After nine divisions, most of the nuclei have migrated to the egg surface. At this time, the soma and germ-line nuclei segregate when  $\approx 15$  nuclei move to the posterior region of the egg, bud off, and eventually become the nuclei in the pole cells. These nuclei divide about twice more and become **pole cells** that will give rise to the germ-line tissues of the fly. Meanwhile, the other nuclei migrate to the surface of the egg and divide four times more in synchrony to produce a syncytial blastoderm.

Finally, the membrane covering the egg invaginates to enclose each nucleus in a separate membrane, to form a cellularized **blastoderm** (Figure 4.5). The blastoderm is the layer of cells in an insect embryo that completely surrounds an internal yolk mass. The cellular blastoderm develops from a syncytial blastoderm by partitioning the cleavage nuclei with membranes derived from infolding of the oolemma (egg membrane). During the cellular blastoderm stage, *D. melanogaster* exhibits the **long-germ band** type of development, in which the pattern of segmentation is established by the end of blastoderm. Some other insects exhibit a **short-germ band** type of development, in which all or most of the metamerized pattern is completed by the sequential addition of segments during elongation of the caudal region of the embryo.



**Figure 4.5** Early embryonic stages of *Drosophila melanogaster* from fertilization to just before gastrulation, showing the appearance of pole and somatic buds and cessation of division of yolk nuclei. Numbers indicate division cycles; each cycle begins with the start of interphase and ends at the conclusion of mitosis. Embryos are in longitudinal section without the vitelline membrane. All nuclei (black circles) are shown for cycles 1–5 and afterwards only some are shown. Stippled areas represent yolk and open areas represent yolk-free cytoplasm. Yolk-free cytoplasm is found both at the periphery (=periplasm) and in islands around the nuclei.

During Cycles 1–7, nuclei multiply exponentially in the central region of the fertilized egg. Cycle 8 illustrates migration of the majority of the nuclei to the periphery, leaving the future yolk nuclei behind in the center. Yolk nuclei continue to divide in synchrony with other nuclei in cycles 8–10; they then cease dividing and become polyploid. Early in cycle 9, a few nuclei appear in the posterior periplasm and cause protrusions of the cytoplasm, called pole buds. During cycle 10, the remaining migrating nuclei enter the periplasmic region, forming somatic buds over the entire embryonic surface. During the tenth cycle, pole buds are pinched off to form pole cells. After this, synchrony between the pole cells and the syncytium is lost. The syncytial nuclei continue to divide synchronously. The periplasm begins to thicken in cycle 13. During cycle 14, the formation of a plasma membrane begins to separate cells over the entire surface of the embryo, with nuclei elongating to match elongated cells formed by late cycle 14A. During 14B, gastrulation movements begin with the infolding of the cephalic furrow (anterior) and posterior midgut furrow, and subsequently the cells no longer divide synchronously.

Before the cellularized-blastoderm stage, the dividing nuclei are equivalent and totipotent, but after the cellularized blastoderm stage, specific body segments have been determined. The cellularized blastoderm stage is a key transition point in embryogenesis in *D. melanogaster* because this is when the products of maternal genes become less important. It is thought that only a few zygotic genes are active before cellularization. After the cellularized blastoderm stage, the genes in the zygote begin to dominate in directing the development of the embryo (Walser and Lipshitz 2011). After additional development, the insect embryo gives rise to a segmented larva with three major tagmata: the head, thorax, and abdomen.

#### 4.13.3 Postembryonic Development

*Drosophila melanogaster* is a holometabolous insect with sequential life stages: egg → larval stage 1 → molt → larval stage 2 → molt → larval stage 3 → molt → pupa → molt → adult. The larva hatches (=ecloses) from the egg, grows, and molts after each larval stadium. After the third larval stadium, the insect molts, pupates, and undergoes metamorphosis to the adult form. During metamorphosis, most of the larval tissues are digested.

Adult structures develop from cells in structures called **imaginal discs** and abdominal histoblast nests that will give rise to the abdominal epithelium. The cells that give rise to the 19 imaginal discs became segregated from surrounding cells during the first half of embryogenesis. By the time the larva hatches, the imaginal discs and histoblast cells are visibly distinct from the surrounding larval cells because they have smaller nuclei and an undifferentiated appearance. The labial, clypeolabral, antennal + eye, thoracic, three leg, wing, and haltere (=wings on metathorax of other insects) imaginal discs are paired. In addition, there is a single fused genital disc. Imaginal cells are diploid and able to divide. By contrast, most of the 6000 somatic cells of *D. melanogaster* grow in size, but they do not undergo cell division. The chromosomes of the larval cells continue to undergo replication and become polytene.

At the end of the third larval stage, the larva transforms into a pupa. During the pupal stage, the imaginal discs, each consisting of about 40 cells, develop into adult structures such as legs, wings, eyes, ovaries or testes, and antennae (Larsen-Rappaport 1986). Because the imaginal discs were determined during embryonic development, the basic body plan of the adult fly was laid down before the larva eclosed from the egg. The wings, halteres, and legs of the adult, with as many as 50,000 cells each, are formed from the imaginal discs. Not only is segmentation in *Drosophila* larvae based on cues obtained from the

mother, but coordinates provided by the mother also determine the position and organization of adult structures (Couso and Gonzalez-Gaitan 1993). After emergence as an adult, the insect mates and their progeny begin this developmental cycle again.

#### 4.14 Dissecting Development with *D. melanogaster* Mutants

The study of development in *Drosophila* has depended upon the availability of mutants so that the process can be dissected; this process is called **reverse genetics**. In fact, in discussing development in *Drosophila*, the genes influencing development are called by names that reflect their mutant form. Nusslein-Volhard and Wieschaus (1980) began a systematic program of mutagenizing *Drosophila* females to obtain many developmental mutants in insect embryos. In addition, E. B. Lewis discovered many useful developmental mutants in his pioneering work on *Drosophila* development involving the bithorax homeotic gene complex. Christiane Nusslein-Volhard, Eric Wieschaus, and Edward Lewis jointly were awarded the 1995 Nobel Prize in Medicine for their pioneering research on development.

The term **homeotic** was coined to describe the replacement of one part of the body by a serially homologous part. Lewis (1978) developed the hypothesis that families of structurally related genes control the specification of the insect body plan because insect bodies are metamerized (composed of serially repeating units or body segments). The body segments differentiate into specific structures according to their position. Likewise, the appendages in each major body segment develop into appropriate structures, with antennae located on the head, legs on the thoracic segments, and wings on the second thoracic segment. However, this normal pattern is disrupted by a number of homeotic mutants.

*Drosophila* embryos seem to go through two phases in their development. During the first phase, many genes seem to encode transcription factors or nuclear proteins, suggesting that a cascade of transcriptional factors regulate other genes in development. There is a successive subdivision of the embryo into smaller and smaller domains that is accomplished by the differential and combinatorial action of transcription factors. The first phase is completed by the time cells and are fully formed at the end of blastoderm.

The second phase begins after the formation of the cellular blastoderm and consists of elaborating the information provided from reference points that have been deposited along the dorsal–ventral and anterior–posterior axes. This requires the communication of information between cells by intercellular signal molecules.

The genes that control *Drosophila* embryonic development can be divided into three classes: 1) **maternal-effect genes** that specify egg polarity and the spatial coordinates of the egg and future embryo; 2) **segmentation genes**, including the gap, pair-rule, and segment-polarity classes of genes that determine the number and polarity of the body segments; and 3) **homeotic genes** that determine the identification and sequence of the segments. Although most genes with a homeo domain are in the homeotic class, a few are found among the segmentation genes (Figure 4.6).

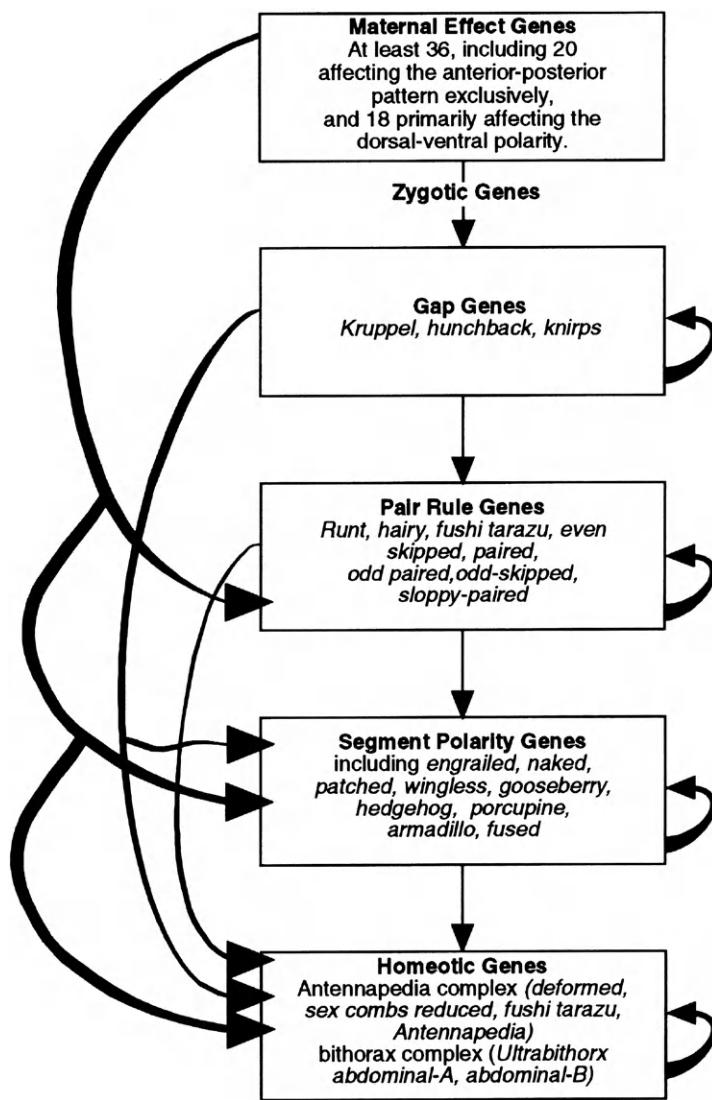
#### 4.14.1 Maternal-Effect Genes

Maternal-effect genes are transcribed in the mother, and their messenger RNAs influence the development of oocytes and embryos. Maternal-effect genes were discovered by determining whether the mother carries a mutant gene that results in an inability to rescue her embryo when the father contributes a wild-type gene at the time the egg is fertilized. Maternal-effect genes are important because the basic organization of the oocyte (front-back, top-bottom) has been accomplished even before the egg has been fertilized. At least 38 maternal-effect genes have been studied;  $\approx 20$  affect the anterior-posterior pattern and  $\approx 18$  affect the dorsal-ventral polarity of the embryo (Figure 4.6).

The 20 genes influencing differentiation into proper anterior-posterior polarity can be subdivided into a group affecting the anterior half of the embryo, a group affecting the posterior half, and a group affecting both the anterior and posterior ends. Of 18 genes affecting dorso-ventral polarity, several are required for the polarity of both the eggshell and embryo, and several are required for the polarity of the embryo only (Wilkins 1993).

Determining the dorsal-ventral and anterior-posterior polarity in the embryo is a highly significant step. The *Toll*<sup>+</sup> gene is a dorsalizing gene, and mutations produce embryos lacking both ventral and lateral structures. The *Toll*<sup>+</sup> gene product appears to be a **morphogen** (molecules whose local concentration directly determine the local pattern of differentiation), because when the wild-type gene product is injected into mutant embryos it can make the affected region become the dorsal region of the fly.

A major determinant of anterior-posterior polarity is the product of the *bicoid*<sup>+</sup> gene. It is transcribed in the nurse cells of the ovary, and the mRNA passes into the oocyte where it becomes localized in the anterior of the egg, apparently aided by components of the cytoskeleton that are encoded by products of the genes *swallow*<sup>+</sup> and *exuperentia*<sup>+</sup>.



**Figure 4.6** Development of segments in the embryo of *Drosophila melanogaster* involves a hierarchy of regulatory genes. The earliest-acting genes are maternal-effect genes that regulate the spatial expression of later-acting genes. Some maternal-gene products are present in concentration gradients and regulate downstream genes in a concentration-dependent manner. In many cases, genes within a given class regulate other members of the class (depicted as arrows beginning and ending at the same class). Many other genes not shown may play a role in segmentation.

Several of the maternal-effect genes (*nanos*<sup>+</sup>, *cappuccino*<sup>+</sup>, *spire*<sup>+</sup>, *staufen*<sup>+</sup>, *oskar*<sup>+</sup>, *vasa*<sup>+</sup>, *valois*<sup>+</sup>, and *tudor*<sup>+</sup>) are required for the localization of factors that determine the germ line. In addition, mitochondrial large ribosomal RNA may be important for pole-cell formation (Kobayashi et al. 1993). Thus, both

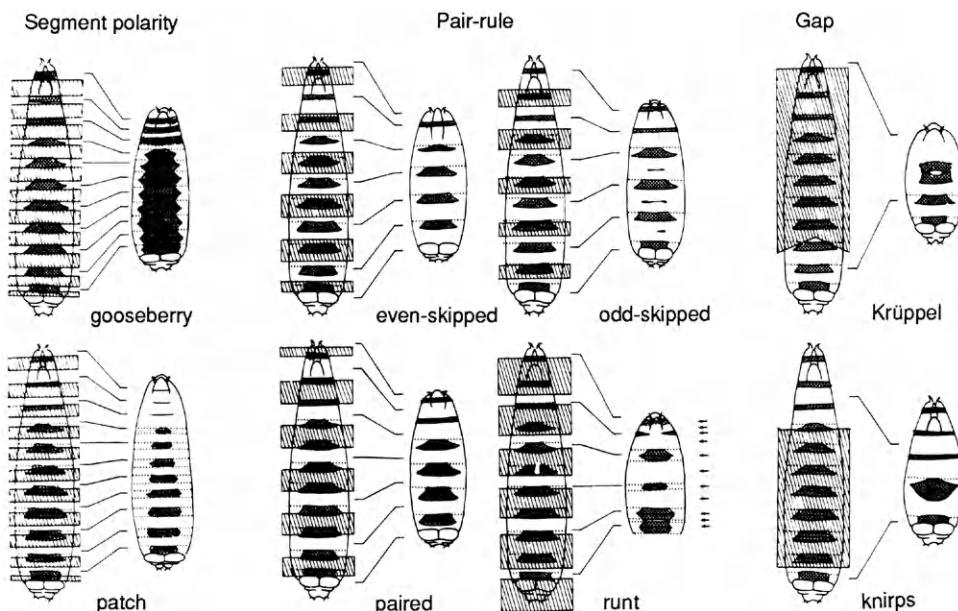
mitochondrial and nuclear genomes are involved in determination of the germ line during embryonic development.

Maternal-effect genes are most important during development of the egg up to the blastoderm stage; before the blastoderm stage, the embryo has discrete nuclei but no cell walls (it is syncytial). After blastoderm, genes inherited by the zygote from both parents become dominant factors in determining development. However, because development is an elaborative process and the adult phenotype is a summation of the developmental effects accrued over the life span of the individual, developmental events early in the life cycle can significantly influence the phenotype at later stages. For example, maternal effects have significant and diverse effects on insect life histories, including incidence and intensity of diapause, production of sexual forms, wing polyphenism, dispersal behavior, developmental time, growth rate, resistance to chemicals and microbial infection, and survival. Some of these influences are caused by maternal age and diet, but some are genetically determined (Mousseau and Dingle 1991).

#### 4.14.2 Zygotic-Segmentation Genes

During blastoderm, the embryo begins to develop a pattern of repeating body segments. Zygotic genes of three types control segmentation: 1) pair-rule, 2) gap, and 3) segment-polarity genes. Segmentation mutants found in *Drosophila* embryos initially were difficult to interpret because they did not affect what appeared to be a single segment; usually they affected half of one "segment" and the adjacent half of the next "segment." Eventually, it was determined that "true segments" are not reflected by the visible cuticular patterns of sclerites and sutures; visible "segments" are, in fact, **parasegments**. There are 14 complete parasegments in *D. melanogaster* that are defined early in development; each is a precise set of cells.

Gap genes subdivide the embryo along the anterior–posterior axis because they encode transcription factors that regulate the expression of the pair-rule genes. Pair-rule genes divide the embryo into pairs of segments and encode transcription factors that regulate the expression of the segment-polarity genes. The segment-polarity genes establish the anterior–posterior axis of each segment (Figure 4.7). Most segmentation mutants are lethal in the zygote, but some gap genes have a maternal effect and are expressed during oogenesis. Four of the segmentation genes (*fushi tarazu*<sup>+</sup>, *even-skipped*<sup>+</sup>, *paired*<sup>+</sup>, and *engrailed*<sup>+</sup>) contain a homeo box (see Section 4.14.2.4). Thus, these genes encode DNA regulatory proteins or transcription factors that bind to specific DNA or RNA sequences.



**Figure 4.7** Embryonic segment pattern defects are illustrated with selected mutants: *gooseberry* and *patch* are segment-polarity mutants; *even-skipped*, *odd-skipped*, *paired* and *runt* are pair-rule mutants; and *Krüppel* and *knirps* are gap genes. Dotted regions represent denticle bands on the developing embryo and dotted lines show the boundaries of segments. Hatched regions indicate the parts of the pattern that are missing in the mutants. Transverse lines link the corresponding regions in mutant and wild-type embryos. Arrows indicate where lines of polarity have been reversed. (From *Nature* 287: 796, 1980.)

#### 4.14.2.1 Gap Genes

**Gap genes** were so named because large areas of the normal cuticular pattern are deleted in individuals with mutant phenotypes (Figure 4.7). The three wild-type versions of the gap genes *Krüppel*<sup>+</sup>, *hunchback*<sup>+</sup>, and *knirps*<sup>+</sup> regionalize the embryo by delimiting domains of homeotic gene expression and affect position-specific regulation of the pair-rule genes *runt*<sup>+</sup>, *fushi tarazu*<sup>+</sup>, *even skipped*<sup>+</sup>, *paired*<sup>+</sup>, and *odd-paired*<sup>+</sup>. All three gap-gene products contain DNA-binding domains.

The gap genes interact to produce sharp boundaries. The establishment of stable domains by the gap genes is a two-step process: 1) a differential response to graded levels of maternal determinants, 2) followed by mutual repression leading to the generation of stable boundaries between adjacent domains. Gap genes regionalize the embryo by delimiting the domains of later homeotic-gene expression and this results in position-specific regulation of the pair-rule class of genes.

#### 4.14.2.2 Pair-Rule Genes

The **pair-rule genes** were so named because mutant flies have a repetitive aberration throughout the germ band, with the removal of integral, alternate segment-width areas (Figure 4.7). The pair-rule genes are transiently expressed in seven or eight stripes during cellularization of the blastoderm. However, each is otherwise unique in its expression. Pair-rule genes are essential, directly or indirectly, for the initial establishment of segmentation.

#### 4.14.2.3 Segment-Polarity Genes

The **segment-polarity genes** seem to determine anterior-posterior domains within each segment (Figure 4.7). Segment-polarity genes are required either continuously or over extensive periods to maintain the segmentation pattern. Most or all are required to maintain patterns in the imaginal tissues.

#### 4.14.2.4 Homeotic Genes

In 1894, William Bateson coined the word **homeosis** to describe the situation in which “something has been changed into the likeness of something else” (Lewis 1994). Bateson was attempting to provide evidence in support of Darwin’s theory of evolution and homeotic variations seemed to Bateson to be the kind of dramatic changes that could explain how evolution occurred. E. B. Lewis (1994) concluded that homeosis provided a rich legacy: “Besides giving us the homeobox, it has opened up a completely new approach to the study of development. And over the past 15 years, it has led to the realization that the body plan of most animals, and presumably of plants as well, is controlled by a set of master regulatory genes, first identified by their homeotic mutations.”

The periodic pattern of body segments generated by segmentation genes (gap genes, pair-rule genes, and segment-polarity genes) has to be converted into segments with wings, legs, and antennae (Figure 4.7). Thus, in insects, thoracic segment 2 is different from thoracic segment 3 and abdominal segment 2 will be different from the terminal abdominal segments, which typically have genital structures. This fine-tuning is determined by **homeotic** or **Hox genes**. Hox genes contain a special sequence called the **homeobox**. The homeobox consists of  $\approx 180$  bp that is translated into a 60-amino-acid domain. The sequences of the different homeoboxes are nearly identical and they mediate the binding of homeotic proteins to specific DNA sequences and thus regulate the expression of many downstream genes. It has been proposed that just two homeotic genes, *even-skipped*<sup>+</sup> and *Fushi tarazu*<sup>+</sup>, directly control the expression of the majority of genes in the *Drosophila* genome (Mannervik 1999). Homeodomain proteins occur in all eukaryotes, where they perform important functions during development.

Since the first homeobox sequence was isolated from the *Antennapedia*<sup>+</sup> gene in late 1983, it has been used as a probe to identify and isolate previously unknown homeotic genes from *Drosophila*. Furthermore, because the homeobox is evolutionarily conserved, this *Drosophila* sequence was used as a probe to identify homeotic genes from other species, including humans (Gehring 1985). One of the homeobox sequences of *Xenopus* is so similar to the *Antennapedia* homeobox in *Drosophila* that only one amino acid of 60 is different. The reason for this extraordinary conservation during evolution is not fully understood (Gehring 1987). The sequences of the different *Hox* components in the vertebrate and insect gene clusters are comparable, as well as the order of those elements on the chromosome (Nam and Nei 2005). This remarkable conservation suggests that there may have been an ancestral sequence common to flies and humans (Lawrence 1992).

Homeotic mutants may have segments that are transformed dramatically. For example, antennal segments may be transformed into leg-like structures, and metathoracic segments with halteres may be transformed into mesothoracic segments with a set of wings (Figure 4.8). The four-winged *D. melanogaster* shown in Figure 4.8 is the result of combining three separate mutated genes in one fly! Normally, of course, a pair of wings is found on the second thoracic segment and a pair of balancing organs, called halteres, is on the third. However, this fly has two essentially normal second thoraces (and no third thoracic segment) because the combined effect of the three mutations is to transform the third thoracic segment into the second without affecting any other parts of the fly.



**Figure 4.8** A four-winged *Drosophila melanogaster* fly showing a complete transformation of the third thoracic segment into a second thoracic segment. The fly carries one chromosome with a deletion of the homeotic *bithorax* complex, while the other chromosome carries mutations of the *bithorax* (*Abx*, *bx3*) and *postbithorax* (*pbx*) loci. (Photograph provided by Edward Lewis, California Institute of Technology.)

Lewis, E., (1978) proposed a combinatorial model that assumes each insect segment is specified by a unique combination of homeotic genes that are expressed in that particular segment. Thus, the fewest number of homeotic genes would be required in thoracic segment 2, which would be the prototypical segment, and progressively more genes would be active in the more-posterior segments. Although this model has been modified, it provided a useful conceptual framework for investigating *Drosophila* development.

Homeotic genes have some unusual characteristics. First, several homeotic genes seem to be very large relative to most other genes in *Drosophila*. For example, the *Antennapedia*<sup>+</sup> primary gene transcript is ≈100 kb long and the *Ultrabithorax*<sup>+</sup> transcript is ≈75 kb. However, after the introns are spliced out, the remaining sequences are only a few kilobases. Many of the exons in homeotic genes seem to encode protein domains with distinct structural or enzymatic functions. As a result, alternative splicing patterns in large genes, such as the *Antennapedia*<sup>+</sup> and *bithorax*<sup>+</sup> gene complexes, may allow organisms to adapt one basic protein structure to different, but related, developmental uses. By adding or subtracting functional protein domains encoded by optional exons, the structural and enzymatic properties of the homeotic gene product can be modified and the ability of the protein to interact with other cellular components can be altered as development proceeds.

#### 4.14.3 Insect Metamorphosis

Insects may be ametabolous (Protura, Diplura, Thysanura, Collembola), hemimetabolous (Ephemeroptera, Odonata, Mantodea, Phasmida, Isoptera, Blattaria, Plecoptera, Dermaptera, Embioptera, Psocoptera, Mallophaga, Anoplura, Hemiptera), or holometabolous (Neoptera, Megaloptera, Mecoptera, Coleoptera, Trichoptera, Lepidoptera, Diptera, Siphonaptera, Hymenoptera). Ametabolous insects change little in form, except adults have fully developed reproductive organs. Hemimetabolous insects undergo a progressive change (egg→nymphal stages→adult) and holometabolous insects undergo dramatic changes at the end of the life cycle (egg→larval stages→pupa→adult). Some entomologists further subdivide the types of metamorphosis, but these are the main categories.

The fossil record indicates that hemimetabolous insects were present ≈300 million years ago and that holometabolous insects appeared ≈280 million years ago (Belles 2010). The changes in morphology during development of both hemi- and holometabolous insects are regulated by molting hormone (which promotes molts) and juvenile hormone (JH), which represses transformation into the adult stage. These two hormones are regulated by many transcription factors, and microRNAs affect expression of these transcription factors (Belles 2011). Gomez-Orte and

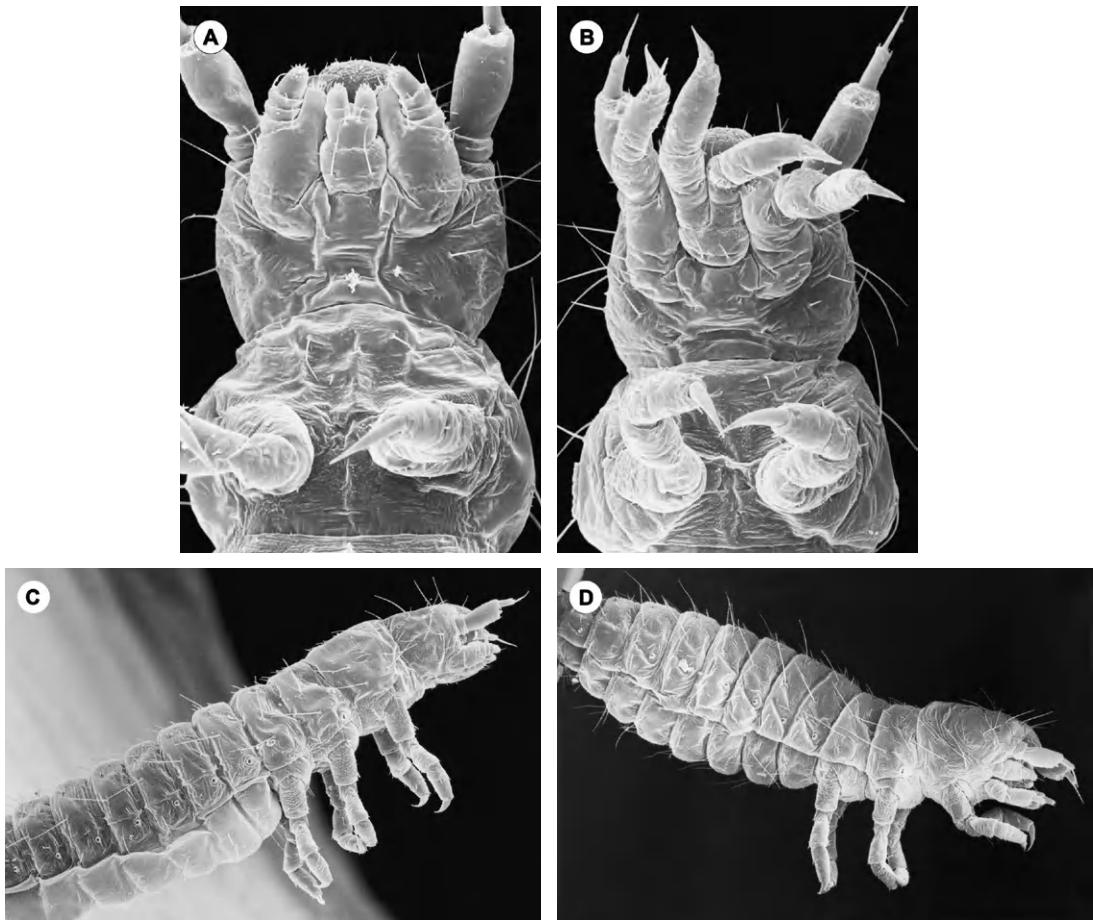
Belles (2009) studied metamorphosis in hemimetabolous insects using RNA interference (RNAi) methods and showed that microRNAs are involved in metamorphosis of *Blattella germanica*, perhaps by regulating genes that are juvenile hormone targets. Zhang et al. (2009) showed that 106 microRNAs are expressed in all stages of *Bombyx mori*, but 248 miRNAs were egg- and pupa-specific, suggesting that they are important in embryogenesis and metamorphosis.

#### 4.15 Interactions During Development

Normal development requires the coordinated expression of thousands of structural genes in a controlled manner. Because independent control of individual structural genes would result in chaotic development, controlling genes regulate the activity of groups of structural genes in a coordinated manner. Such genes are presumably arranged hierarchically or form a controlling network that ensures the correct timing of development so that the proper pattern develops. Although development in *Drosophila* is not fully understood, many of the genes and their interactions are known, including those that regulate the development of mesoderm (Furlong et al. 2001), appendages (Morata 2001), and the eye (Thomas and Wasserman 1999). It is likely that understanding development in *D. melanogaster* will elucidate many of the principles by which other higher eukaryotes develop. The molecular genetics of sex determination is discussed in Chapter 10 and provides another example of the hierarchical nature of development.

#### 4.16 Similarities and Differences in Development in Other Insects

Although developmental studies are most advanced in *D. melanogaster*, it is important to carry out comparable research with other insects to determine whether our knowledge of development in *D. melanogaster* can be extrapolated to other insects (Klingler 1994, Buning 1994, Davis and Patel 2002, Lynch et al. 2012). *Drosophila* represents a relatively specialized type of development, the long-germ development pattern, in which segmentation occurs essentially simultaneously along the anterior-posterior axis with the process of segmental specification under the control of homeotic genes in the Antennapedia and bithorax complexes. Lynch et al. (2012) compared the embryonic development of *Drosophila*, *Nasonia*, and *Tribolium*. *Nasonia* has a long-germ band mode of embryogenesis, similar to that of *Drosophila*, but *Tribolium* has an intermediate mode of embryogenesis, which is more like the ancestral mechanism. In this system, only the most anterior segments are specified before gastrulation and the posterior segments are patterned from a posterior region called the growth zone. Understanding the development of *Nasonia* and *Tribolium* has been enhanced by the availability of fully sequenced genomes, so that **orthologs**



**Figure 4.9** A) A ventral view of a wild type embryo of *Tribolium castaneum*. B) A ventral view showing a homeotic mutation, *maxillopedia*, in which the maxillary and labial palps are modified into leg-like appendages. C) A side view of a mature wild-type embryo. D) A side view showing the homeotic mutant *cephalothorax*, which results in incorporation of the prothorax into the head and transformation of the labial palps into antennae. *maxillopedia* and *cephalothorax* are in the HOM-C complex and correspond to *proboscipedia* and *sex combs reduced* in *Drosophila*. (Photographs provided by Richard W. Beeman.)

(genes in different species that originated by descent from a single gene of the last common ancestor) to those in *Drosophila* can be identified.

Beeman (1987) showed that six loci of homeotic genes in a single cluster (HOM-C) of *T. castaneum* contain elements homologous to the homeotic genes in the Antennapedia and bithorax complexes of *Drosophila*. These genes map along the chromosome in the same order from anterior to posterior as their effects occur, but they occur in a single cluster rather than in two in *Drosophila* and with a different gene order (Beeman et al. 1989, 1993a,b). *Tribolium* with

the *maxillopedia* mutation have labial and maxillary palps transformed into leg-like structures (Figure 4.9). *Tribolium* with mutations of *Cephalothorax* have the head and first thoracic segment fused and the labial palps transformed into antennae (Figure 4.9). *Tribolium* with *prothoraxless* mutations exhibit fusion of head segments with the entire thorax, and transformation of all three pairs of thoracic legs into antennae.

Comparative studies on the molecular evolution of genes involved in development in additional insects provide information about the evolution of gene families regulating development, as well as contribute to our understanding of the basic mechanisms underlying the genetic control of development (Davis and Patel 2002, Roth and Hartenstein 2008, Schroder et al. 2008, Shippy et al. 2008, Lynch et al. 2012).

#### 4.17 Evo-Devo and the Revolution in Developmental Studies

“Evolutionary developmental biology,” also called “evo-devo,” recently emerged to study the diversity of development in animal and plant forms from an evolutionary point of view (Pennisi and Roush 1997, West-Eberhard 1998, Holland 1999, von Dassow and Munro 1999, Dalton 2000, Jenner 2000, Muller 2007, Brakefield 2011). Evo-devo combines comparative embryology, paleontology, molecular phylogenetics, and genome analysis. The goals of **evo-devo** are to understand the following (Hall 2000):

- The origin and evolution of embryonic development,
- How modifications of development and developmental processes lead to the production of novel features,
- The adaptive plasticity of development in life-history evolution,
- How ecology affects development to modulate evolutionary change, and
- The developmental basis of homoplasy and homology.

The assumption underlying evo-devo is that analysis of the evolution of developmental stages, processes, and mechanisms will enable us to understand how organisms, organs, tissues, cells, and genes evolve. Evo-devo successes include learning the following (Hall 2000):

- Genes that control major developmental processes (establishment of body plans, formation of appendages and sense organs) are shared across the animal kingdom and arose early in metazoan evolution.
- New knowledge of developmental mechanisms underlying the formation of organs or major body parts has led to an understanding of the mechanisms involved in their origin from structures in ancestral organisms.

- Loss of organs does not imply loss of the developmental potential to form those organs.
- Life-history stages (embryos, larvae, adults) can develop and evolve separately, providing opportunities to modify and modulate embryonic development, for specialization or diversification of adult structure, and for the evolution of novel structures.
- Homology is now seen as hierarchical, with homologous genes initiating development of structures that are not homologous (such as arthropod and vertebrate eyes) and homologous structures developing by processes that are not homologous.

For example, studies seem to confirm that there was an inversion of the dorso-ventral axis during animal evolution, with the ventral region of *Drosophila* homologous to the dorsal side of vertebrates ([De Robertis and Sasai 1996](#)). Thus, developmental systems that control patterns from eggs to adults are remarkably similar across a wide range of phyla despite at least half a billion years of evolution since their origin from their last common ancestor ([Erwin et al. 1997](#)). Despite their genetic similarities, the developmental systems produce very different body plans. The basic aspects of the developmental control systems of long-extinct animals can now be reconstructed and the diversification of animal form and the evolution of the genetic controls that regulate it are becoming understood ([Erwin et al. 1997](#)).

The field of evo-devo uses the concept of homology. However, **homology** is a difficult and, sometimes, fuzzy term ([Bolker and Raff 1996](#)). There are at least nine homology concepts in current use in the comparative biology literature, which makes evo-devo a particularly contentious topic ([Janies and DeSalle 1999](#)). The concept of homology originated from classical studies of comparative adult morphology and one classical definition is “the same organ in different animals under every variety of form and function.” Another is “derived from an equivalent characteristic of the common ancestor.” Homology has been used recently to be “shared patterns of gene expression,” a controversial use of the term ([Bolker and Raff 1996](#)), and more often is limited to “similarities due to descent from the same ancestral source.” Because most genes play multiple roles in development, it is difficult to resolve homologies. For example, arthropods, annelids and chordates all possess segments. It remains unclear whether the segments evolved independently or are derived from a common ancestor ([Davis and Patel 1999](#)). A review of the successes of evo-devo by [Brakefield \(2011\)](#) noted “the evolution of endless forms is more about the evolution of the regulatory machinery of ancient genes than the origin and elaboration of new genes.”

New technology allows new approaches to understanding development, and the development of vast amounts of DNA sequence information allows researchers to determine which genes are functioning from a global perspective (Janies and DeSalle 1999, White et al. 1999). For example, it is possible, using microarray analysis (described in Chapter 6), to evaluate the role of hundreds of genes involved in the process of metamorphosis of *Drosophila*.

## References Cited

- Acuna, R., Padilla, B.E., Florez-Ramos, C.P., Rubio, J.D., Herrera, J.C., and Benavides, P., et al. (2012). Adaptive horizontal transfer of a bacterial gene to an invasive insect pest of coffee. *Proc. Natl. Acad. Sci. USA* **109**: 4197–4202.
- Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D., and Amanatides, P.G., et al. (2000). The genome sequence of *Drosophila melanogaster*. *Science* **287**: 2185–2195.
- Agarwal, M., Bensaadi, N., Salvado, J.C., Campbell, K., and Mouches, C., (1993). Characterization and genetic organization of full-length copies of a LINE retroposon family dispersed in the genome of *Culex pipiens* mosquitoes. *Insect Biochem. Mol. Biol.* **23**: 621–629.
- Aksoy, S., (2000). Tsetse-a haven for microorganisms. *Parasitol. Today* **16**: 114–118.
- Alam, U., Medlock, J., Brelsfoard, C., Pais, R., Lohs, C., and Balmand, S., et al. (2011). *Wolbachia* symbiont infections induce strong cytoplasmic incompatibility in the tsetse fly *Glossina morsitans*. *PLoS Pathogens* **7**: e1002415.
- Aminetzach, Y.T., Macpherson, J.M., and Petrov, D.A., (2005). Pesticide resistance via transposition-mediated adaptive gene truncation in *Drosophila*. *Science* **309**: 764–767.
- Antoine, M., and Niessing, J., (1984). Intron-less globin genes in the insect *Chironomus thummi thummi*. *Nature* **310**: 795–798.
- Ashburner, M., (1989). *Drosophila. A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Bachtrog, D., Weiss, S., Zangerl, B., Brem, G., and Schlotterer, C., (1999). Distribution of dinucleotide microsatellites in the *Drosophila melanogaster* genome. *Mol. Biol. Evol.* **16**: 602–610.
- Bandi, C., Slatko, B., and O'Neill, S.L., (1999). *Wolbachia* genomes and the many faces of symbiosis. *Parasitol. Today* **15**: 428–429.
- Bartholomay, L.C., Waterhouse, R.M., Mayhew, G.F., Campbell, C.L., Michel, K., and Zou, Z., et al. (2010). Pathogenomics of *Culex quinquefasciatus* and meta-analysis of infection responses to diverse pathogens. *Science* **330**: 88–90.
- Baumann, P., Moran, N.A., and Baumann, L., (1997). The evolution and genetics of aphid endosymbionts. *BioScience* **47**: 12–20.
- Bazzocchi, C., Jamnongluk, W., O’Neil, S.L., Anderson, T.J.C., Genchi, C., and Bandi, C., (2000). *wsp* gene sequences from the *Wolbachia* of filarial nematodes. *Curr. Microbiol.* **41**: 96–100.
- Beckage, N.E., (1998). Parasitoids and polydnaviruses. *BioScience* **48**: 305–310.
- Beeman, R.W., (1987). A homeotic gene cluster in the red flour beetle. *Nature* **327**: 247–249.
- Beeman, R.W., Stuart, J.J., Haas, M.S., and Denell, R.E., (1989). Genetic analysis of the homeotic gene complex (Hom-C) in the beetle *Tribolium castaneum*. *Dev. Biol.* **133**: 196–209.
- Beeman, R.W., Brown, S.J., Stuart, J.J., and Denell, R., (1993a). The evolution of genes regulating developmental commitments in insects and other animals *Evolutionary Conservation of Developmental Mechanisms*. Wiley-Liss, New York, 71–84.
- Beeman, R.W., Stuart, J.J., Brown, S.J., and Denell, R.E., (1993b). Structure and function of the homeotic gene complex (HOM-C) in the beetle, *Tribolium castaneum*. *BioEssays* **15**: 439–444.
- Belles, X., (2010). Beyond *Drosophila*. RNAi in vivo and functional genomics in insects. *Annu. Rev. Entomol.* **55**: 111–128.

- Belles, X., (2011). Origin and evolution of insect metamorphosis *Encyclopedia of Life Sciences*. (ELS). John Wiley, Chichester, UK.
- Bello, B., and Couble, P., (1990). Specific expression of a silk-encoding gene of *Bombyx* in the anterior salivary gland of *Drosophila*. *Nature* **346**: 480–482.
- Berg, D.E., and Howe, M.M., Eds., (1989). *Mobile DNA*, ASM Press, Washington, DC.
- Berry, S.J., (1985). Insect nucleic acids, pp. 219–253 In: Kerkut, G.A., and Gilbert, L.I., Eds., *Comprehensive Insect Physiology Biochemistry and Pharmacology*, Vol. 10, Pergamon Press, Oxford, UK.
- Bian, G., Xu, Y., Lu, P., Xie, Y., and Xi, Z., (2010). The endosymbiotic bacterium *Wolbachia* induces resistance to dengue virus in *Aedes aegypti*. *PLoS Pathogens* **6** (4): e10000833.
- Biernat, C., and Vieira, C., (2006). Junk DNA as an evolutionary force. *Nature* **443**: 521–524.
- Biernat, C., Vieira, C., Borie, N., and Lepetit, D., (1999). Transposable elements and genome evolution: the case of *Drosophila simulans*. *Genetica* **107**: 113–120.
- Biessmann, H., Kasravi, B., Jakes, K., Bui, T., Ikenaga, K., and Mason, J.M., (1993). The genomic organization of HET-A retroposons in *Drosophila melanogaster*. *Chromosoma* **102**: 297–305.
- Blagrove, M.S.C., Arias-Goeta, C., Falloux, A.B., and Sinkins, S.P., (2012). *Wolbachia* strain wMel induces cytoplasmic incompatibility and blocks dengue transmission in *Aedes albopictus*. *Proc. Natl. Acad. Sci. USA* **109**: 255–260.
- Block, K., Ising, G., and Stahl, F., (1990). Minichromosomes in *Drosophila melanogaster* derived from the transposing element TE1. *Chromosoma* **99**: 336–343.
- Blumenstiel, J.P., (2011). Evolutionary dynamics of transposable elements in a small RNA world. *Trends Genet* **27**: 23–31.
- Bolker, J.A., and Raff, R.A., (1996). Developmental genetics and traditional homology. *BioEssays* **18**: 489–494.
- In: Bourtzis, K., and Miller, T.A., Eds., (2003). *Insect Symbiosis*, Vol. 1, CRC Press, Boca Raton, FL.
- In: Bourtzis, K., and Miller, T.A., Eds., (2006). *Insect Symbiosis*, Vol. 2, CRC Press, Boca Raton, FL.
- In: Bourtzis, K., and Miller, T.A., Eds., (2009). *Insect Symbiosis*, Vol. 3, CRC Press, Boca Raton, FL.
- Bourtzis, K., Dobson, S.L., Braig, H.R., and O'Neill, S.L., (1998). Rescuing *Wolbachia* have been overlooked. *Nature* **391**: 852–853.
- Bowen, N.J., and McDonald, J.F., (2001). *Drosophila* euchromatic LTR retrotransposons are much younger than the host species in which they reside. *Genome Res* **11**: 1527–1540.
- Bownes, M., (1986). Expression of the genes coding for vitellogenin (yolk protein). *Annu. Rev. Entomol.* **31**: 507–531.
- Boyle, L.S., O'Neill, L., Robertson, H.M., and Karr, T.L., (1993). Interspecific and intraspecific horizontal transfer of *Wolbachia*. *Drosophila*. *Science* **260**: 1796–1799.
- Brakefield, P.M., (2011). Evo-devo and accounting for Darwin's endless forms. *Philos. Trans. R. Soc. B* **366**: 3069–3075.
- Britten, R.J., (1997). Mobile elements inserted in the distant past have taken on important functions. *Gene* **205**: 177–182.
- Broderick, N.A., Raffa, K.F., and Handelsman, J., (2006). Midgut bacteria required for *Bacillus thuringiensis*. *Proc. Natl. Acad. Sci. USA* **103**: 15196–15199.
- Brody, T., (1999). The Interactive Fly: gene networks, development and the Internet. *Trends Genet.* **15**: 333–334.
- Brosius, J., (1991). Retroposons—seeds of evolution. *Science* **251**: 753.
- Brown, S.W., and Nur, U., (1964). Heterochromatic chromosomes in coccids. *Science* **145**: 130–136.
- Brownlie, J.C., Cass, B.N., Riegler, M., Witsenburg, J.J., Iturbe-Ormaetxe, I., McGraw, E.A., and O'Neill, S.L., (2009). Evidence for metabolic provisioning by a common invertebrate endosymbiont, *Wolbachia pipiensis*, during periods of nutritional stress. *PLoS Pathogens* **5** (4): e1000368.
- Brune, A., and Friedrich, M., (2000). Microecology of the termite gut: structure and function on a microscale. *Curr. Opin. Microbiol.* **3**: 263–269.

- Buning, J., (1994). *The Insect Ovary. Ultrastructure, Previtellogenesis Growth and Evolution*, Chapman and Hall, New York.
- Burke, G.R., and Moran, N.A., (2011). Massive genomic decay in *Serratia symbiotica*, a recently evolved symbiont of aphids. *Genome Biol. Evol.* **3**: 195–208.
- Burke, W.D., Eickbush, D.G., Xiong, Y., Jakubczak, J., and Eickbush, T.H., (1993). Sequence relationship of retrotransposable elements R1 and R2 within and between divergent insect species. *Mol. Biol. Evol.* **10**: 163–185.
- Caldera, E.J., Poulsen, M., Suen, G., and Currie, C.R., (2009). Insect symbioses: A case study of past, present, and future fungus-growing ant research. *Environ. Entomol.* **39**: 78–92.
- Calvi, B.R., Hong, T.J., Findley, S.D., and Gelbart, W.M., (1991). Evidence for a common evolutionary origin of inverted repeat transposons in *Drosophila* and plants: *hobo*, *activator*, and *Tam3*. *Cell* **66**: 465–471.
- Carminati, J.L., Johnston, C.G., and Orr-Weaver, T.L., (1992). The *Drosophila* ACE3 chorion element autonomously induces amplification. *Mol. Cell. Biol.* **12**: 2444–2453.
- Carton, Y., and Nappi, A., (2001). Immunogenetic aspects of the cellular immune response of *Drosophila* against parasitoids. *Immunogenetics* **52**: 157–164.
- Cazemier, A.E., Hackstein, J.H.P., Op den Camp, H.J.M., Rosenberg, J., and van der Drift, C., (1997). Bacteria in the intestinal tract of different species of arthropods. *Microb. Ecol.* **33**: 189–197.
- Chang, J., Masters, A., Avery, A., and Werren, J.H., (2010). A divergent *Cardinium* found in daddy long-legs (Arachnida: Opiliones). *J. Invertebr. Pathol.* **105**: 220–227.
- Chigira, A., and Miura, K., (2005). Detection of 'Candidatus Cardinium' bacteria from the haploid host *Brevipalpus californicus* (Acari: Tenuipalpidae) and effect on the host. *Exp. Appl. Acarol.* **37**: 107–116.
- Cook, J.M., and Butcher, R.D.J., (1999). The transmission and effects of *Wolbachia* bacteria in parasitoids. *Res. Popul. Ecol.* **41**: 15–28.
- Couso, J.P., and Gonzalez-Gaitan, M., (1993). Embryonic limb development in *Drosophila*. *Trends Genet.* **9**: 371–373.
- Craig, C.L., (1997). Evolution of arthropod silks. *Annu. Rev. Entomol.* **42**: 231–267.
- Currie, C.R., (2001). A community of ants, fungi, and bacteria: a multilateral approach to studying symbiosis. *Annu. Rev. Microbiol.* **55**: 357–380.
- Currie, C.R., Poulsen, M., Mendenahall, J., Boomsma, J.J., and Billen, J., (2006). Coevolved crypts and exocrine glands support mutualistic bacteria in fungus-growing ants. *Science* **311**: 81–83.
- Daborn, P.J., Yen, J.L., Bogwitz, M.R., Le Goff, G., Feil, E., and Jeffers, S., et al. (2002). A single P450 allele associated with insecticide resistance in *Drosophila*. *Science* **297**: 2253–2256.
- Dalton, R., (2000). Biologists flock to 'evo-devo' in a quest to read the recipes of life. *Nature* **403**: 125.
- Daly, H.V., Doyen, J.T., and Purcell III, A.H., (1998). *Introduction to Insect Biology and Diversity*, 2nd ed. Oxford Univ. Press, Oxford, UK.
- Daniels, S.B., Peterson, K.R., Strausbaugh, L.D., Kidwell, M.G., and Chovnick, A., (1990). Evidence for horizontal transmission of the P transposable element between *Drosophila* species. *Genetics* **124**: 339–355.
- Davey, M.E., and O'Toole, G.A., (2000). Microbial biofilms: from ecology to molecular genetics. *Microbiol. Mol. Biol. Rev.* **64**: 847–867.
- Davis, G.K., and Patel, N.H., (1999). The origin and evolution of segmentation. *Trends Genet.* **15**: M68–72.
- Davis, G.K., and Patel, N.H., (2002). Short, long, and beyond: molecular and embryological approaches to segmentation. *Annu. Rev. Entomol.* **47**: 669–699.
- DeGroote, F., Pont, G., Micard, D., and Picard, G., (1989). Extrachromosomal circular DNAs in *Drosophila melanogaster*: comparison between embryos and KcO% cells. *Chromosoma* **98**: 201–206.

- Delpuech, J.M., Aquadro, C.F., and Roush, R.T., (1993). Noninvolvement of the long terminal repeat of transposable element 17.6 in insecticide resistance in *Drosophila*. *Proc. Natl. Acad. Sci. USA* **90**: 5643–5647.
- De Robertis, E.M., and Sasai, Y., (1996). A common plan for dorsoventral patterning in Bilateria. *Nature* **380**: 37–40.
- Devonshire, A.L., and Field, L.M., (1991). Gene amplification and insecticide resistance. *Annu. Rev. Entomol.* **36**: 1–23.
- Dion, E., Polin, S.E., Simon, J.C., and Outreman, Y., (2011). Symbiont infection affects aphid defensive behaviours. *Biol. Lett.* **7**: 743–746.
- Dobson, S.L., Bourtzis, K., Braig, H.R., Jones, B.F., Zhou, W., Rousset, F., and O'Neill, S.C., (1999). *Wolbachia* infections are distributed throughout insect somatic and germ line tissues. *Insect Biochem. Mol. Biol.* **29**: 153–160.
- Douglas, A.E., (1992). Symbiotic microorganisms in insects. *Encycl. Microbiol.* **4**: 165–178.
- Douglas, A.E., (1998). Nutritional interactions in insect-microbial symbioses: aphids and their symbiotic bacteria. *Annu. Rev. Entomol.* **43**: 17–37.
- Douglas, A.E., (2010). *The Symbiotic Habit*. Princeton Univ. Press, Princeton, NJ.
- Dupuy, C., Huguet, E., and Drezen, J.M., (2006). Unfolding the evolutionary story of polydnaviruses. *Virus Res.* **117**: 81–89.
- Dupuy, C., Periquet, G., Serbielle, C., Bezier, A., Louis, F., and Drezen, J.M., (2011). Transfer of a chromosomal *Maverick* to endogenous bracovirus in a parasitoid wasp. *Genetica* **139**: 489–496.
- Durajlija Zinic, S., Ugarkovic, D., Cornudella, L., and Plohl, M., (2000). A novel interspersed type of organization of satellite DNAs in *Tribolium madens* heterochromatin. *Chromosome Res.* **8**: 201–212.
- Duron, O., Hurst, G.D.D., Hornett, E.A., Josling, J.A., and Engelstadter, J., (2008). High incidence of the maternally inherited bacterium *Cardinium* in spiders. *Mol. Ecol.* **17**: 1427–1437.
- Edgar, B.A., and Orr-Weaver, T.L., (2001). Endoreplication cell cycles: more for less. *Cell* **105**: 297–306.
- Eickbush, T.H., and Burke, W.D., (1985). Silkmoth chorion gene families contain patchwork patterns of sequence homology. *Proc. Natl. Acad. Sci. USA* **82**: 2814–2818.
- Eickbush, M.T., and Eickbush, T.H., (2011). Retransposition of R2 elements in somatic nuclei during the early development of *Drosophila*. *MobileDNA* **2**: 11.
- Eickbush, D.G., Eickbush, T.H., and Werren, J.H., (1992). Molecular characterization of repetitive DNA sequences from a B chromosome. *Chromosoma* **101**: 575–583.
- Enigl, M., and Schausberger, P., (2007). Incidence of the endosymbionts *Wolbachia*, *Cardinium* and *Spiroplasma* in phytoseiid mites and associated prey. *Exp. Appl. Acarol.* **42**: 75–85.
- Erwin, D., Valentine, J., and Jablonski, D., (1997). The origin of animal body plans. *Am. Sci.* **85**: 126–137.
- Fallon, A.M., (1984). Methotrexate resistance in cultured mosquito cells. *Insect Biochem.* **14**: 697–704.
- Feder, M.E., and Krebs, R.A.A., (1997). Ecological and evolutionary physiology of heat shock proteins and the stress response in *Drosophila*: complementary insights from genetic engineering and natural variation. In: Bijlsma, R., and Loeschke, V., Eds., *Environmental Stress, Adaptation and Evolution*, Birkhauser Verlag, Basel, Switzerland, pp. 155–173.
- Feschotte, C., and Pritham, E.J., (2007). DNA transposons and the evolution of eukaryotic genomes. *Annu. Rev. Genet.* **41**: 331–368.
- Field, L.M., (2000). Methylation and expression of amplified esterase genes in the aphid *Myzus persicae* (Sulzer). *Biochem. J.* **348**: 863–868.
- Finnegan, D.J., (1989). Eukaryotic transposable elements and genome evolution. *Trends Genet.* **5**: 103–107.

- Fitch, D.H.A., Strausbaugh, L.D., and Barrett, V., (1990). On the origins of tandemly repeated genes: Does histone gene copy number in *Drosophila* reflect chromosomal location? *Chromosoma* **99**: 118–124.
- Fleming, J.G.W., (1992). Polydnaviruses: mutualists and pathogens. *Annu. Rev. Entomol.* **37**: 401–425.
- Fleming, J.G.W., and Summers, M.D., (1991). Polydnavirus DNA is integrated in the DNA of its parasitoid wasp host. *Proc. Natl. Acad. Sci. USA* **88**: 9770–9774.
- Frank, J., (2000). The ribosome—a macromolecular machine par excellence. *Chem. Biol.* **7**: R133–141.
- Fraser, M.J., (1985). Transposon-mediated mutagenesis of baculoviruses: transposon shuttling and implications for speciation. *Ann. Entomol. Soc. Am.* **79**: 773–783.
- Furlong, E.E.M., Andersen, E.C., Null, B., White, K.P., and Scott, M.P., (2001). Patterns of gene expression during *Drosophila* mesoderm development. *Science* **293**: 1629–1633.
- Gaubatz, J.W., (1990). Extrachromosomal circular DNAs and genomic sequence plasticity in eukaryotic cells. *Mutat. Res.* **237**: 271–292.
- Gazla, I.N., and Carracedo, M.C., (2011). *Wolbachia* induces sexual isolation in *Drosophila melanogaster* and *Drosophila simulans*. *Open J. Genet.* **1**: 18–26.
- Gehring, W.J., (1985). The homeo box: a key to the understanding of development? *Cell* **40**: 3–5.
- Gehring, W.J., (1987). Homeo boxes in the study of development. *Science* **236**: 1245–1252.
- Gilbert, S.F., (2000). *Developmental Biology*, 6th ed. Sinauer, Sunderland, MA.
- Gilbert, C., Schaack, S., Pace, J.K., Brindley, P.J., and Feschotte, C., (2010). A role for host-parasite interactions in the horizontal transfer of transposons across phyla. *Nature* **464**: 1347–1350.
- Gillespie, J.P., Kanost, M.R., and Trenczek, T., (1997). Biological mediators of insect immunity. *Annu. Rev. Entomol.* **42**: 611–643.
- Goldsmith, M.R., Shimada, T., and Abe, H., (2005). The genetics and genomics of the silkworm *Bombyx mori*. *Annu. Rev. Entomol.* **50**: 71–100.
- Gomez-Orte, E., and Belles, X., (2009). MicroRNA-dependent metamorphosis in hemimetabolous insects. *Proc. Natl. Acad. Sci. USA* **106**: 21678–21682.
- Gooding, R.H., and Kraszna, E.F., (2005). Tsetse genetics: contributions to biology, systematics, and control of tsetse flies. *Annu. Rev. Entomol.* **50**: 101–123.
- Gotoh, T., Noda, H., and Ito, S., (2007). *Cardinium* symbionts cause cytoplasmic incompatibility in spider mites. *Heredity* **98**: 13–20.
- Gray, M.W., (1989). Origin and evolution of mitochondrial DNA. *Annu. Rev. Cell Biol.* **5**: 25–50.
- Groot, T.V.M., and Breeuwer, J.A.J., (2006). *Cardinium* symbionts induce haploid thelytoky in most clones of three closely related *Brevipalpus* species. *Exp. Appl. Acarol.* **(39)**: 257–271.
- Haig, D., (1993). The evolution of unusual chromosomal systems in sciarid flies: intragenomic conflict and the sex ratio. *J. Evol. Biol.* **6**: 249–261.
- Hales, D.F., Tomiuk, J., Wohrmann, K., and Sunnucks, P., (1997). Evolutionary and genetic aspects of aphid biology: a review. *Eur. J. Entomol.* **94**: 1–55.
- Hall, B.K., (2000). Evo-devo or devo-evo—does it matter? *Evol. Dev.* **2**: 177–178.
- Hankeln, T., and Schmidt, E.R., (1991). The organization, localization and nucleotide sequence of the histone genes of the midge *Chironomus thummi*. *Chromosoma* **101**: 25–31.
- Hansen, A.K., and Moran, N.A., (2011). Aphid genome expression reveals host-symbiont cooperation in the production of amino acids. *Proc. Natl. Acad. Sci. USA* **108**: 2849–2854.
- Hartl, D.L., and Brown, S.W., (1970). The origin of male haploid genetic systems and their expected sex ratio. *Theor. Popul. Biol.* **1**: 165–190.
- Hawley, R.S., and Walker, M.Y., (2000). *Drosophila melanogaster*: A genome sequenced. *Genome Res.* **10**: 593–594.
- Heath, B.D., Butcher, R.D.J., Whitfield, W.G.F., and Hubbard, S.F., (1999). Horizontal transfer of *Wolbachia* between phylogenetically distant insect species by a naturally occurring mechanism. *Curr. Biol.* **9**: 313–316.

- Heck, M.M.S., and Spradling, A.C., (1990). Multiple replication origins are used during *Drosophila* chorion gene amplification. *J. Cell Biol.* **110**: 903–914.
- Heckel, D.G., (1993). Comparative genetic linkage mapping in insects. *Annu. Rev. Entomol.* **38**: 381–408.
- Heddi, A., Grenier, A.M., Khatchadourian, C., Charles, H., and Nardon, P., (1999). Four intracellular genomes direct weevil biology: nuclear mitochondrial, principal endosymbiont and *Wolbachia*. *Proc. Natl. Acad. Sci. USA* **95**: 6814–6819.
- Heddi, A., Charles, H., and Khatchadourian, C., (2001). Intracellular bacterial symbiosis in the genus *Sitophilus*: the ‘biological individual’ concept revisited. *Res. Microbiol.* **152**: 431–437.
- Hemingway, J., Hawkes, N., Prapanthadara, L., Indrananda Jayawardena, K.G., and Ranson, H., (1998). The role of gene splicing, gene amplification and regulation in mosquito insecticide resistance. *Philos. Trans. R. Soc. Lond. B* **353**: 1695–1699.
- Holland, P.W.H., (1999). The future of evolutionary developmental biology. *Nature* **402**: C41–44.
- Hotopp, J.C.D., (2011). Horizontal gene transfer between bacteria and animals. *Trends Genet.* **27**: 157–163.
- Hotopp, J.C.D., Clark, M.E., Oliveira, D.C.S.G., Foster, J.M., Fischer, P., and Torres, M.C.M., et al. (2007). Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes. *Science* **317**: 1753–1756.
- Houck, M.A., Clark, J.B., Peterson, K.R., and Kidwell, M.G., (1991). Possible horizontal transfer of *Drosophila* genes by the mite *Proctolaelaps regalis*. *Science* **253**: 1125–1129.
- Hoy, M.A., (2000). Deploying transgenic arthropods in pest management programs: risks and realities. In: Handler, A.M., and James, A.A., Eds., *Insect Transgenesis. Methods and Applications*, CRC Press, Boca Raton, FL, pp. 335–367.
- Hughes, G.L., Koga, R., Xue, P., Fukatsu, T., and Rasgon, J.L., (2011). *Wolbachia* infections are virulent and inhibit the human malaria parasite *Plasmodium falciparum* in *Anopheles gambiae*. *PLoS Pathogens* **7** (5): e1002043.
- Huijser, P., Kirchhoff, C.J., Lankenau, D.H., and Hennig, W., (1988). Retrotransposon-like sequences are expressed in the Y chromosomal lampbrush loops of *Drosophila hydei*. *J. Mol. Biol.* **203**: 689–697.
- Hultmark, D., (1993). Immune reactions in *Drosophila* and other insects: a model for innate immunity. *Trends Genet.* **9**: 178–183.
- Hunter, M.S., Perlman, S.J., and Kelly, S.E., (2003). A bacterial symbiont in the *Bacteroidetes* induces cytoplasmic incompatibility in the parasitoid wasp *Encarsia pergandiella*. *Proc. R. Soc. Lond. B* **270**: 2185–2190.
- Hurst, G.D.D., and Schilthuizen, M., (1998). Selfish genetic elements and speciation. *Heredity* **80**: 2–8.
- Husseneder, C., (2010). Symbiosis in subterranean termites: a review of insights from molecular studies. *Environ. Entomol.* **39**: 378–388.
- Jabbari, K., and Bernardi, G., (2000). The distribution of genes in the *Drosophila* genome. *Gene* **247**: 287–292.
- Jakubczak, J.L., Burke, W.D., and Eickbush, T.H., (1991). Retrotransposable elements R1 and R2 interrupt the rRNA genes of most insects. *Proc. Natl. Acad. Sci. USA* **88**: 3295–3299.
- Janies, D., and DeSalle, R., (1999). Development, evolution and corroboration. *Anatom. Rec. (New Anat.)* **257**: 6–14.
- Jenner, R.A., (2000). Evolution of animal body plans: the role of metazoan phylogeny at the interface between pattern and process. *Evol. Dev.* **2**: 208–221.
- Jeyaprakash, A., and Hoy, M.A., (1995). Complete sequence of a *mariner* transposable element from the predatory mite *Metaseiulus occidentalis* isolated by an inverse PCR approach. *Insect Mol. Biol.* **4**: 31–39.
- Jeyaprakash, A., and Hoy, M.A., (2000). Long PCR improves *Wolbachia* DNA amplification: *wsp* sequences found in 76% of 63 arthropod species. *Insect Mol. Biol.* **9**: 393–405.

- Jurka, J., Kapitonov, V.V., Pavlicek, A., Klonowski, P., Kohany, O., and Walichiewicz, J., (2005). Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet. Genome Res.* **110**: 267–462.
- Kafatos, F.C., (1981). Structure, evolution and developmental expression of the silkworm chorion multigene families. *Am. Zool.* **21**: 707–714.
- Kafatos, F.C., Delidakis, C., Orr, W., Thireos, G., Komitopoulou, K., and Wong, Y., (1986). Studies on the developmentally regulated amplification and expression of *Drosophila* chorion genes *Molecular Developmental Biology*. Alan R. Liss, New York, 85–101.
- Kambris, Z., Blagborough, A.M., Pinto, S.B., Blagrove, M.S.C., Godfray, H.C.J., and Sinden, R.E., et al. (2010). *Wolbachia* stimulates immune gene expression and inhibits *Plasmodium* development in *Anopheles gambiae*. *PLoS Pathogens* **6** (10): e1001143.
- Kapitonov, V.V., and Jurka, J., (2001). Rolling-circle transposons in eukaryotes. *Proc. Natl. Acad. Sci. USA* **98**: 8714–8719.
- Kapitonov, V.V., and Jurka, J., (2003). Molecular paleontology of transposable elements in the *Drosophila melanogaster* genome. *Proc. Natl. Acad. Sci. USA* **94**: 196–201.
- Kapitonov, V.V., and Jurka, J., (2007). Helitrons on a roll: eukaryotic rolling-circle transposons. *Trends Genet.* **23**: 521–529.
- Kapitonov, V.V., and Jurka, J., (2008). A universal classification of eukaryotic transposable elements implemented in Repbase. *Nat. Rev. Genet.* **8**: 973–982.
- Kaufman, M.G., Walker, E.D., Odelson, D.A., and Klug, M.J., (2000). Microbial community ecology and insect nutrition. *Amer. Entomol.* **46**: 173–184.
- Kaulenas, M.S., (1985). Molecular Biology: protein synthesis In: Kerkut, G.A., and Gilbert, L.I., Eds., *Comprehensive Insect Physiology Biochemistry and Pharmacology*, Vol. 10 Pergamon, Oxford, UK, pp. 255–305.
- Kenyon, S.G., and Hunter, M.S., (2007). Manipulation of oviposition choice of the parasitoid wasp, *Encarsia pergandiella*, by the endosymbiotic bacterium *Cardinium*. *J. Evol. Biol.* **20**: 707–716.
- Kerrebrock, A.W., Srivastava, R., and Gerbi, S.A., (1989). Isolation and characterization of ribosomal DNA variants from *Sciara coprophila*. *J. Mol. Biol.* **210**: 1–13.
- Khush, R.S., and Lemaitre, B., (2000). Genes that fight infection: what the *Drosophila* genome says about animal immunity. *Trends Genet.* **16**: 442–468.
- Kidwell, M.G., (1992). Horizontal transfer. *Curr. Opin. Genet. Dev.* **2**: 868–873.
- Kikuchi, Y., Hayatsu, M., Hosokawa, T., Nagayama, A., Tago, K., and Fukatsu, T., (2012). Symbiont-mediated insecticide resistance. *Proc. Natl. Acad. Sci. USA* **109**: 8618–8622.
- King, R.E. Ed., 1975. *Handbook of Genetics*. Vol. 3. *Invertebrates of Genetic Interest*. Plenum. New York.
- Klingler, M., (1994). Segmentation in insects: how singular is *Drosophila*? *BioEssays* **16**: 391–392.
- Kobayashi, S., Amikura, R., and Okada, M., (1993). Presence of mitochondrial large ribosomal RNA outside mitochondria in germ plasm of *Drosophila melanogaster*. *Science* **260**: 1521–1524.
- Komaki, K., and Ishikawa, H., (1999). Intracellular bacterial symbionts of aphids possess many genomic copies per bacterium. *J. Mol. Evol.* **48**: 717–722.
- Komaki, K., and Ishikawa, H., (2000). Genomic copy number of intracellular bacterial symbionts of aphids varies in response to developmental stage and morph of their host. *Insect Biochem. Mol. Biol.* **30**: 253–258.
- Konig, H., and Varma, A., Eds., (2006). *Intestinal Microorganisms of Termites and Other Invertebrates*, Springer, New York.
- Konsolaki, M., Komitopoulou, K., Tolias, P.P., King, D.L., Swimmer, C., and Kafatos, F.C., (1990). The chorion genes of the medfly, *Ceratitis capitata* l; structural and regulatory conservation of the s36 gene relative to two *Drosophila* species. *Nucleic Acids Res.* **18**: 1731–1737.
- Krell, P.J., (1991). The polydnaviruses: multipartite DNA viruses from parasitic Hymenoptera. In: Kurstak, E. Ed., *Viruses of Invertebrates*. Marcel Dekker, New York, pp. 141–177.
- Langworthy, N.G., Renz, A., Mackenstedt, U., Henkle-Duhrsen, K., Bronsvort, M.B.D., and Tanya, V.N., et al. (2000). Macrofilaricidal activity of tetracycline against the filarial nematode

- Onchocerca ochengi*: elimination of *Wolbachia* precedes worm death and suggests a dependent relationship. *Proc. R. Soc. Lond. B* **267**: 1063–1069.
- Larsen-Rapport, E.W., (1986). Imaginal disc determination: molecular and cellular correlates. *Annu. Rev. Entomol.* **31**: 145–175.
- Lauzon, C.R., Sjogren, R.E., Wright, S.E., and Prokopy, R.J., (1998). Attraction of *Rhagoletis pomonella* (Diptera: Tephritidae) flies to odor of bacteria: apparent confinement to specialized members of Enterobacteriaceae. *Environ. Entomol.* **27**: 853–857.
- Lauzon, C.R., Sjogren, R.E., and Prokopy, R.J., (2000). Enzymatic capabilities of bacteria associated with apple maggot flies: a postulated role in attraction. *J. Chem. Ecol.* **26**: 953–967.
- Lecanidou, R., and Papantonis, A., (2010). Silkmoth chorion gene regulation revisited: promoter architecture as a key player. *Insect Mol. Biol.* **19**: 141–151.
- Lawrence, P.A., (1992). *The Making of a Fly The Genetics of Animal Design*. Blackwell Scientific, Oxford, UK.
- Lazzaro, B.P., and Rolff, J., (2011). Danger, microbes, and homeostasis. *Science* **332**: 43–44.
- Levin, H.L., and Moran, J.V., (2011). Dynamic interactions between transposable elements and their hosts. *Nature Rev. Genet.* **12**: 615–627.
- Lewis, E., (1978). A gene complex controlling segmentation in *Drosophila*. *Nature* **176**: 141–152.
- Lewis, E.B., (1994). Homeosis: the first 100 years. *Trends Genet.* **10**: 341–343.
- Lewis, Z., Champion de Crespigny, F.E., Salt, S.M., Tregenza, T., and Wedell, N., (2011). *Wolbachia* infection lowers fertile sperm transfer in a moth. *Biol. Lett.* **7**: 187–189.
- Lilburn, T.G., Schmidt, T.M., and Breznak, J.A., (1999). Phylogenetic diversity of termite gut spirochaetes. *Environ. Microbiol.* **1**: 331–345.
- Lu, L., Zhang, H., and Tower, J., (2001). Functionally distinct, sequence-specific replicator and origin elements are required for *Drosophila* chorion gene amplification. *Genes Dev.* **15**: 134–146.
- Lynch, J.A., El-Sherif, E., and Brown, S.J., (2012). Comparisons of the embryonic development of *Drosophila*, *Nasonia*, and *Tribolium*. *WIREs. Dev. Biol.* **1**: 16–39.
- Mannervik, M., (1999). Target genes of homeodomain proteins. *BioEssays* **21**: 267–270.
- Martin, W., (1999). A briefly argued case that mitochondria and plastids are descendants of endosymbionts, but the nuclear compartment is not. *Proc. R. Soc. Lond. B* **266**: 1387–1395.
- Martin, O.Y., and Goodacre, S.L., (2009). Widespread infections by the bacterial endosymbiont *Cardinium* in arachnids. *J. Arachnol.* **37**: 106–108.
- Mason, J.M., Haoudi, A., Konev, A.Y., Kurenova, E., Walter, M.F., and Biessmann, H., (2000). Control of telomere elongation and telomeric silencing in *Drosophila melanogaster*. *Genetica* **109**: 61–70.
- Masui, S., Kamoda, S., Sasaki, T., and Ishikawa, H., (1999). The first detection of the insertion sequence ISW1 in the intracellular reproductive parasite *Wolbachia*. *Plasmid* **42**: 13–19.
- Masui, S., Kamoda, S., Sasaki, T., and Ishikawa, H., (2000). Distribution and evolution of bacteriophage WO in *Wolbachia*, the endosymbiont causing sexual alterations in arthropods. *J. Mol. Evol.* **51**: 491–497.
- Mayer, M.P., and Bukau, B., (1999). Molecular chaperones: the busy life of *Hsp90*. *Curr. Biol.* **9**: R322–325.
- Miller, W.J., McDonald, J.F., and Pinsker, W., (1997). Molecular domestication of mobile elements. *Genetica* **100**: 261–270.
- Mitsialis, S.A., and Kafatos, F.C., (1985). Regulatory elements controlling chorion gene expression are conserved between flies and moths. *Nature* **317**: 453–456.
- Mitsialis, S.A., Veletza, S., and Kafatos, F.C., (1989). Transgenic regulation of moth chorion gene promoters in *Drosophila*: tissue, temporal, and quantitative control of four bidirectional promoters. *J. Mol. Evol.* **29**: 486–495.
- Mizrokhi, L.J., and Mazo, A.M., (1990). Evidence for horizontal transmission of the mobile element *jockey* between distant *Drosophila* species. *Proc. Natl. Acad. Sci. USA* **87**: 9216–9220.

- Moran, N.A., and Baumann, P., (2000). Bacterial endosymbionts in animals. *Curr. Opin. Microbiol.* **3**: 270–275.
- Moran, N.A., and Jarvik, T., (2010). Lateral transfer of genes from fungi underlies carotenoid production in aphids. *Science* **328**: 624–627.
- Morata, G., (2001). How *Drosophila* appendages develop. *Nat. Rev. Mol. Cell Biol.* **2**: 89–98.
- Morimoto, R.I., Sarge, K.D., and Abravaya, K., (1992). Transcriptional regulation of heat shock genes: a paradigm for inducible genomic responses. *J. Biol. Chem.* **31**: 21987–21990.
- Mouches, C., Pauplin, Y., Agarwal, M., Lemieux, L., Herzog, M., and Abadon, M., et al. (1990). Characterization of amplification core and esterase B1 gene responsible for insecticide resistance in *Culex*. *Proc. Natl. Acad. Sci. USA* **87**: 2574–2578.
- Mousseau, T.A., and Dingle, H., (1991). Maternal effects in insect life histories. *Annu. Rev. Entomol.* **36**: 511–534.
- Muller, G.B., (2007). Evo-devo: extending the evolutionary synthesis. *Nat. Rev. Genet.* **8**: 943–949.
- Nakamura, Y., Kawai, S., Yukihiko, F., Ito, S., Gotoh, T., and Kisimoto, R., et al. (2009). Prevalence of *Cardinium* bacteria in planthoppers and spider mites and taxonomic revision of “*Candidatus Cardinium hertigii*” based on detection of a new *Cardinium* group from biting midges. *Appl. Environ. Microbiol.* **75**: 6757–6763.
- Nakamura, Y., Gotoh, T., Imanishi, S., Mita, K., Kurtti, T.J., and Noda, H., (2011). Differentially expressed genes in silkworm cell cultures in response to infection by *Wolbachia* and *Cardinium* endosymbionts. *Insect Mol. Biol.* **20**: 279–289.
- Nam, J., and Nei, M., (2005). Evolutionary change in the numbers of homeobox genes in bilateral animals. *Mol. Biol. Evol.* **22**: 2386–2394.
- Nikoh, N., McCutcheon, J.P., Kudo, T., Miyagishima, S., Moran, N.A., and Nakabachi, A., (2010). Bacterial genes in the aphid genome: absence of functional gene transfer from *Buchnera* to its host. *PLoS Genet.* **6** (2): e1000827.
- Nur, U., (1990). Heterochromatization and euchromatization of whole genomes in scale insects (Coccoidea: Homoptera). *Development* **1990 (Suppl)**: 29–34.
- Nusslein-Volhard, C., and Wieschaus, E., (1980). Mutations affecting segment number and polarity in *Drosophila*. *Nature* **187**: 795–801.
- Ohno, S., and Yomo, T., (1991). The grammatical rule for all DNA: Junk and coding sequences. *Electrophoresis* **12**: 103–108.
- Okazaki, S., Tsuchida, K., Maekawa, H., Ishikawa, H., and Fujiwara, H., (1993). Identification of a pentanucleotide telomeric sequence, (TTAGG)<sub>n</sub>, in the silkworm *Bombyx mori* and in other insects. *Mol. Cell. Biol.* **13**: 1424–1432.
- Oliver, K.M., Degnan, P.H., Burke, G.R., and Moran, N.A., (2010). Facultative symbionts in aphids and the horizontal transfer of ecologically important traits. *Annu. Rev. Entomol.* **55**: 247–266.
- O'Neill, S.L., Giordano, R., Colbert, A.M.E., Karr, T.L., and Robertson, H.M., (1992). 16S rRNA phylogenetic analysis of the bacterial symbiont associated with cytoplasmic incompatibility in insects. *Proc. Natl. Acad. Sci. USA* **89**: 2699–2702.
- O'Neill, S.L., Hoffmann, A.A., and Werren, J.H., Eds., (1997). *Influential Passengers. Inherited Microorganisms and Arthropod Reproduction*, Oxford Univ. Press, Oxford, UK.
- Orr-Weaver, T.L., (1991). *Drosophila* chorion genes: cracking the eggshell's secrets. *BioEssays* **113**: 97–105.
- Osada, N., and Innan, H., (2008). Duplication and gene conversion in the *Drosophila melanogaster* genome. *PLoS Genet.* **4** (12): e1000305.
- Otto, P.A., (2000). *Drosophila* viewer: a program on the formal genetics, anatomy and developmental biology of *Drosophila melanogaster* for students and specialists. *Gen. Mol. Biol.* **23**: 835–839.
- Otto, S.P., and Whitton, J., (2000). Polyploid incidence and evolution. *Annu. Rev. Genet.* **34**: 401–437.
- Pan, X., Zhou, G., Wu, J., Bian, G., Lu, P., Raikhel, A.S., and Xi, Z., (2012). *Wolbachia* induces reactive oxygen species (ROS)-dependent activation of the Toll pathway to control dengue virus in the mosquito *Aedes aegypti*. *Proc. Natl. Acad. Sci. USA* **209**: e23–E31.

- Pardue, M.L., (1991). Dynamic instability of chromosomes and genomes. *Cell* **66**: 427–431.
- Pardue, M.L., and Hennig, W., (1990). Heterochromatin: junk or collectors item? *Chromosoma* **100**: 3–7.
- Pasteur, N., and Raymond, M., (1996). Insecticide resistance genes in mosquitoes: their mutations, migration, and selection in field populations. *J. Hered.* **87**: 444–449.
- Paton, M.G., Karunaratne, S.H.P.P., Giakoumaki, E., Roberts, N., and Hemingway, J., (2000). Quantitative analysis of gene amplification in insecticide-resistant *Culex* mosquitoes. *Biochem. J.* **346** (17–24).
- Pauli, D., Arrigo, A.P., and Tissieres, A., (1992). Heat shock response in *Drosophila*. *Experientia* **48**: 623–629.
- Pennisi, E., and Roush, W., (1997). Developing a new view of evolution. *Science* **277**: 34–37.
- Perrimon, N., Mohler, D., Engstrom, L., and Mahowald, A., (1986). X-linked female-sterile loci in *Drosophila melanogaster*. *Genetics* **113**: 695–712.
- Pinto-Tomas, A.A., Anderson, M.A., Suen, G., Stevenson, D.M., Chu, F.S.T., and Cleland, W.W., et al. (2009). Symbiotic nitrogen fixation in the fungus gardens of leaf-cutter ants. *Science* **326**: 1120–1123.
- Pistone, D., Sacchi, L., Lo, N., Epis, S., Pajoro, M., and Favia, G., et al. (2012). *Candidatus Midichloria mitochondrii*, symbiont or parasite of tick mitochondria? In: Zchori-Fein, E., and Bourtzis, K., Eds., *Manipulative Tenants Bacteria Associated with Arthropods*, CRC Press, Boca Raton, FL, pp. 113–126.
- Plasterk, R.H.A., Izsvák, Z., and Ivics, Z., (1999). Resident aliens. The *Tc1/mariner* superfamily of transposable elements. *Trends Genet.* **15**: 326–332.
- Poulsen, M., and Currie, C.R., (2010). Symbiont interactions in a tripartite mutualism: exploring the presence and impact of antagonism between two fungus-growing ant mutualists. *PLoS One* **5** (1): e8748.
- Powell, J.R., (1997). *Progress and Prospects in Evolutionary Biology: The Drosophila Model*. Oxford Univ. Press, New York.
- Prody, C.A., Dreyfus, P., Zamir, R., Zakut, H., and Soreq, H., (1989). *De novo* amplification within a “silent” human cholinesterase gene in a family subjected to prolonged exposure to organophosphorous insecticides. *Proc. Natl. Acad. Sci. USA* **86**: 690–694.
- Prudhomme, J.C., Couble, P., Garel, J.P., and Daillie, J., (1985). Silk synthesis, pp. 571–94 In: Kerkut, G.A., and Gilbert, L.I., Eds., *Comprehensive Insect Physiology Biochemistry and Pharmacology*, Vol. 10, Pergamon, Oxford, UK.
- Puinean, A.M., Foster, S.P., Oliphant, L., Denholm, I., Field, L.M., and Millar, N.S., et al. (2010). Amplification of a cytochrome P450 gene is associated with resistance to neonicotinoid insecticides in the aphid *Myzus persicae*. *PLoS Genet.* **6** (6): e1000999.
- Raihel, A.S., and Dhadialla, T.S., (1992). Accumulation of yolk proteins in insect oocytes. *Annu. Rev. Entomol.* **37**: 217–251.
- Retnakaran, A., and Percy, J., (1985). Fertilization and special modes of reproduction In: Kerkut, G.A., and Gilbert, L.I., Eds., *Comprehensive Insect Physiology Biochemistry and Pharmacology*, Vol. 1, Pergamon, Oxford, UK, pp. 231–293.
- Richard, F.J., Poulsen, M., Hefetz, A., Errard, C., Nash, D.R., and Boomsma, J.J., (2007). The origin of the chemical profiles of fungal symbionts and their significance for nestmate recognition in *Acromyrmex* leaf-cutting ants. *Behav. Ecol. Sociobiol.* **61**: 1637–1649.
- Rigaud, T., (1999). Further *Wolbachia* endosymbiont diversity: a tree hiding in the forest? *Trends Ecol. Evol.* **14**: 212–213.
- Rigaud, T., and Rousset, F., (1996). What generates the diversity of *Wolbachia*-arthropod interactions? *Biodiver. Conserv.* **5**: 999–1013.
- Rina, M., and Savakis, C., (1991). A cluster of vitellogenin genes in the Mediterranean fruit fly *Ceratitis capitata*: sequence and structural conservation in dipteran yolk proteins and their genes. *Genetics* **127**: 769–780.
- Rinderer, T.E. Ed., 1986. *Bee Genetics and Breeding*. Academic Press. Orlando, FL.

- Robertson, H.M., (1993). The *mariner* transposable element is widespread in insects. *Nature* **362**: 241–245.
- Robertson, H.M., and Lampe, D.J., (1995). Distribution of transposable elements in arthropods. *Annu. Rev. Entomol.* **40**: 333–357.
- Robinson, R., (1971). *Lepidoptera Genetics*. Pergamon, Oxford, UK.
- Robinson, G.E., Hackett, K.J., Purcell-Miramontes, M., Brown, S.J., Evans, J.D., and Goldsmith, M.R., et al. (2011). Creating a buzz about insect genomes. *Science* **331**: 1386.
- Rokas, A., (2000). *Wolbachia* as a speciation agent. *Trends Ecol. Evol.* **15**: 44–45.
- Ros, V.I.D., and Breeuwer, J.A.J., (2009). The effects of, and interactions between, *Cardinium* and *Wolbachia* in the doubly infected spider mite *Bryobia sarothonami*. *Heredity* **102**: 413–422.
- Roth, S., and Hartenstein, V., (2008). Development of *Tribolium castaneum*. *Dev. Genes Evol.* **218**: 115–118.
- Rutherford, S.L., and Lindquist, S., (1998). Hsp90 as a capacitor for morphological evolution. *Nature* **396**: 336–342.
- Sapienza, C., Peterson, A.C., Rossant, J., and Balling, R., (1987). Degree of methylation of transgenes is dependent on gamete of origin. *Nature* **328**: 251–254.
- Schaack, S., Gilbert, C., and Feschotte, C., (2010). Promiscuous DNA: horizontal transfer of transposable elements and why it matters for eukaryotic evolution. *Trends Ecol. Evol.* **25**: 537–546.
- Schroder, R., Beermann, A., Wittkopp, N., and Lutz, R., (2008). From development to biodiversity—*Tribolium castaneum*, an insect model organism for short germband development. *Dev. Genes Evol.* **218**: 119–126.
- Schwemmler, W., and Gassner, G., Eds., (1989). *Insect Endocytobiosis: Morphology, Physiology, Genetics, Evolution*, CRC Press, Boca Raton, FL.
- Scott, J.J., Oh, D.C., Yuceer, M.C., Klepzig, K.D., Clardy, J., and Currie, C.R., (2008). Bacterial protection of beetle-fungus mutualism. *Science* **322**: 63.
- Sezutsu, H., and Yukihiko, K., (2000). Dynamic rearrangement within the *Antherea pernyi* silk fibroin gene is associated with four types of repetitive units. *J. Mol. Evol.* **51**: 329–338.
- Shapiro, J.A., (1999). Transposable elements as the key to a 21st century view of evolution. *Genetica* **107**: 171–179.
- Shelby, K.S., and Webb, B.A., (1999). Polydnavirus-mediated suppression of insect immunity. *J. Insect Physiol.* **45**: 507–514.
- Shigenobu, S., Watanabe, H., Hattori, M., Sakaki, Y., and Ishikawa, H., (2000). Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. APS. *Nature* **407**: 81–86.
- Shippy, T.D., Ronshaugen, M., Cande, J., He, J.P., Beeman, R.W., and Levine, M., et al. (2008). Analysis of the *Tribolium* homeotic complex: insights into mechanisms constraining insect Hox clusters. *Dev. Genes Evol.* **218**: 127–139.
- Shoemaker, D.D., Katju, V., and Jaenike, J., (1999). *Wolbachia* and the evolution of reproductive isolation between *Drosophila recens* and *Drosophila subquadrata*. *Evolution* **53**: 1157–1164.
- Simon, J.C., Boutin, S., Tsukada, T., Koga, R., Le Gallic, J.F., and Frantz, A., et al. (2011). Facultative symbiont infections affect aphid reproduction. *PLoS ONE* **6** (7): e21831.
- Snook, R.R., Cleland, S.Y., Wolfner, M.F., and Karr, T.L., (2000). Offsetting effects of *Wolbachia* infection and heat shock on sperm production in *Drosophila simulans*: analyses of fecundity, fertility and accessory gland proteins. *Genetics* **155**: 167–178.
- Sokoloff, A., (1966). *The Genetics of Tribolium and Related Species*. Academic, New York.
- Sokoloff, A., (1977). *The Biology of Tribolium, with Special Emphasis on Genetic Aspects*, Vol. 3. Clarendon, Oxford, UK.
- Solter, D., (1988). Differential imprinting and expression of maternal and paternal genomes. *Annu. Rev. Genet.* **22**: 127–146.
- Stark, G.R., and Wahl, G.M., (1984). Gene amplification. *Annu. Rev. Biochem.* **53**: 91–447.

- Steiner, W.W.M., Tabachnick, W.J., Rai, K.S., and Narang, S., Eds., (1982). *Recent Developments in the Genetics of Insect Disease Vectors*, Stipes Publishing, Champaign, IL.
- Stoltz, D., and Whitfield, J.B., (1992). Viruses and virus-like entities in the parasitic Hymenoptera. *J. Hymenopt. Res.* **1**: 125–139.
- Sugimoto, T.N., and Ishikawa, Y., (2012). A male-killing *Wolbachia* carries a feminizing factor and is associated with degradation of the sex-determining system of its host. *Biol. Lett.* **8** (3): 412–415.
- Sun, L.V., Foster, J.M., Tzertzinis, G., Ono, M., Bandi, C., Slatko, B.E., and O'Neill, S.L., (2001). Determination of *Wolbachia* genome size by pulse-field gel electrophoresis. *J. Bacteriol.* **183**: 2219–2225.
- Sutherland, T.D., Young, J.H., Weisman, S., Hayashi, C.Y., and Merritt, D.J., (2010). Insect silk: one name, many materials. *Annu. Rev. Entomol.* **55**: 171–188.
- Tazima, Y., (1964). *Genetics of the Silkworm*. Logos, London, UK.
- Tazima, Y., Doira, H., and Akai, H., (1975). The domesticated silkworm, *Bombyx mori*. In: King, R.C. Ed., *Handbook of Genetics. Vol. 3. Invertebrates of Genetic Interest*. Plenum. New York, pp. 63–124.
- Teixeira, L., Ferreira, A., and Ashburner, M., (2008). The bacterial symbiont *Wolbachia* induces resistance to RNA viral infections in *Drosophila melanogaster*. *PLoS Biol.* **6**: 2e2.
- Thomas, B.J., and Wassarman, D.A., (1999). A fly's eye view of biology. *Trends Genet.* **15**: 184–190.
- Thomas, J., Schaack, S., and Pritham, E.J., (2010). Pervasive horizontal transfer of rolling-circle transposons among animals. *Genome Biol. Evol.* **2**: 656–664.
- Tram, U., and Sullivan, W., (2002). Role of delayed nuclear envelope breakdown and mitosis in *Wolbachia*-induced cytoplasmic incompatibility. *Science* **296**: 1124–1126.
- Tsuchida, T., Koga, R., Horikawa, M., Tsunoda, T., Maoka, T., and Matsumoto, S., et al. (2010). Symbiotic bacterium modifies aphid body color. *Science* **330**: 1102–1104.
- Tsuchida, T., Koga, R., Matsumoto, S., and Fukatsu, T., (2011). Interspecific symbiont transfection confers a novel ecological trait to the recipient insect. *Biol. Lett.* **7**: 245–248.
- Trewitt, P.M., Heilmann, L.J., Degruyillier, S.S., and Kumaran, A.K., (1992). The boll weevil vitellogenin gene: nucleotide sequence, structure, and evolutionary relationship to nematode and vertebrate vitellogenin genes. *J. Mol. Evol.* **34**: 478–492.
- van Meer, M.M.M., and Stouthamer, R., (1999). Cross-order transfer of *Wolbachia* from *Muscidifurax uniraptor* (Hymenoptera: Pteromalidae) to *Drosophila simulans* (Diptera: Drosophilidae). *Heredity* **82**: 163–169.
- von Dassow, G., and Munro, E., (1999). Modularity in animal development and evolution: elements of a conceptual framework for EvoDevo. *J. Exp. Zool. (Mol. Dev. Evol.)* **285**: 307–325.
- von Dohlen, D.C., Kohler, S., Alsop, S.T., and McManus, W.R., (2001). Mealybug β-proteobacterial endosymbionts contain γ-proteobacterial symbionts. *Nature* **412**: 433–435.
- Wagner, R.P., Maguire, M.P., and Stallings, R.L., (1993). *Chromosomes: A Synthesis*. Wiley-Liss, New York.
- Walser, C.B., and Lipshitz, H.D., (2011). Transcript clearance during the maternal-to-zygotic transition. *Curr. Opin. Gen. Dev.* **21**: 431–443.
- Warnecke, F., Luginbuhl, P., Ivanova, N., Ghassanian, M., Richardson, T.H., and Stege, J.T., et al. (2007). Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature* **450**: 560–565.
- Waters, L.C., Zelhof, A.C., Shaw, B.J., and Chang, L.Y., (1992). Possible involvement of the long terminal repeat of transposable element 17.6 in regulating expression of an insecticide resistance-associated P450 gene in *Drosophila*. *Proc. Natl. Acad. Sci. USA* **89**: 4855–4859.
- Webb, B.A., (1998). Polydnavirus biology, genome structure and evolution. In: Miller, L.K., and Ball, A., Eds., *The Insect Viruses*, Plenum, New York, pp. 105–139.

- Webb, B.A., and Cui, L., (1998). Relationships between polydnavirus genomes and viral gene expression. *J. Insect Physiol.* **44**: 785–793.
- Weeks, A.R., and Breeuwer, J.A.J., (2001). *Wolbachia*-induced parthenogenesis in a genus of phytophagous mites. *Proc. R. Soc. Lond. B* **268**: 2245–2251.
- Weiss, B., and Aksoy, S., (2011). Microbiome influences on insect host vector competence. *Trends Parasitol.* **27**: 514–522.
- Welburn, S.C., Arnold, K., Maudlin, I., and Gooday, G.W., (1993). Rickettsia-like organisms and chitinase production in relation to transmission of trypanosomes by tsetse flies. *Parasitology* **107**: 141–145.
- Werren, J.H., (2011). Selfish genetic elements, genetic conflict and evolutionary innovation. *Proc. Natl. Acad. Sci. USA* **108** (Suppl. 2): 10863–10870.
- Werren, J.H., Guo, L.R., and Windsor, D.W., (1995). Distribution of *Wolbachia* in Neotropical arthropods. *Proc. R. Soc. Lond. B* **262**: 197–204.
- Werren, J.H., Richards, S., Desjardins, C.A., Niehuls, O., Gadau, J., Colbourne, J.K., and The Nasonia Genome Working Group, 2010. Functional and evolutionary insights from the genomes of three parasitoid *Nasonia* species. *Science* **327**: 343–348.
- West-Eberhard, M.J., (1998). Evolution in the light of developmental and cell biology, and vice versa. *Proc. Natl. Acad. Sci. USA* **95**: 8417–8419.
- White, M.J.D., (1973). *Animal Cytology and Evolution*, 3rd ed. Cambridge Univ. Press, Cambridge, UK.
- White, K.P., Rifkin, S.A., Hurban, P., and Hogness, D.S., (1999). Microarray analysis of *Drosophila* development during metamorphosis. *Science* **286**: 2179–2184.
- Whitfield, J.B., (1997). Molecular and morphological data suggest a single origin of the polydnaviruses among braconid wasps. *Naturwissenschaften* **84**: 502–507.
- Whitfield, J.B., and Purcell III, A.H., (2013). *Daly and Doyen's Introduction to Insect Biology and Diversity*, 3d ed. Oxford Univ. Press, New York.
- Wilkins, A.S., (1993). *Genetic Analysis of Animal Development*, 2nd ed. Wiley-Liss, New York.
- Williams, S.M., and Robbins, L.G., (1992). Molecular genetic analysis of *Drosophila* rDNA arrays. *Trends Genet.* **8**: 335–340.
- Wilson, T.G., (1993). Transposable elements as initiators of insecticide resistance. *J. Econ. Entomol.* **86**: 645–651.
- Wolf, K.W., Mertl, H.G., and Traut, W., (1991). Structure, mitotic and meiotic behaviour, and stability of centromere-like elements devoid of chromosome arms in the fly *Megaselia scalaris* (Phoridae). *Chromosoma* **101**: 99–108.
- Wright, J.W., and Pal, R., Eds., (1967). *Genetics of Insect Vectors of Disease*, Elsevier, Amsterdam, The Netherlands.
- Wu, K., and Hoy, M.A., (2012). *Cardinium* is associated with reproductive incompatibility in the predatory mite *Metaseiulus occidentalis* (Acari: Phytoseiidae). *J. Invertebr. Pathol.* **110**: 359–365.
- Xiong, Y., Burke, W.D., and Eickbush, T.H., (1993). Pao, a highly divergent retrotransposable element from *Bombyx mori* containing long terminal repeats with tandem copies of the putative R region. *Nucleic Acids Res.* **21**: 2117–2123.
- Ye, J., Perez-Gonzalez, C.E., Eickbush, D.G., and Eickbush, T.H., (2005). Competition between R1 and R2 transposable elements in the 28S rRNA genes of insects. *Cytogenet. Genome Res.* **110**: 299–306.
- Yoshida, N., Oeda, K., Watanabe, E., Mikami, T., Fukita, Y., and Nishimura, K., et al. (2001). Chaperonin turned insect toxin. *Nature* **411**: 44.
- Yuan, Y.W., and Wessler, S.R., (2011). The catalytic domain of all eukaryotic cut-and-paste transposase superfamilies. *Proc. Natl. Acad. Sci. USA* **108**: 7884–7889.
- Zchori-Fein, E., and Bourtzis, K., Eds., (2011). *Manipulative Tenants: Bacteria Associated with Arthropods*, CRC Press, Boca Raton, FL.

- Zchori-Fein, E., and Perlman, S.J., (2004). Distribution of the bacterial symbiont *Cardinium* in arthropods. *Mol. Ecol.* **13**: 2009–2016.
- Zchori-Fein, E., Gottlieb, Y., Kelly, S.E., Brown, J.K., Wilson, J.M., and Karr, T.L., et al. (2001). A newly discovered bacterium associated with parthenogenesis and a change in host selection behavior in parasitoid wasps. *Proc. Natl. Acad. Sci. USA* **98**: 12555–12560.
- Zchori-Fein, E., Perlman, S.J., Kelly, S.E., Katzir, N., and Hunter, M.S., (2004). Characterization of a 'Bacteroidetes' symbiont in *Encarsia* wasps (Hymenoptera: Aphelinidae): proposal of *Candidatus Cardinium hertigii*. *Int. J. Syst. Evol. Microbiol.* **54**: 961–968.
- Zhang, Y., Zhou, X., Ge, X., Jiang, J., Li, M., and Jia, S., et al. (2009). Insect-specific microRNA involved in the development of the silkworm *Bombyx mori*. *PLoS ONE* **4** (3): e4677.
- Zhou, W., Rousset, F., and O'Neill, S.L., (1998). Phylogeny and PCR-based classification of *Wolbachia* strains using *wsp* gene sequences. *Proc. R. Soc. Lond. B* **265**: 509–515.

# Some Basic Tools: How to Cut, Paste, Copy, Measure, Visualize, and Clone DNA

## Chapter Outline

5.1 Overview	183
5.2 Introduction to a Basic Molecular Biology Experiment	184
5.2.1 The Pros and Cons of Kits	185
5.2.2 A Simple Cloning Experiment	186
5.3 Extracting DNA from Insects	187
5.3.1 DNA Extraction Resulting in Loss of the Specimens	187
5.3.2 DNA Extraction That Does Not Require Destroying the Specimens	188
5.3.3 Assessing the Quality of Extracted DNA	189
5.4 Precipitating Nucleic Acids	190
5.5 Shearing DNA	192
5.6 Cutting DNA with Restriction Endonucleases	192
5.7 Joining DNA Molecules	195
5.8 Growth, Maintenance, and Storage of <i>E. coli</i>	196
5.9 Plasmids for Cloning in <i>E. coli</i>	196
5.10 Transforming <i>E. coli</i> with Plasmids	201
5.11 Purifying Plasmid DNA from <i>E. coli</i>	202
5.12 Electrophoresis in Agarose or Acrylamide Gels	204
5.13 Detecting, Viewing, and Photographing Nucleic Acids in Gels	206
5.14 Identifying Specific DNA by Southern Blot Analysis	206
5.15 Labeling DNA or RNA Probes	209
5.16 Removing DNA from Agarose Gels after Electrophoresis	210
5.17 Restriction-Site Mapping	211
General References	212
References Cited	213

## 5.1 Overview

Genetic engineers use several techniques to isolate DNA, cut and join molecules, and monitor the results. To clone a gene, determine its sequence, or alter the genetic makeup of an arthropod, microbiological techniques are used,

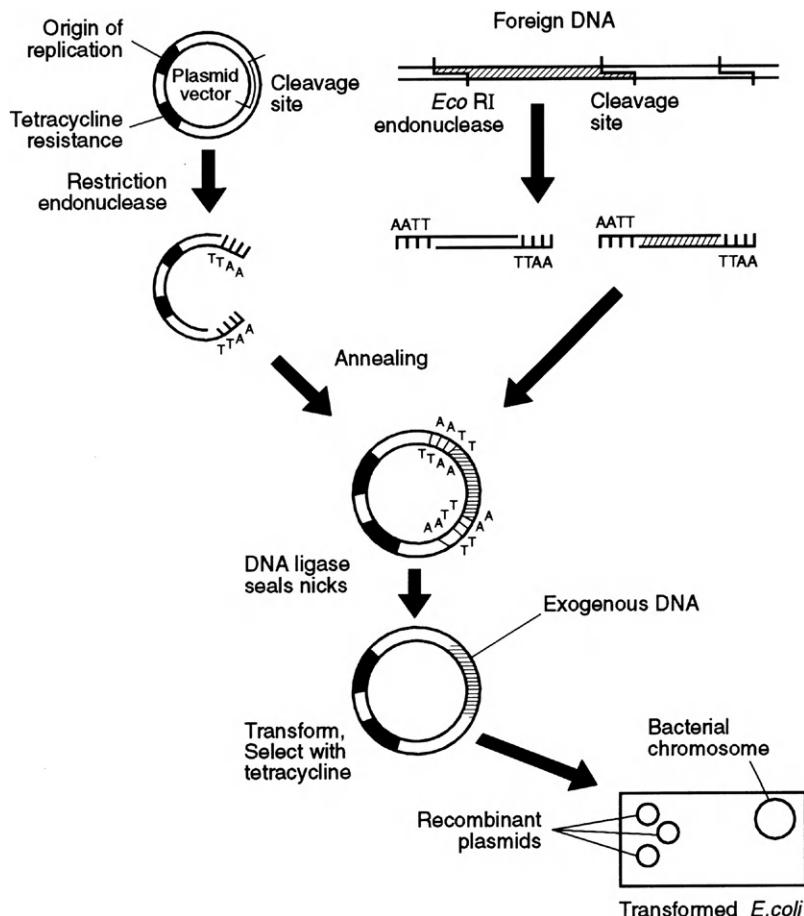
particularly with the bacterium *Escherichia coli*. This bacterium has become a molecular biology workhorse because it can be induced to produce large amounts of recombinant DNA molecules by inserting plasmids, the bacteriophage  $\lambda$  or genetically engineered variants of these agents into it. Genetic engineers use enzymes from different organisms to modify, ligate, or splice DNA. Basic tools include purifying plasmids from *E. coli*, visualizing DNA by electrophoresis through agarose or polyacrylamide gels, Southern blot analyses, and producing labeled probes. This chapter describes the steps involved in inserting a foreign gene into a plasmid, inserting the plasmid into *E. coli*, and isolating and analyzing the amplified DNA by Southern blot analysis and restriction-site mapping. "Northern" and "Western" blot analyses allow the researcher to evaluate RNA and proteins, respectively. Many of these procedures have been simplified by the availability of commercial kits, but understanding the concepts behind them will allow you to "trouble shoot." Cloning of DNA has five essential components: 1) a method for generating exogenous DNA fragments, 2) reactions to join exogenous DNA fragments to a vector, 3) a method to introduce the vector into a host cell where the vector ensures the exogenous DNA is replicated, 4) methods for selecting or identifying the vectors that contain the introduced DNA (recombinant molecules), and 5) methods for analyzing the cloned DNA.

## 5.2 Introduction to a Basic Molecular Biology Experiment

Molecular genetic techniques, some arising from research on basic topics, have become crucial tools that allow scientists to manipulate DNA from living organisms. As a result, it is difficult to predict to what use some basic research can be applied. The molecular genetic revolution began in 1970. Before 1970, there was no way to cut a DNA molecule into discrete and predictable fragments, nor could specific DNA fragments be joined together. The discovery of enzymes called **restriction endonucleases** and **ligases** changed this. Much of genetic engineering technology is dependent upon our ability to cut DNA molecules at specific sites and combine them into new molecules by base pairing and ligation.

Another significant development was the harnessing of **plasmids** and **bacteriophages** as vehicles (**vectors**) to replicate foreign DNA within the bacterium *Escherichia coli*, which allowed nearly unlimited amounts of specific DNA to be produced for study and manipulation. Techniques for monitoring the results of such manipulations were developed so that researchers could identify as small as a single-base modification in DNA.

A simple cloning project, as would have been conducted in 1985 when many of the current kits were unavailable, is outlined in [Figure 5.1](#). The methods



**Figure 5.1** Outline of a simple cloning project involving insertion of foreign DNA into a plasmid vector, and its subsequent insertion into a bacterial host, *E. coli*, to produce multiple copies (clones) of the foreign DNA in recombinant plasmids. Both foreign and plasmid DNA is cut with a restriction endonuclease to produce ends that will allow annealing of the plasmid and foreign DNA fragments. The addition of DNA ligase combines the two DNA molecules and the plasmid is inserted into *E. coli* where it will produce multiple copies (clones) of the new, recombinant DNA molecule. Subsequently, the plasmids will be extracted from their host cells and used for other studies.

described provide the theoretical background to many of the current kit protocols, although companies may change details or may not explain the steps or chemical components for proprietary reasons.

### 5.2.1 The Pros and Cons of Kits

Understanding the theoretical bases of the techniques described in this chapter provides a background helpful in understanding the procedures one carries out

with commercially available kits. As noted in *Nature Methods* (2011), “Modern molecular biology would be impossible without commercial kits.” These kits are essential for isolating DNA and plasmids, extracting RNA, conducting restriction digests, cloning, and many other tasks. Kits save time and may be cost-effective. Only a relatively few years ago, molecular biologists had to produce their own enzymes, make buffers, and spend a great deal of time doing very basic biochemistry. However, although kits save valuable time, “The use of kits is blamed for undercutting researchers’ ability to recognize artifacts and for making young scientists more inclined to trust their results than to question them,” and “Any self-respecting scientist should be able to say what is happening at every step of a protocol” (*Nature Methods* 2011).

When using kits, carefully read the methods provided by the manufacturer. Manufacturers should explain what to do at each step, what is happening, and what the reagents are. When choosing a kit, consider the quality of the information provided and any troubleshooting information. Examine each step before you begin to be sure you understand what you will be doing, and why. If you have questions, most companies provide a “hot line” that you can telephone or an email address for advice and information. Because improvements in kits are made regularly, it is a good idea to contact the manufacturer’s website to check on methods updates; some won’t be in the printed directions. Once in a very rare while, kits will not perform properly because they are defective, so don’t hesitate to contact the company if you have problems.

### 5.2.2 A Simple Cloning Experiment

The project described here is simple and basic (Figure 5.1). It involves inserting a piece of **exogenous** (or foreign) DNA extracted from one organism into a plasmid that has been engineered to serve as a **vector** to carry the exogenous DNA into *E. coli*. The *E. coli* cells with the exogenous DNA (in the plasmid vector) will be able to yield large numbers of the desired DNA molecules. Once the DNA has been multiplied (**cloned**), the exogenous DNA can be studied in detail.

Although the experiment in Figure 5.1 conceptually is simple, cloning a fragment of foreign DNA into a vector demands that several steps be achieved: 1) The circular vector DNA must be purified and cut. 2) The exogenous DNA must be extracted, purified, and cut. 3) The vector DNA and exogenous DNA must be joined together. 4) The reactions should be monitored. 5) The recombinant plasmid or vector containing the exogenous DNA must be put back into *E. coli* to be multiplied. 6) The recombinant plasmid then must be removed from *E. coli* and purified for analysis or use of the exogenous DNA.

This chapter provides an introduction to the procedures that could be used to carry out the experiment illustrated in [Figure 5.1](#). Simplified protocols of some procedures are provided for those interested in knowing something about the steps involved, although the methods described are illustrative rather than complete laboratory protocols. There are many excellent laboratory manuals available that provide detailed techniques (see General References) as well as websites with up-to-the-minute protocols. As noted, many techniques have been simplified and are available in kits provided by commercial sources. This led one molecular biologist to conclude that the availability of kits has led to a “paradigm shift” in biology ([Gilbert 1991](#)).

## 5.3 Extracting DNA from Insects

Most DNA extraction protocols require that the insect or specific tissues be ground up. However, methods have been developed that allow the specimen to be preserved (Section 5.3.2). Whatever method is used to extract the DNA from insects, it must be purified before it can be used in most molecular biology projects. The degree of purity needed is determined by the goals of the experiment. Any time DNA is to be cut, or ligated, or cloned, it must be *very clean*. Proteins, lipids, sugars and nucleases that could interfere with these procedures should be removed. Insects should be killed rapidly so that nucleases (DNases or RNases) are not allowed to damage the DNA or RNA being extracted. Insects can be stored in 95% ethanol (EtOH) at room temperature or in a freezer ( $-20^{\circ}\text{C}$ ) until their DNA is extracted. Alternatively, the living insect can be ground up in an extraction buffer that stops degradation of the DNA or RNA.

### 5.3.1 DNA Extraction Resulting in Loss of the Specimens

One of the most common methods formerly used to extract and purify nucleic acids required phenol ([Table 5.1](#)). Cells were disrupted mechanically by grinding to release the genomic DNA; usually the insect sample was ground in a solution containing protease K ([Jowett 1986](#), [Sambrook and Russell 2001](#)). Once the DNA is released, phenol can be used to extract DNA (or RNA) in large- or small-scale procedures. A plethora of different phenol-extraction methods have been published, but the primary function of phenol is to remove proteins from an aqueous solution containing nucleic acids. Proteins need to be removed because some of them are nucleases that could damage the DNA, whereas others simply could interfere with later manipulations. Ethylenediaminetetra-acetic acid (EDTA) is often added; it is a chelating agent that binds magnesium ( $\text{Mg}^{2+}$ ) ions that are required for nucleases to act on the DNA.

**Table 5.1: Rapid Phenol Extraction of Genomic DNA from *D. melanogaster*.**

1. Homogenize 50–200 flies (frozen in liquid nitrogen) in a 15-ml polypropylene tube with a Teflon pestle in 2 ml of lysis buffer.  
(Lysis buffer contains 100 mM Tris-HCl [pH 8.0], 50 mM NaCl, 50 mM EDTA, 1% SDS, 0.15 mM spermine, and 0.5 mM spermidine.)
2. Add 20 µl of proteinase K solution (10 mg/ml).
3. Leave at 37°C for 1–2 hours, but occasionally swirl and invert the tube to mix.
4. Extract once with an equal volume of phenol + chloroform + isoamyl alcohol.  
(The phenol, chloroform, and isoamyl alcohol should be in a 24:24:1 ratio. The isoamyl alcohol serves as an antifoaming agent.)
5. Spin in a bench centrifuge for 5 minutes at room temperature.
6. Decant the aqueous layer with a Pasteur pipette into a new tube.
7. Extract twice more with phenol + chloroform + isoamyl alcohol. Respin and decant the aqueous layer each time.
8. Extract the aqueous layer with chloroform and isoamyl alcohol (24:1). The interface between the organic and aqueous layer should now be clean.

(Modified from Jowett 1986.)

Highly purified phenol is mixed with the sample under conditions that favor the dissociation of proteins from the nucleic acids and the sample is then centrifuged (Table 5.1). Centrifugation yields two phases: 1) a lower organic phenol phase carrying the protein, and 2) the less-dense aqueous phase containing the nucleic acids. Some phenol-extraction protocols include chloroform; chloroform denatures proteins, removes lipids, and improves the efficiency of the extractions. To reduce foaming caused by chloroform, isoamyl alcohol is usually added. Phenol must be handled with great care and used only in a fume hood because it is toxic. Extraction of DNA from cells or organisms should be carried out as quickly as possible in ice with refrigerated buffers to minimize the activity of nucleases in the cells that can degrade the DNA.

Kits that are purchased usually involve the following steps: 1) breaking the cells open (cell lysis) to expose the DNA, by grinding or sonication; 2) removing membrane lipids by adding a detergent; 3) removing proteins by adding a protease; 4) removing RNA by adding an RNase (not always done); and 5) precipitating the DNA with an alcohol, usually ice-cold ethanol or isopropanol, resulting in a pellet upon centrifugation. In addition, a chelating agent may be added to inactivate Mg<sup>2+</sup> or Ca<sup>2+</sup> to prevent DNases from degrading the DNA. Some kits include chaotropic salts to disrupt cells.

### **5.3.2 DNA Extraction That Does Not Require Destroying the Specimens**

It is often the case that retaining voucher specimens is important in ecological and taxonomic studies. However, once the specimen(s) have been ground up

using the protocol outlined in Section 5.3.1, there are no true vouchers (Rohland et al. 2004, Rowley et al. 2007). A less well-known method to extract DNA and RNA involves using a **chaotropic salt**, a salt that disrupts the structure of or denatures proteins and nucleic acids. Guanidinium thiocyanate (GuSCN) or guanidinium hydrochloride (GuHCl) are both chaotropic salts that can be used to denature proteins and therefore can destroy nucleases that can degrade DNA or adhere to the DNA (Cox 1968, Bowtell 1987). The addition of a silica-based purification method can improve extraction efficiency (Hoss and Paabo 1993).

Jeyaprakash and Hoy (2010) evaluated four different protocols for extracting DNA from single predatory mites (Phytoseiidae) using a “soaking” method and evaluated the resultant nuclear and mitochondrial DNA by polymerase chain reaction (PCR). (For a description of the PCR, see Chapter 8.) The DNA obtained was suitable for amplification by the Random Amplification of Polymorphic DNA (RAPD)-PCR protocol, as well. One extraction protocol used the GuSCN buffer alone, the second protocol used a GuSCN-PTB buffer, the third protocol used a GuSCN buffer followed by isolation using a silica matrix, and the fourth protocol used a GuSCN-PTB buffer followed by isolation of DNA with the silica matrix. Sufficient DNA was extracted after soaking intact specimens using the third protocol (GuSCN buffer and a silica matrix that binds the DNA) to yield DNA for the PCR, although eggs of the mites had to be pricked to provide sufficient DNA. The quality of DNA produced was high, and the DNA produced stored well at –20°C. The adult mites soaked in the GuSCN buffer appeared normal after they were slide mounted and examined under high magnification.

Rowley et al. (2007) evaluated whether mitochondrial DNA could be extracted from tenuipalpid mites, a spider (Araneidae), a lady beetle (Coccinellidae), a dipteran (Ulidiidae), and a eurytomid wasp without damaging the specimens. Specimens were removed from 80% EtOH, air-dried, and soaked in GuSCN-based extraction buffer. DNA was precipitated with isopropanol, stored overnight at –20°C, and centrifuged for 20 minutes at 40°C. The DNA was then vacuum-dried, resuspended in TE buffer, and stored at –20°C. Specimens were generally well preserved and suitable as voucher specimens, and the DNA obtained was used to amplify mitochondrial DNA by the PCR.

### ***5.3.3 Assessing the Quality of Extracted DNA***

The accurate determination of DNA concentration and purity is essential for many molecular genetic protocols. Knowing the concentration of DNA is essential in ensuring reproducibility, enhancing amplification of DNA for the PCR, and cloning or sequencing protocols. Ultraviolet (UV) spectrometry often is used to determine DNA purity and quantity. The nucleic-acid sample, which may include

both DNA and RNA if an RNase has not been used to eliminate the RNA, is placed into a quartz cuvette that is then placed into the UV spectrophotometer. UV light is passed through the sample at a specified path length, and the absorbance of the sample at specific wavelengths is measured ( $A_{260}$ , or absorbance at 260 nm and  $A_{280}$ ) to measure contaminating protein in the sample. Based on the readings, the concentration of the sample is determined and the  $A_{260}/A_{280}$  ratios are calculated to indicate sample purity. The maximum absorbance of nucleic acids occurs at 260 nm, and DNA should yield an absorbance ratio of  $>1.8$ , with a maximum of 2.0. The  $A_{260}/A_{230}$  ratio is a key measure of purity.  $A_{230}$  is the wavelength that can measure contaminants such as chaotropic salts, which can inhibit PCR, and the  $A_{260}/A_{230}$  ratio should be  $>1.4$ , on a scale of 2. UV spectrophotometry is common and simple and can only detect concentrations of nucleic acids of  $\geq 1.5\text{ }\mu\text{g/ml}$ . The presence of proteins, RNA, and chaotropic salts can result in false estimates of DNA concentration, as can Tris buffer, EDTA, and GuSCN salts. The presence of free nucleotides and changes in pH also affect quantitation of DNA (Wilfinger et al. 1997).

An improved version of UV spectrophotometry is available in the NanoDrop series of equipment. It requires only 1–2  $\mu\text{l}$  of sample, which is retained on an optical fiber that assesses the UV absorbance of the sample between 220 and 750 nm. Software provided with the machine allows analysis of small samples. The concentration of the sample and its entire absorbance spectrum are provided in graphical form, thereby allowing contaminants to be detected and identified based on their absorbance wavelengths. The NanoDrop technology can evaluate a wide range of sample concentrations (Boesenberg-Smith et al. 2012).

Another method for estimating nucleic acids quantity is fluorometric. Dyes, which intercalate and bind nucleic acid grooves, bind nonspecifically or selectively to nucleic material. Ethidium bromide is one fluorometric dye that binds double-stranded DNA by intercalation, but it has a high level of natural fluorescence so it is limited in its sensitivity. Other fluorometric methods involve the use of Hoeschst 33258 and PicoGreen dyes. These methods require specific instruments and a variety of reagent kits.

## 5.4 Precipitating Nucleic Acids

During cloning, it is often necessary to concentrate DNA samples or change the solvent in which the nucleic acid is dissolved. DNA isolated by phenol contains trace amounts of phenol that could disrupt the activity of enzymes in subsequent manipulations if it were not purified further. Purification can be achieved

**Table 5.2: Sample Protocol for Precipitating DNA with Ethanol.**

For recovery of DNA from a typical reaction (1 µg of DNA in 20 µl):
1. To 20 µl of aqueous DNA sample in a microcentrifuge tube, add 2 µl of 3 M sodium acetate (pH 5.5) and 40 µl of EtOH.
2. Mix well by vortexing and immerse the tube in a –70°C bath composed of methanol plus dry ice for 15 minutes. The mixture should freeze or form a slurry.
3. Centrifuge the DNA precipitate in a bench top microcentrifuge at maximum speed for 10 minutes in a cold room. A whitish pellet of DNA should appear at the bottom of the tube. In general, pellets of 10 µg are visible, whereas pellets of 2 µg are invisible.
4. Remove the EtOH supernatant by using a micropipet, taking care not to disturb the pellet or the area of the tube where the pellet should be located.
5. Add 100 µl 70% EtOH (chilled to –20°C) to the sample and vortex. This step removes any solute trapped in the precipitate.
6. Reprecipitate the DNA by centrifugation for 2 minutes and remove the supernatant as described in step 4.
7. Dry the pelleted DNA for 1–2 minutes in a vacuum desiccator, taking care to release the vacuum gently so as not to dislodge the dried sample.
8. Resuspend the DNA in TE buffer (pH 8) (TE buffer contains 10 mM Tris-HCl at pH 8.0; 1 mM Na <sub>2</sub> EDTA) or in sterile water.

(Modified from Berger and Kimmel 1987.)

by ethanol (EtOH) precipitation, isopropanol precipitation, or several other methods. The most versatile is probably EtOH precipitation, because it can concentrate both DNA and RNA and purify DNA after phenol extractions.

DNA can be precipitated by combining the DNA sample with a salt and EtOH at –20°C or lower (**Table 5.2**). The precipitated salt of the nucleic acid is then pelleted by centrifugation, the EtOH supernatant is removed, and the nucleic-acid pellet is resuspended in a buffer. Which salt or which buffer is used is determined by the nature of the sample and by the planned use for the nucleic acid. Once the DNA is purified, it can be stored at 4°C in TE buffer (pH 8.0) or sterile water. EDTA in the TE buffer helps to prevent degradation of the DNA during storage because it chelates heavy metal ions that are commonly required for DNase activity. For very long-term storage ( $\geq 5$  years), the DNA can be frozen at –80°C but should not be subjected to freeze-thaw cycles or it will be damaged (sheared).

The now-purified DNA can be cut, either by shearing or, preferably, with a restriction endonuclease. Shearing produces random fragments, but restriction endonucleases can generate fragments of a desired size and with termini, or ends, appropriate to the annealing and ligation steps of the experiment outlined in **Figure 5.1**.

## 5.5 Shearing DNA

A variety of protocols are available to mechanically produce fragmented DNA. DNA can be sheared by sonication or by high-speed stirring. Sonication with ultrasound can produce DNA fragments of  $\approx 300$  nucleotides. High-speed stirring of cells in a blender at 1500 revolutions/minute for 30 minutes will produce DNA molecules with a mean size of 8kb. Breakage occurs essentially at random with respect to DNA sequence, and the broken ends consist of short, single-stranded regions. These single-stranded termini must be modified before the DNA can be joined to a vector, so mechanically sheared DNA rarely is used in experiments. More often, DNA is cut in a specific manner with a restriction endonuclease to make it easier to manipulate.

## 5.6 Cutting DNA with Restriction Endonucleases

Most cloning projects use restriction endonucleases to cut DNA. Restriction enzymes were discovered as an outcome of basic research aimed at understanding how bacteria control infections by **bacteriophages** (viruses that invade bacteria). Most bacteria contain a variety of specific endonucleases that guard against invasion of foreign DNA (Frank 1994). These endonucleases make cuts in the double-stranded DNA invading the cell unless the DNA has been modified in a specific manner by methylation. Thus, “foreign” DNA has an inappropriate methylation pattern, but the bacterium’s own DNA is protected because its methylation pattern is recognized. More than 3681 restriction enzymes (restriction endonucleases) have been identified, and 588 are commercially available (Roberts et al. 2005). Endonucleases with at least 221 different sequence specificities are commercially available (Kessler and Manta 1990, Pingoud et al. 1993, Roberts et al. 2005).

Restriction enzymes are identified by three-letter abbreviations for the parent organism (for example, *Hin* for *Haemophilus influenzae* or *Bam* for *Bacillus amyloliquefaciens*). An additional letter is added, if needed, to identify a specific strain or serotype (*Hind* or *BamH*). A Roman numeral is added to reflect the order of identification or characterization of the specific endonuclease (*HindIII* or *Bam HI*, Table 5.3).

Restriction endonucleases recognize specific sequences in DNA. They cleave the DNA in a precise manner, producing either blunt or staggered cuts (Table 5.3). Most endonucleases recognize sequences of four to six nucleotides, but some have seven- and eight-base recognition sites, and a very few recognize 12-base recognition sites. [Which endonucleases will produce the larger DNA fragments?] Restriction enzymes cleave DNA to form 5'-phosphate and

**Table 5.3: Target Sites for Selected Restriction Endonucleases.**

Organism from which endonuclease was isolated	Abbreviation	Recognition sequences indicates cleavage site (  indicates cleavage site)
<i>Bacillus amyloliquefaciens</i> H	<i>Bam</i> HI	G   GATCC
<i>Bacillus globigii</i>	<i>Bgl</i> II	A   GATCT
<i>Escherichia coli</i> RY13	<i>Eco</i> RI	G   AATTC
<i>Haemophilus aegyptius</i>	<i>Hae</i> III	GG   CC
<i>Klebsiella pneumoniae</i>	<i>Kpn</i> II	GGTAC   C
<i>Nocardiad otitidis-caviarum</i>	<i>Not</i> I	GC   GGCCGC
<i>Providencia stuartii</i>	<i>Pst</i> I	CTGCA   G
<i>Serratia marcescens</i>	<i>Smal</i>	CCC   GGG

Recognition sequences are written from 5' to 3' with only one strand given.

(Modified from Kessler and Manta 1990.)

**Table 5.4: Three Types of Termini are Created by Restriction Enzyme Cleavage of Double-Stranded DNA: a 5' Overhang, Blunt Ends, or a 3' Overhang.**

<i>Eco</i> RI	<i>Pvu</i> II	<i>Kpn</i> I
G   AATTC	CAG   CTG	GGTAC   C
CTTAA   G	GTC   GAC	C   CATGG
G <sup>3'</sup> AATTC	CAG <sup>3'</sup> CTG	GGTAC <sup>3'</sup> C
CTTAA <sup>5'</sup> G	GTC <sup>5'</sup> GAC	C <sup>5'</sup> 3'CATGG
5' overhang	Blunt end	3' overhang

The | indicates where cleavage occurs when DNA is restricted, or cut, by three representative enzymes.

3'-hydroxyl termini on each strand (Table 5.3). Endonucleases that produce staggered breaks generate either 5'-phosphate extensions or 3'-hydroxyl extensions. Other endonucleases produce "blunt" breaks (Table 5.4).

It is often desirable to generate DNA fragments of a specific length, with a specific sequence, and with a particular type of end. This precision is possible with well-characterized DNA that has been sequenced. Such precision is not possible with uncharacterized DNA, except to predict whether the ends will be blunt, or with 5' or 3' overhangs. It is difficult to predict precisely the length of the DNA fragments that will be generated after digesting unknown DNA sequences with a particular restriction endonuclease, although we can predict the sequences at the ends of each fragment (Table 5.4). [Why is that?]

Predictable fragment lengths would occur if all DNA sequences contained 50% guanine and cytosine (G + C) base pairs, and if all bases were distributed randomly in the DNA. Under these conditions, a four-base sequence that is recognized by the restriction endonuclease would occur approximately every

256 bases ( $4^4$ ), a six-base sequence would occur approximately every 4 kb ( $4^6$  or 4096 bases), and an eight-base sequence would occur approximately every 65 kb. However, many segments of DNA are not random in their G + C content. For example, highly repetitive DNA may have several nucleotides repeated millions of times, which obviously would bias restriction site frequencies significantly. The percentage of G + C in DNA from different sources may vary from 22 to 73%.

Many different restriction endonucleases are commercially available either in native or cloned form. Most manufacturers provide standardized buffers (high-, medium-, or low-salt) for optimizing the reaction conditions and protocols for carrying out the digestions. Among those available, some recognize identical sequences, although they may vary with respect to their sensitivity to methylation and cleavage rates (Roberts et al. 2005). Restriction endonucleases can be degraded if not properly stored at  $-20^{\circ}\text{C}$  and should be subdivided into small amounts so they do not undergo multiple freeze-thaw cycles before being used.

Choosing which endonuclease to use is determined by the goals of the project, and a computer program is available to assist in designing experiments using restriction enzymes (Martin et al. 2006). Enzymes that produce small segments of a few hundred bases are useful for restriction mapping or for sequencing. Enzymes that produce fragments of 1–10 kb are useful for mapping large DNA regions and for cloning whole genes with their introns and control sequences. Generating even larger fragments (5–50 kb) is necessary for cloning into some vectors or for genome walking (discussed in Chapter 6).

Digestion reactions with restriction endonucleases contain: the DNA substrate, the restriction endonuclease(s), Tris buffer,  $\text{Mg}^{2+}$ , NaCl, 2-mercaptoethanol, and bovine serum albumin. All endonucleases require  $\text{Mg}^{2+}$  as a cofactor, and most are active at pH values ranging from 7.2 to 7.6. The major difference among the endonucleases is their dependence on ionic strength and their temperature optima. Most digestions are done at  $37^{\circ}\text{C}$ , but a few restriction endonucleases perform better at lower temperatures. The manufacturer usually measures endonuclease activity with bacteriophage lambda ( $\lambda$ ) DNA as a substrate, but activity of the endonuclease varies with different DNA substrates and also can be modified by the neighboring sequences. Activity rates can vary by a factor of 10- to 50-fold in your laboratory experiments.

The number and variety of endonucleases available for genetic manipulations continue to increase. Endonucleases that recognize longer recognition sequences are useful if large DNA fragments are to be separated by pulsed-field gel electrophoresis. New microbial sources of enzymes are being sought, especially those that tolerate high temperatures. Catalogs obtained from suppliers

contain useful information on restriction endonuclease activity and their appropriate reaction conditions.

## 5.7 Joining DNA Molecules

Different DNA fragments cleaved by restriction endonucleases can be joined by ligases. Ligases make it possible to insert exogenous DNA into plasmid vectors. Two **DNA ligases** are commonly used. One DNA ligase is derived from *E. coli* and the other DNA ligase is from the bacteriophage T4. Their requirements for cofactors differ. T4 ligase requires ATP, whereas *E. coli* ligase requires NAD<sup>+</sup>. Both catalyze the joining of a 5'-phosphate and a 3'-OH group to form a phosphodiester bond. T4 DNA ligase will catalyze the joining of blunt-ended DNA molecules and cohesive-ended molecules, although more enzyme is required for blunt-ended ligation of DNA molecules.

If the restriction endonuclease used generated DNA fragment ends with uneven ends or overhangs (Table 5.4), then the sequences of the DNA within the single-stranded regions of the two molecules have to be complementary for ligation to occur. Ligation of four-base extensions is easier than ligation of two-base extensions. Extensions that consist of G + C bases ligate more readily than those with A + T base pairs. [Can you explain why?]

Blunt ends are more difficult to ligate, requiring 20–100 times more T4 DNA ligase and higher DNA concentrations. The surrounding DNA sequences do not affect ligation efficiency, but ligation is negatively influenced by the presence of contaminating endonucleases or by phosphatase.

The optimum temperature for ligating DNA is 37°C, but the hydrogen-bonded joint between sticky ends is unstable at this temperature. As a result, the ligation reaction is carried out at a temperature that is a compromise between the optimum for the rate of the enzyme action and the association of the termini, usually between 4 and 15°C. Ligation reactions often are allowed to take place overnight at these low temperatures. The ligation reaction can be promoted by adjusting the ratio of insert DNA and vector DNA. When a linear DNA fragment is produced by a restriction endonuclease from a circular vector, the linear fragment will often recircularize and hydrogen bond to itself, or to other linear vector sequences. To prevent this, the linearized plasmid vector DNA can be treated with alkaline phosphatase to remove 5'-terminal phosphate groups. Alkaline phosphatase prevents recircularization of the plasmid or formation of plasmid dimers, although the phosphatase must be eliminated if the vector and linear DNA are to be ligated. Circularization of the vector will then occur if the foreign DNA (untreated with phosphatase) joins the ends of the vector.

**Table 5.5: Blunt-End Ligation when the Vector-to-Insert Molar Ratio = 3.**

In a siliconized Eppendorf tube in ice mix:	
Dephosphorylated vector DNA ( $\approx$ 4 kb)	160 ng
DNA fragment ( $\approx$ 1 kb)	13 ng
10× ligase buffer I (250 mM Tris-HCl [pH 7.5], 50 mM MgCl <sub>2</sub> , 25% w/v polyethylene glycol 8000, 5 mM DTT, 4 mM ATP)	4 $\mu$ l
T4 DNA ligase	1 Weiss unit
Water to a final volume of	20 $\mu$ l
Incubate at 23°C for 4 hours and stop the reaction by adding 1 $\mu$ l of 0.5 M EDTA.	
Dilute five-fold before adding the mixture to competent <i>E. coli</i> cells for transformation.	

Only T4 DNA ligase is able to join blunt-ended DNA molecules. A typical blunt-end ligation reaction is described in [Table 5.5](#).

## 5.8 Growth, Maintenance, and Storage of *E. coli*

*Escherichia coli* has been studied extensively and the genome completely sequenced ([Snyder and Champness 1997](#), [Neidhardt 1999](#)). DNA manipulations require manipulating bacteria, primarily derivatives of *E. coli* K12 strains. Different *E. coli* strains are used for different purposes ([Miller 1992](#)).

Standard microbial techniques are used: pure cultures of *E. coli* are obtained by propagating cultures from single, isolated colonies on agar plates. Dilution streaking with an inoculating loop readily produces isolated colonies and an isolated colony can be restreaked to obtain a pure master plate that can be stored at 4°C for a month. It is important that *E. coli* cultures are kept pure, the phenotypes are verified before use, and the cultures are stored properly. For long-term storage, cultures can be stored in stab vials, as frozen glycerol cultures, or as lyophilized cultures.

Overnight cultures of most strains of *E. coli* produce  $\approx$ 4  $\times$  10<sup>9</sup> bacteria/ml, depending upon the medium, degree of aeration, strain, and temperature. To determine the cell concentration, dilutions of the culture should be plated. Detailed methods for manipulating *E. coli* are readily available in laboratory manuals ([Miller 1992](#), [Sambrook and Russell 2001](#)).

## 5.9 Plasmids for Cloning in *E. coli*

Plasmids are widely used as cloning vectors. Many plasmids have undergone extensive genetic engineering to enhance their value as vectors ([Figure 5.2](#)). The complete sequence of the vector usually is known, including the location of unique restriction sites (sites where a specific endonuclease can cut the plasmid).

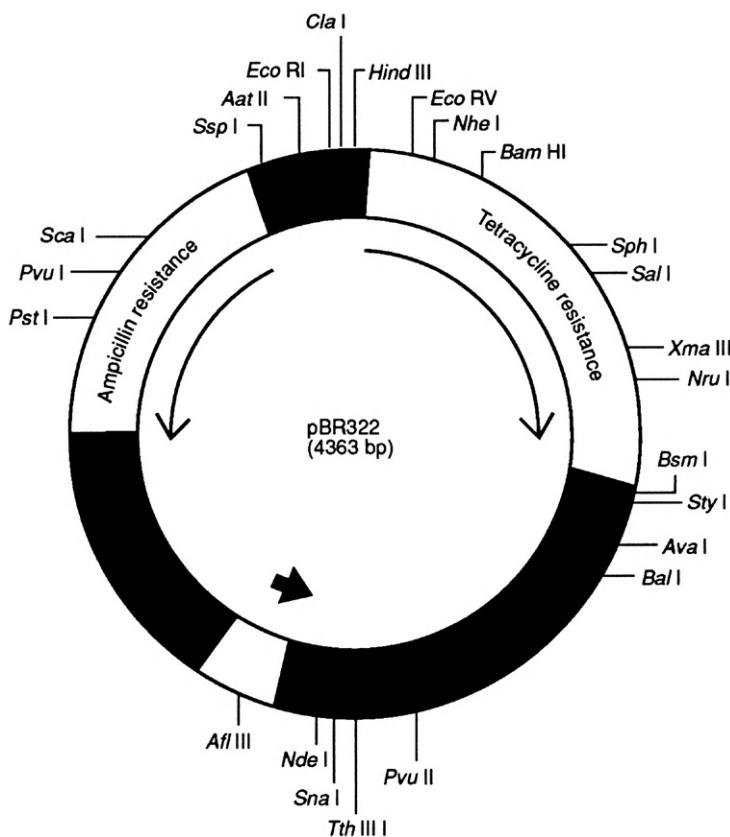
"Wild-type" **plasmids** are small DNA molecules that are stably inherited as extra-chromosomal units in many kinds of bacteria. Plasmids are widely found in bacteria, but usually are not essential to their host. Many plasmids carry genes for antibiotic resistance, antibiotic production, heavy-metal resistance, an ability to degrade aromatic compounds, sugar fermentation, enterotoxin production, or hydrogen-sulfide production. Most are covalently closed DNA circles, but some are linear.

Plasmids can be classified into two types depending upon whether they carry a set of genes that promote **bacterial conjugation**. Plasmids also can be categorized as to whether they are maintained in multiple copies in host cells or in limited numbers per cell. Generally, plasmids that promote bacterial conjugation are relatively large and are present in one to three copies per bacterial cell. Plasmids that do not promote bacterial conjugation are smaller and multiple copies are found in a cell. Plasmids are "promiscuous" if they can promote their own transfer to a wide range of bacteria and can be maintained stably in their new hosts. Promiscuous plasmids can transfer cloned DNA molecules into different bacteria. Wild-type plasmids could be used for cloning in *E. coli*, but they suffer from several disadvantages, and genetically engineered plasmids have been developed that have many desirable attributes.

The first genetic improvement of plasmids involved removing excess DNA so that the plasmid is easier to manipulate *in vitro*, resistant to damage by shearing, and readily isolated from bacterial cells. Smaller size is an advantage because bacterial cells usually can sustain several smaller plasmids, which will increase the yield of the recombinant DNA molecules.

A second improvement was the addition of one or more **selectable marker** genes to the plasmid ([Figure 5.2](#)). A selectable marker allows the experimenter to identify those bacterial cells that have taken up the plasmid during the transformation process. Many selectable markers are antibiotic resistance genes (for example, ampicillin and tetracycline) that allow the transformed bacteria to be grown on selective media.

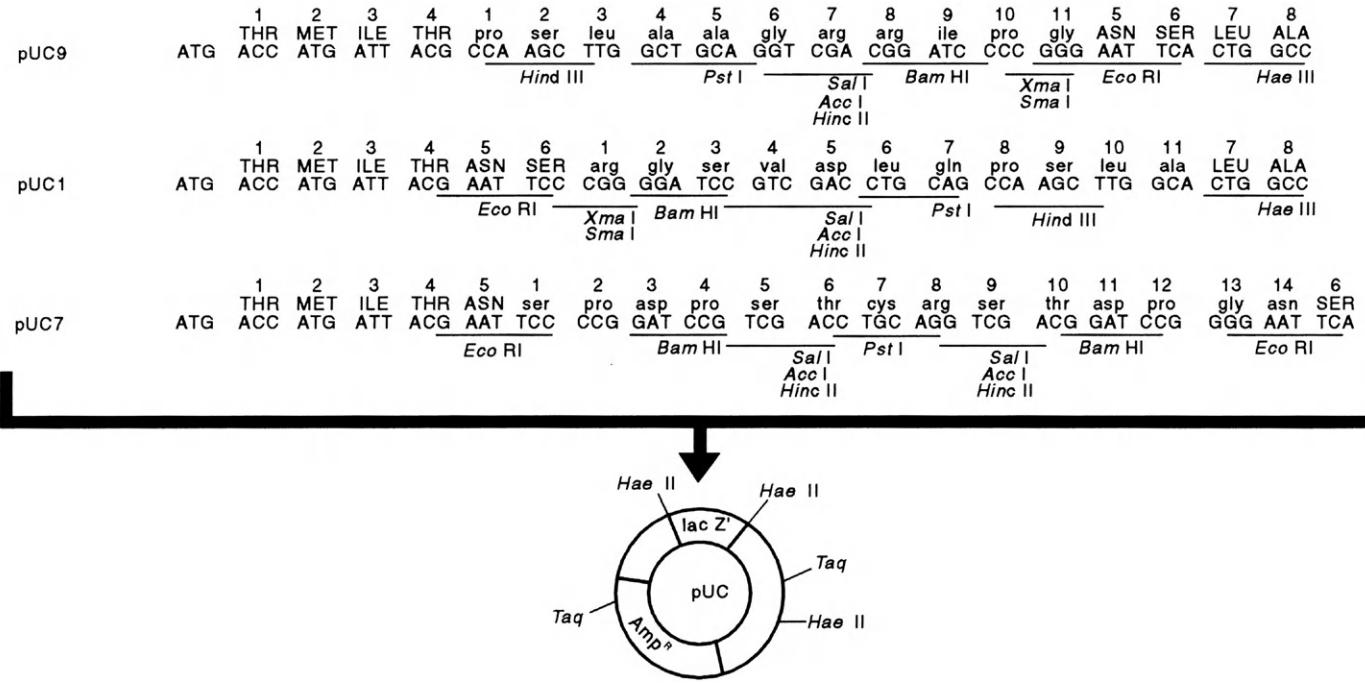
A third improvement involved adding DNA containing a **multiple cloning site** or **polylinker** that can be cut by several restriction endonucleases ([Figure 5.3](#)). The presence of these unique restriction or cloning sites is helpful because cloning requires that both the vector and the exogenous DNA be cut with the same endonuclease (or with endonucleases that produce the same kinds of ends) so that the ends can be ligated together. If the plasmid had more than one site that was cut by a specific endonuclease, the plasmid vector would be cut into several fragments, resulting in defective vectors. A polylinker is a short segment of DNA with sites where several different restriction endonucleases can cut.



**Figure 5.2** The structure of the cloning plasmid pBR322 showing the unique sites where restriction endonucleases can cleave the DNA. The thin arrows inside the circle indicate the direction of transcription of the ampicillin and tetracycline resistance genes, which serve as selectable markers. The thick arrow shows the direction of DNA replication.

This gives the genetic engineer options as to which restriction enzyme to use. If the polylinker is placed within a selectable marker gene such as *lacZ*, gene function is disrupted when exogenous DNA is cloned into the polylinker site at any of the restriction sites and the recombinant colonies can be identified by their color.

Plasmids have been engineered by sophisticated techniques to perform a variety of defined tasks. **Expression vectors** facilitate expression of proteins; for example, baculovirus vectors are used to produce large amounts of foreign proteins in insect cells (see Chapter 6). Some vectors help identify regulatory signals that turn genes on or off, some are used for direct selection of recombinants, some have increased stability so that they are not eliminated from their host cells, and others are genetically altered so that high copy numbers/host cell can



**Figure 5.3** Genetic maps of the multiple cloning sites of some pUC plasmids. The multiple cloning site is inserted in the *lacZ* gene but does not disrupt *lacZ* gene function. However, if exogenous DNA is cloned into this site, then *lacZ* gene function is disrupted, which indicates that insertion was successful. This allows the genetic engineer to detect colonies of *E. coli* containing recombinant plasmids because these colonies are white rather than blue if the *E. coli* are grown on a substrate with 5-bromo-4-chloro-3-indolyl-β-D-galactoside (X-Gal). The pUC plasmids also contain an ampicillin resistance gene.

be maintained. **Sequencing vectors** can be used for DNA sequencing because they produce a single-stranded copy of DNA rather than a double-stranded molecule (see Chapters 6 and 7). Some vectors are modified so that proteins are secreted through the host cell wall to facilitate purification of proteins, and others are modified to produce fusion proteins to facilitate protein purification. Many versatile vectors are available from commercial sources.

A key to effective genetic engineering is the ability to identify those cells that have been genetically transformed. To aid in this, most plasmid vectors contain at least two selectable markers, which often are antibiotic resistance genes. If two markers are present, then one is often the site into which the exogenous DNA is cloned. Insertion of exogenous DNA should *inactivate* that resistance gene, so that the *E. coli* cells containing this insertion can be identified because they are newly susceptible to the antibiotic. Another commonly used selectable marker is *lacZ*, which allows selection of blue/white bacterial colonies.

Plasmid vectors, such as pBR322 and its derivatives, are widely used because they can be produced in multiple copies within a cell; they are easily purified, and can produce large amounts of the cloned gene (Figure 5.2). pBR322 carries both ampicillin and tetracycline resistance genes and an **origin of replication**, which is a sequence at which replication of the DNA molecule is initiated. pBR322 has been completely sequenced and its restriction sites characterized. This means that the exact length of each fragment from a restriction digest can be predicted and that these fragments can serve as DNA markers for sizing other DNA fragments. pBR322 fragments produced after digestion with restriction enzymes range in size from several base pairs to the entire 4.3-kb plasmid.

Some unique restriction sites occur within both the ampicillin and tetracycline resistance genes of pBR322 (Figure 5.2), and these are very useful in cloning. If exogenous DNA is inserted into a site in the ampicillin resistance gene where a restriction enzyme cuts uniquely, the ampicillin resistance gene will be inactivated. The recombinant plasmids contained within their *E. coli* host can then be identified by spreading the transformed *E. coli* first onto culture plates with media containing tetracycline and then replica plating them onto plates with ampicillin.

**Replica plating** is a procedure in which a particular pattern of *E. coli* colonies on an agar surface is reproduced on another agar surface. The pattern is obtained by pressing a piece of sterile velvet upon the original agar surface, thereby transferring cells from each colony to the cloth and pressing this pattern onto another agar surface. It is important to carefully mark the orientation of the patterns on the original and the replica plates. The recombinant *E. coli* colonies that are unable to survive on the ampicillin can be recovered by finding the colony growing on the original

**Table 5.6: Producing Competent *E. coli* and Transforming them with Plasmid DNA Using CaCl<sub>2</sub>.**

Production of competent cells
<ol style="list-style-type: none"> <li>1. Grow a fresh overnight culture of <i>E. coli</i> in Luria-Bertani (LB) broth at 37°C.</li> <li>2. Dilute the cells 40-fold into 1 liter of fresh medium. Incubate at 37°C with good aeration until their density produces an absorbency rating at 550 nm of 0.4–0.5.</li> <li>3. Immediately chill the culture by swirling in a ice-water bath.</li> <li>4. When the cells are chilled, centrifuge the culture at 4°C at 5000 rpm for 10 minutes.</li> <li>5. Decant the supernatant and place the pellet in ice.</li> <li>6. Resuspend the pellet in 500 ml of ice-cold 100 mM CaCl<sub>2</sub>. It is easier to resuspend the pellets if they are vortexed before the CaCl<sub>2</sub> is added. The cells can be suspended by sucking them up and down in a 25-ml pipet.</li> <li>7. Once the cells are resuspended, incubate in ice for 30 minutes with occasional swirling.</li> <li>8. Pellet the cells once again at 5000 rpm for 10 minutes at 4°C.</li> <li>9. Resuspend in 40 ml of ice-cold 100 mM CaCl<sub>2</sub> and 15% glycerol.</li> <li>10. Distribute aliquots of 0.2 ml of cells into sterile Eppendorf tubes in ice.</li> <li>11. Keep in ice at 0–4°C for 12–24 hours. This is essential for maximal competency, although the cells are competent at this stage.</li> <li>12. Freeze the tubes in ethanol-dry ice or liquid nitrogen and place immediately at –70°C. The cells remain competent for months if stored at –70°C.</li> </ol>
Transformation of competent cells
<ol style="list-style-type: none"> <li>1. Thaw a tube of frozen competent cells at 4°C.</li> <li>2. Add DNA in buffer.</li> <li>3. Incubate in ice for 30 minutes.</li> <li>4. Heat shock for 2–5 minutes in a 42°C water bath.</li> <li>5. Add 0.4 ml of LB broth at room temperature to each tube and incubate for 1 hour at 37°C.</li> <li>6. Spread on agar plate with appropriate antibiotics. Incubate plates overnight at 37°C.</li> </ol>

(Modified from Berger and Kimmel 1987.)

tetracycline plate based on their location. Many derivatives of pBR322 have been constructed to fulfill particular cloning goals (Balbas et al. 1986).

## 5.10 Transforming *E. coli* with Plasmids

A plasmid carrying exogenous DNA must be inserted into *E. coli* to amplify, or clone, the DNA. The process of inserting a plasmid into *E. coli* is called bacterial **transformation**. For many years efforts to transform *E. coli* were unsuccessful, and it was only in 1970 that transformation methods were developed. The ability to transform *E. coli* required understanding its genetics and having the ability to manipulate the physiological status of the *E. coli* cells to optimize the transformation reaction.

A simple transformation procedure involves suspending *E. coli* cells that are in the log phase of their growth cycle in an ice-cold solution containing membrane-disrupting agents such as polyethylene glycol (PEG), dimethyl sulfoxide (DMSO), or divalent cations such as calcium chloride (Table 5.6). Plasmid DNA is added

to a small aliquot of these **competent cells** (competent for transformation) and the incubation on ice is continued for another 30 minutes. A heat shock is then administered by putting the cells into 42°C for 2 minutes. The cells are then transferred to nutrient broth and incubated for 30–60 minutes to allow the plasmid to express its phenotypic properties (plasmids often carry antibiotic resistance genes as selectable markers). The cells then are plated onto agar plates containing a selective medium. Only those bacteria that have taken up the plasmid with the selectable marker should survive and reproduce on the selective medium.

How transformation occurs is not entirely understood. Various agents affect the bacterial cell wall and, in the case of  $\text{CaCl}_2$ , also may be responsible for binding DNA (the plasmid) to the cell wall. The actual uptake of DNA is stimulated by the brief heat shock. Large DNA molecules are taken up less efficiently than smaller DNA molecules, and the efficiency of transformation varies with the strain of *E. coli* used. Efficiency is typically expressed as the number of transformant cells per microgram of plasmid DNA. Various protocols produce efficiencies of  $10^7$  or  $10^8$  transformants/ $\mu\text{g}$  of plasmid DNA.

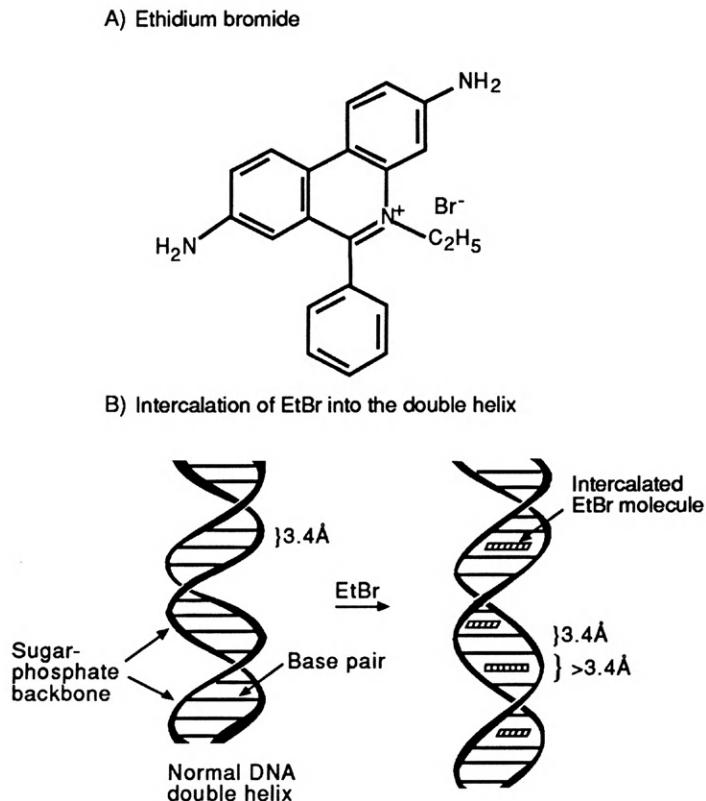
**Electroporation** also can be used to insert DNA into bacterial cells. Electroporation involves disrupting the cell membrane briefly with an electric current so that DNA can be incorporated. Commercial units and protocols can be purchased for electroporation of *E. coli*.

## 5.11 Purifying Plasmid DNA from *E. coli*

Removing the plasmids from *E. coli* is necessary if experiments are to be conducted on the now-cloned DNA. The trick is to lyse (break open) the *E. coli* cells just sufficiently so the plasmids can escape without too much contamination by the bacterial chromosome. If the bacterial cell is lysed gently, most of the bacterial chromosomal material released will be of higher molecular weight than the plasmids and can be removed, along with the cell debris, by complexing with detergents and high-speed centrifugation. The plasmid DNA is left in the clear liquid remaining, and it can be extracted by one of two traditional methods.

In the first method, cesium chloride centrifugation with **ethidium bromide** (EtBr) yields bands in the centrifuge tube that contain chromosomal and plasmid DNA at different levels due to the different densities of linear and supercoiled DNA in the presence of EtBr.

Ethidium bromide stains DNA by intercalating between the double-stranded DNA base pairs and in so doing causes the DNA to unwind ([Figure 5.4](#)). A plasmid DNA molecule that has not been nicked is a circular double-stranded



**Figure 5.4** Ethidium bromide (EtBr) can intercalate into DNA and cause the DNA to unwind. DNA containing EtBr will fluoresce if exposed to UV irradiation. EtBr is used to visualize DNA fragments on gels after electrophoresis. EtBr also can serve as a mutagen. Handle EtBr with great care and protect your eyes from UV irradiation.

supercoil that has no free ends and can only unwind to a limited extent, thus limiting the amount of EtBr that it binds. A linear DNA molecule, such as fragmented bacterial chromosomal DNA, can bind more EtBr and become stiffer, extending the molecule and reducing its buoyant density.

A second method for extracting and purifying plasmid DNA exploits the observation that within a pH range of 12.0 to 12.5, linear DNA will completely denature, but closed circular (plasmid) DNA will not. Plasmid-containing bacteria are treated with lysozyme to weaken the cell wall and then lysed with sodium hydroxide and sodium dodecyl sulfate (SDS). The chromosomal DNA is denatured, but upon neutralization with acidic potassium acetate, the chromosomal DNA renatures and aggregates to form an insoluble network. The high concentration of potassium acetate also causes the protein-SDS complexes and high

**Table 5.7: Small-Scale Plasmid Preparations (Minipreps).**

1. Prepare 5-ml transformed *E. coli* cultures in LB broth containing the appropriate antibiotic. The cultures can be grown in disposable 14-ml plastic centrifuge tubes by picking colonies with a sterile toothpick and dropping the toothpick into the tube. Cap the tube and incubate at 37°C with shaking (250 rpm) for 16 hour.
2. Pellet the cells by centrifugation at 5000 rpm for 5 minutes. Discard supernatant and toothpick.
3. Add 100 µl of 50 mM glucose, 25 mM Tris-HCl (pH 8.0), 10 mM EDTA, and 2 mg/ml lysozyme (freshly prepared). Resuspend pellet and incubate for 10 minutes.
4. Add 200 µl 0.2 N NaOH, and 1% SDS. Mix gently. Incubate on ice for 10 minutes. The SDS-NaOH solution must be made just before use and kept at room temperature. Mix 3.5 ml of water, 1 ml of 1 N NaOH, and 0.5 ml of 10% SDS.
5. Add 150 µl of 3 M potassium acetate (pH 4.8). Mix gently. Incubate for 10 minutes in a freezer. A white precipitate will form.
6. Centrifuge for 15 min at 15,000 rpm at 4°C.
7. Pour supernatant into Eppendorf tubes and fill with cold ethanol. Incubate in ice for 10 minutes.
8. Pellet the DNA for 1 minutes in a microcentrifuge and aspirate off supernatant. Add 0.5 ml of cold 70% ethanol and aspirate off.
9. Dry under vacuum. Resuspend in 50 µl of distilled water containing 10 µg/ml pancreatic ribonuclease (RNase) to remove RNA.
10. The DNA is suitable for restriction analysis or fragment preparation. Use 5–10 µl per reaction. If the DNA does not cut well it can be re extracted with phenol and precipitated with ethanol.

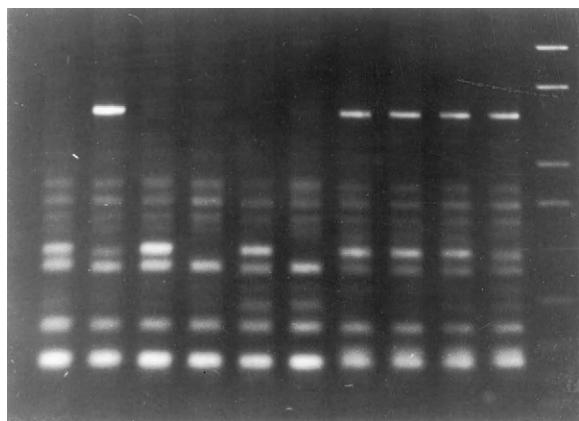
molecular weight RNA to precipitate. If the pH of the alkaline denaturation step has been controlled carefully, the plasmid DNA molecules will remain circularized and in solution, but the contaminating molecules will precipitate. The precipitate can be removed by centrifugation and the plasmid can be purified and concentrated by ethanol precipitation (**Table 5.7**).

However, many kits are now available that can be used to extract plasmids from transformed *E. coli*, and automated work stations even are available for laboratories with a high-volume workload.

## 5.12 Electrophoresis in Agarose or Acrylamide Gels

DNA and RNA molecules can be separated by size and visualized by agarose or acrylamide gel electrophoresis. **Gel electrophoresis** provides a powerful method for resolving mixtures of single- or double-stranded (ss or ds) nucleic-acid molecules. The nucleic acids can be visualized *in situ* in the gel by staining with ethidium bromide (EtBr) and examining it under UV light (**Figure 5.5**).

At a pH near neutrality, linear DNA is negatively charged and migrates from cathode to anode in a gel, with its mobility dependent on fragment size, voltage applied, composition of the electrophoresis buffer, base composition, gel concentration, and temperature.



**Figure 5.5** Photograph of DNA stained with ethidium bromide and illuminated with UV light. The bands are a pink-purple color.

Agarose gel is used for longer DNA molecules and polyacrylamide for shorter. Nondenaturing polyacrylamide gels can be used to separate double-stranded DNA fragments between 6 bp (20% acrylamide) and 1000 bp (3% acrylamide). Nondenaturing agarose gels can separate DNA fragments between 70 bp (3% agarose) and 10,000 bp (0.1% agarose). Single-stranded DNA can be separated by agarose or polyacrylamide gel electrophoresis by including a denaturing reagent in the gel.

DNA from 60 kb to 0.1 kb can be detected with agarose gels containing different percentages of agarose. Agarose gels are usually electrophoresed at room temperature, except for low-percentage agarose gels (<0.5%) or low-melting temperature gels that are easier to handle at cooler temperatures.

Agarose powder comes in grades that vary in purity and melting temperature. Type II agarose is generally used, although it contains contaminants that coelute with DNA and inhibit most commonly used enzymes. This means that DNA must be purified after elution from this gel if it is to be ligated or cut with restriction enzymes. An alternative involves using a high-quality, low-melting temperature agarose that melts at 65°C and sets at 30°C. Low-melt agarose allows DNA to remain double-stranded and also allows many enzymes to be used in the liquid agar.

Polyacrylamide gels result from polymerization of acrylamide monomers into linear chains and the linking of these chains with *N,N'*-methylenebis-acrylamide (often called bis). The concentration of acrylamide and the ratio of acrylamide to bis determine the pore size of the resultant three-dimensional network and thus its sieving effect on nucleic acids. Polyacrylamide gels can be used to purify

synthetic oligonucleotides, isolate or analyze DNA<1 kb, and resolve small RNA molecules by two-dimensional or pulsed-field gel electrophoresis. CAUTION: Polyacrylamide gel ingredients are highly toxic and should not be inhaled or touched without gloves.

### 5.13 Detecting, Viewing, and Photographing Nucleic Acids in Gels

Ethidium bromide is a useful dye to detect both single- and double-stranded nucleic acids in both agarose and polyacrylamide gels (Figure 5.5). Agarose gels are less sensitive in detecting small amounts of DNA than are polyacrylamide gels. The sensitivity for DNA is 5- to 10-fold less.

Ethidium bromide (EtBr) can be incorporated into the gel and running buffer during electrophoresis. Alternatively, gels can be stained after electrophoresis by placing them in buffer containing EtBr for 30 minutes. CAUTION: EtBr is mutagenic so the experimenter must be extremely cautious when handling it. Gloves must be worn and care must be taken to avoid contaminating laboratory surfaces.

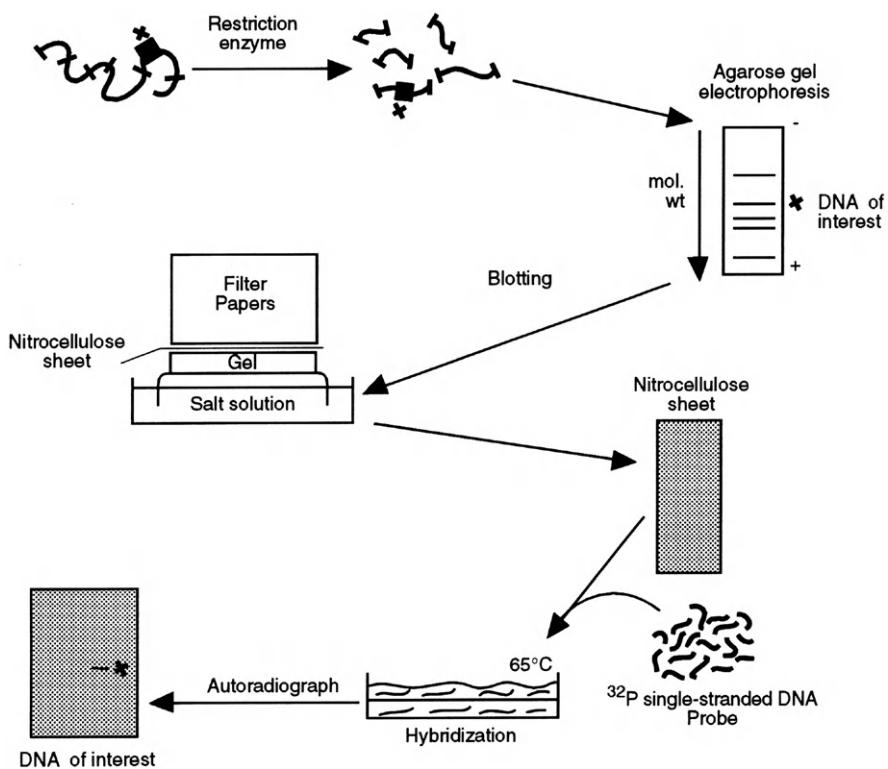
As little as 0.05 µg of DNA can be visualized in one band when the gel is exposed to ultraviolet (UV) light (Figure 5.5). The EtBr–nucleic acid complex absorbs UV irradiation at ≈260 nanometers (nm) or 300 nm. The fluorescence of EtBr stacking in duplex DNA is 10 times greater than that of free EtBr, and the emission is at 590 nm (red orange). UV light sources (transilluminators) are used to view DNA stained with EtBr at 254 or 366 nm. Although the short-wave model can detect smaller amounts of sample DNA, it damages DNA by nicking or dimerization, and the DNA is unsuitable for most cloning work. CAUTION: Experimenters should wear safety glasses around UV light sources to protect their eyes. Agarose and polyacrylamide gels can be photographed to document the results of the experiment.

### 5.14 Identifying Specific DNA by Southern Blot Analysis

It is often necessary to identify specific DNA sequences. One method to do so was invented by [Southern \(1975\)](#), and the “**Southern blot**” has been a fundamental and versatile tool for genetic engineers ever since. The original paper published by Southern was the most-cited paper ever published in the *Journal of Molecular Biology* ([Southern 2000](#)). Variations on Southern blots have been developed to identify RNA or proteins in gels. These modifications are called **Northern** (RNA) and **Western** (protein) blotting, respectively. There is no “Eastern” blot.

In Southern blotting, DNA fragments that have been separated by electrophoresis in an agarose gel are denatured into a single-stranded form by alkali treatment (Figure 5.6). The gel is then laid on top of buffer-saturated filter paper. The top of the gel is covered with a nitrocellulose filter membrane. This

membrane is then overlain with dry filter paper. Additional layers of dry filter paper or absorbent papers are stacked on top. Buffer passes out of the gel, drawn by capillary action into the dry filter papers. As the buffer moves from the gel, it elutes some of the denatured DNA from the gel. When the single-stranded DNA comes in contact with the nitrocellulose lying on top of the gel, it binds to the membrane. The blotting process is carried out over several hours, and results in the transfer of part of the DNA from the gel onto the nitrocellulose membrane. It results in an (at this point) invisible pattern of bands on the membrane surface that resembles the original bands in the gel, with a minimal loss of resolution. The stack of filter papers is then removed and the



**Figure 5.6** Outline of a Southern blot procedure. DNA is cut with restriction enzymes, electrophoresed, and blotted onto nitrocellulose by capillary action. The nitrocellulose sheet containing the DNA is baked to bind the DNA to the nitrocellulose. Specific DNA is identified by the binding of a labeled probe (here it is  $^{32}\text{P}$ -labeled single-stranded DNA) in a hybridization procedure. Excess probe is washed off, and the nitrocellulose sheet is then exposed to X-ray film to visualize the location of the DNA that contains sequences that are homologous to the radiolabeled probe. The conditions (stringency) under which hybridization occurs can be varied to increase or decrease the specificity of the reaction between DNA and probe. Southern blots allow scientists to locate specific DNA sequences.

nitrocellulose membrane is baked at 80°C in a vacuum to bind the DNA permanently onto the surface of the nitrocellulose filter. Electroblotting or vacuum blotting are alternative methods for transferring the DNA to a membrane and require specialized equipment.

To determine whether the DNA of interest is present on the blot requires probing the DNA on the nitrocellulose membrane ([Figure 5.6](#)). The **probes** can be labeled by  $^{32}\text{P}$  or by nonradioactive methods. Probes can consist of RNA, ss DNA, or a synthetic oligonucleotide that is *complementary* in sequence to the DNA of interest. The labeled probe must bind specifically to the DNA of interest, but not bind to nontarget DNA or the nitrocellulose.

Pretreating the bound DNA on the nitrocellulose by placing it in a solution containing 0.2% each of Ficoll (an artificial polymer of sucrose), polyvinylpyrrolidone, and bovine serum albumin (also known as Denhardt's solution) will prevent nonspecific binding. The mixture often includes an irrelevant nucleic acid such as salmon sperm DNA, which may act by occupying all the available non-specific binding sites on the membrane.

The temperature at which Southern blotting is conducted is adjusted to maximize the rate of hybridization of the probe with the immobilized DNA on the nitrocellulose and to minimize the amount of nonspecific binding. This aspect of planning the Southern blot is called **stringency**, and a highly stringent Southern blot will be more specific. After the hybridization step, in which the labeled probe binds to the immobilized DNA on the membrane, the membrane is washed to remove any unbound probe. The temperature at which the washing takes place also determines the stringency of the Southern blot.

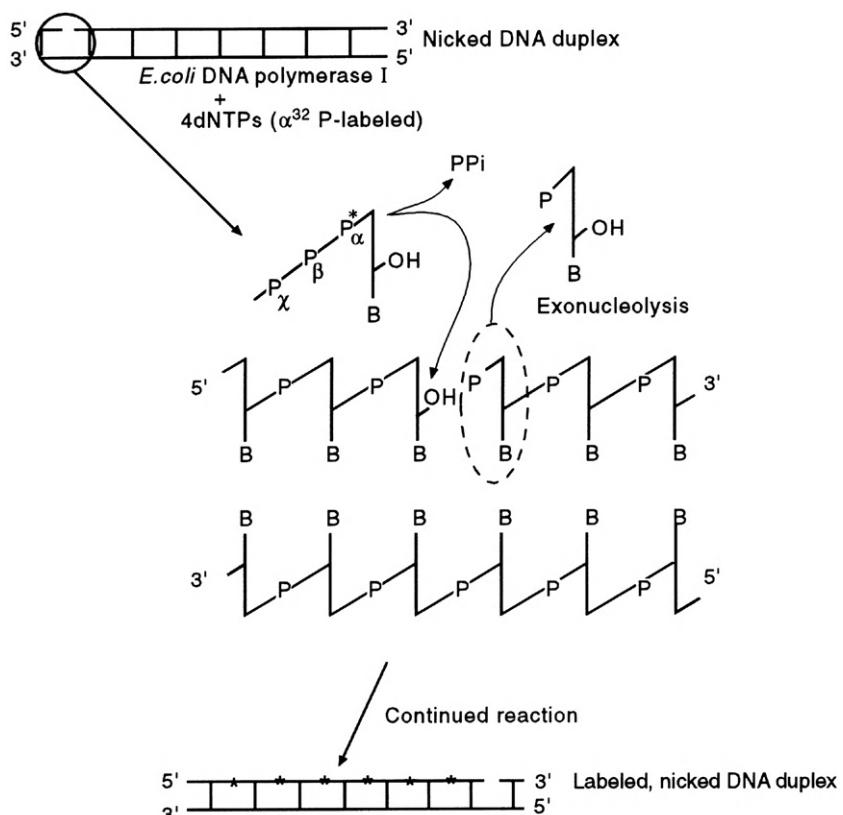
The regions on the membrane where hybridization of the labeled probe and target DNA took place are detected by placing the membrane in contact with X-ray film if the probe was radiolabeled. The length of time the X-ray film is exposed to the radioactivity is determined by the amount of DNA in the blot and the degree of homology between the DNA and the probe. If there is only a small amount of DNA present, as would be expected for a single-copy gene in a blot of genomic DNA, the film may be exposed for 2–3 weeks to the blot.

Southern blots can detect single-copy gene sequences in genomic DNA digests. Modifications of the Southern blot method use nylon membranes as substrates because they are more robust and can be reused. Thus, one probe can be removed by high-temperature washing, and the DNA can be reprobed with a different probe. Another advantage to nylon membranes is that the DNA can be permanently fixed to the membrane by a brief exposure to UV light, which cross-links the DNA and fixes it to the membrane.

## 5.15 Labeling DNA or RNA Probes

In molecular biology, many DNA manipulation techniques depend on hybridizing a nucleic-acid probe to a target DNA or RNA sequence. Probes are required in Southern and Northern blots, dot blots, colony/plaque blots, and *in situ* hybridization. Dot blots can be used to identify unfractionated DNA or RNA molecules that have been immobilized on a nitrocellulose membrane. Plaque/colony blots detect DNA released from lysed bacteria or phage after immobilizing the DNA on a nitrocellulose membrane. *In situ* hybridization is used to detect DNA or RNA molecules in cytological preparations.

Nucleic-acid probes can be labeled by several methods. One method to uniformly label double-stranded DNA probes is described here. **Nick translation** describes the incorporation of a nick (or break in one strand) of a ds DNA molecule (Figure 5.7). Nicks are introduced at widely separated, random sites along



**Figure 5.7**  $^{32}\text{P}$ -labeling of double-stranded DNA can be carried out by nick translation. The asterisks indicate the location of radiolabeled phosphate groups that are inserted into the strand.

the DNA molecule by treating the DNA with small amounts of DNase. A nick exposes a free 3'-OH group and DNA polymerase I of *E. coli* will then remove nucleotides from the 5' side of the nick. The simultaneous removal of nucleotides from the 5' side and the addition of labeled nucleotides to the 3' side by DNA polymerase I results. If deoxynucleoside-5'-triphosphates (dNTPs) are radiolabeled with  $^{32}\text{P}$  and the nicks are random, the duplex DNA molecule will become labeled uniformly along its length as it incorporates radiolabeled dNTPs. The reaction may be carried out to label all four dNTPs or only one dNTP. Nick translation is particularly useful for producing large amounts of probe for multiple hybridization reactions and where a high probe concentration is required.

Nick translation kits are available from a number of commercial sources and provide instructions, a stock mixture of DNA polymerase I and DNase I, and a series of buffers lacking one or more unlabeled dNTPs. The radioactive dNTPs must be obtained fresh (within a few days) before the labeling reaction is set up because  $^{32}\text{P}$ -labeled dNTPs decay rapidly.

The use of radioactive probes requires that the experimenter obtain training in their safe use. Safety protocols are required to prevent contamination of the laboratory and to ensure safe disposal of the radioactive dNTPs. Most organizations require the training and licensing of scientists using radioactivity and regular inspections and reports to confirm its safe use and disposal (Davies 1994).

Kits are available to label DNA without using radioactivity, thereby reducing potential risks to the experimenter and eliminating the need for special disposal methods and the short half-life associated with radioactivity (Allefs et al. 1990). Although nonradioactive labels may not be as sensitive in detecting very small amounts of target DNA, they can be more rapid and less hazardous. Chemiluminescence and fluorescence provide a signal when an enzyme reacts with a chemiluminescent or chromogenic substrate. Alternatively, a signal can occur through excitation and emission of a fluorophore-labeled probe.

## 5.16 Removing DNA from Agarose Gels after Electrophoresis

Several methods have been developed to recover DNA from agarose gels: 1) electrophoresis onto a diethylaminoethyl(DEAE)-cellulose membrane, 2) electroelution into dialysis bags, and 3) low-melting temperature agarose gels. With method 1, fragments of DNA are separated by electrophoresis, a slit is cut in the gel immediately ahead of the DNA fragment of interest, and a sliver of DEAE-cellulose membrane is inserted. Electrophoresis is continued until the DNA in the band has been transferred to the membrane. The membrane is removed from the slit and washed and the DNA is eluted from the membrane.

Method 2, electroelution, allows recovery of a wide size range of DNA, but it is inconvenient. DNA is separated by electrophoresis in agarose gel containing ethidium bromide (EtBr). The band of interest is located with an ultraviolet lamp, and the band is cut from the gel with a razor blade. The gel fragment containing the DNA of interest is then placed in a piece of dialysis tubing, sealed, and placed into an electrophoresis tank. Electric current is passed through the bag to elute the DNA out of the gel and onto the inner wall of the bag. The polarity of the current is reversed to release the DNA from the wall of the bag, the bag is opened and the buffer containing the DNA is transferred to a clean tube.

Method 3 uses low-melting-temperature agarose gels. DNA of interest is electrophoresed, the band of interest is detected by staining with ethidium bromide, cut out, and placed into a clean tube. A buffer is added, and then the mixture is heated to 65°C to melt the agarose and centrifuged to leave the DNA in the aqueous phase and the agarose in the interface. The DNA in the aqueous phase can be purified with phenol. The DNA in the aqueous phase is then precipitated by ammonium acetate and cold (temperature) ethanol. At this point, the DNA is sufficiently pure to be digested by restriction enzymes or modified by ligases.

None of these methods is fully satisfactory in producing large amounts of pure DNA. Problems include the presence of inhibitors of enzymatic inhibitors in the agarose, which can affect subsequent DNA manipulations. Large fragments (>5 kb) of DNA are often inefficiently recovered from agarose gels because these longer fragments bind very tenaciously to purification matrices such as the DEAE-cellulose membrane. Small (<500 nanograms [ng]) amounts of DNA are recovered inefficiently from gels. The methods are labor-intensive so that recovery of DNA from gels is not readily performed on large numbers of samples. Kits, such as spin-columns, can be purchased from commercial sources to purify DNA from gels and remove impurities more efficiently than the three methods described here.

## 5.17 Restriction-Site Mapping

So far in this experiment, DNA has been cloned into a plasmid, amplified in *E. coli*, and specific sequence(s) have been identified by Southern blot analysis using a probe. Specific bands of DNA have been isolated and the DNA purified from the agarose gel.

Information about the cloned DNA can now be obtained by restriction-site mapping, DNA sequencing, and translation of the DNA into proteins using an expression vector. DNA sequencing is described in Chapter 7, and translation and expression of DNA are discussed in Chapter 6.

**Restriction-site mapping** is a relatively simple technique that provides a physical map of the sites in the DNA at which different restriction enzymes cut. One method for constructing restriction maps involves digesting the DNA with a series of different endonucleases in separate reactions. The products of each digestion are electrophoresed on agarose or polyacrylamide gels. DNA marker fragments of known size are electrophoresed in lanes adjacent to the DNA being examined to provide estimates of the lengths of the DNA fragments generated. DNA molecular markers of known size are available from commercial sources. The DNA is stained with EtBr, and the bands that were produced are examined under UV light and photographed.

After the single digestions are done, the DNA can be digested simultaneously with two restriction enzymes. Again, the size of the digestion products is analyzed by gel electrophoresis, using size markers and the samples of the first digest for comparison. If the products of the first and second digests are electrophoresed in adjacent lanes on the gel, it is possible to detect small differences in migration rate. Maps are built up from these data by a process of trial and error and basic logic. Based on the sizes of the DNA fragments generated, it is possible to define the *order* of the restriction sites and the approximate *distances* between them. The resolution of map distances depends on the accuracy with which the sizes of the DNA fragments can be estimated relative to those of the size markers. However, restriction maps are rarely accurate to <100–200 bp.

Restriction maps of DNA provide the experimenter with useful information for additional experiments. Furthermore, such restriction site maps can be used as in systematics or population genetics studies (see Chapters 12 and 13). You will use many of the techniques described in this chapter for other purposes, including preparing a genomic library, as described in Chapter 6.

## General References

- Ausubel, F. M., R. Brent, R. E., Kingston, D. M., Moore, J. G. Seldman, and J. A., Smith, et al., Eds., *Current Protocols in Molecular Biology*. Updated regularly since 1987. Humana Press, Totowa, NJ. Online versions are also available so that up-to-date protocols are provided.
- Barker, K., (2005). *At the Bench: A Laboratory Navigator*, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Berger, S.L., and Kimmel, A.R., Eds., (2005). *Guide to Molecular Cloning Techniques. Methods in Enzymology*, Vol. 152, Academic Press, San Diego, CA. An electronic version is available.
- Brown, T.A., Ed., (1998). *Molecular Biology LabFax, Volume 1: Recombinant DNA* (2nd ed.). Academic Press, San Diego, CA.
- Brown, T.A., Ed., (2002). *Essential Molecular Biology. A Practical Approach*, Vols. I and II (2nd ed.). IRL Press, Oxford, UK.

- Hames, B.D.H., and Higgins, S.J., Eds., (1985). *Nucleic Acid Hybridisation—A Practical Approach*, IRL Press, Oxford, UK.
- Jones, P., (1996). *Gel Electrophoresis: Nucleic Acids: Essential Techniques*. John Wiley & Sons, New York.
- Miller, J., (1992). *A Short Course in Bacterial Genetics. A Laboratory Manual and Handbook for Escherichia coli and Related Bacteria*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Primrose, S.B., and Twyman, R., (2006). *Principles of Gene Manipulation and Genomics*, 7th Edition. Blackwell Publishing, Malden, MA.
- Rio, D.C., Ares, M., Hannon, G.J., and Nilsen, T.W., (2011). *RNA: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Sambrook, J., and Russell, D., (2001). *Molecular Cloning: A Laboratory Manual*, 3rd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY. Three volumes provide detailed protocols of methods.
- Sambrook, J., and Russell, D., (2006). *Condensed Protocols from Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Sealey, P.G., and Southern, E.M., Eds., (1982). *Gel Electrophoresis—A Practical Approach*, IRL Press, Washington, DC.
- Streips, U.N., and Yasbin, R.E., Eds., (2002). *Modern Microbial Genetics* (2nd Ed.), Wiley, New York.
- Sullivan, W., Ashburner, M., and Hawley, R.S., Eds., (2000). *Drosophila Protocols*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Walker, J.M., and Rapley, R., Eds., (2008). *Molecular Biomechanics* (2nd ed.), Humana Press, Totowa, NJ.
- Williams, J.G., and Patient, R.K., (1988). *Genetic Engineering*. IRL Press, Oxford, UK.
- Wilson, K., and Walker, J., Eds., (2010). *Principles and Techniques of Biochemistry and Molecular Biology* (7th ed.), Cambridge Univ. Press, Cambridge, UK.

## References Cited

- Allefs, J.J.H.M., Salentijn, E.M.J., Krens, F.A., and Rouwendal, G.J.A., (1990). Optimization of non-radioactive Southern blot hybridization: single copy detection and reuse of blots. *Nucleic Acids Res.* **18**: 3099–3100.
- Balbas, P., Soberon, X., Merino, E., Zurita, M., Lomeli, H., and Valle, F., et al. (1986). Plasmid vector pBR322 and its special-purpose derivatives—a review. *Gene* **50**: 3–40.
- Boesenberg-Smith, K.A., Pessarakli, M.M., and Wolk, D.M., (2012). Assessment of DNA yield and purity: an overlooked detail of PCR troubleshooting. *Clin. Microbiol. News.* **34**: 1–6.
- Bowtell, D.D.L., (1987). Rapid isolation of eukaryotic DNA. *Anal. Biochem.* **162**: 463–465.
- Cox, R.A., (1968). The use of guanidinium chloride in the isolation of nucleic acids. *Methods Enzymol.* **12B**: 120–129.
- Davies, R.W. (1994). Working safely with radioactivity, pp. 471–477. In: *Cell Biology: A Laboratory Handbook*. Academic Press, San Diego, CA.
- Frank, S.A., (1994). Polymorphism of bacterial restriction-modification systems: the advantage of diversity. *Evolution* **48**: 1470–1477.
- Gilbert, W., (1991). Towards a paradigm shift in biology. *Nature* **349**: 99.
- Hoss, M., and Paabo, S., (1993). DNA extraction from Pleistocene bones by a silica-based purification method. *Nucleic Acids Res.* **21**: 3913–3914.
- Jeyaprakash, A., and Hoy, M.A., (2010). A DNA extraction procedure that allows mite specimens to be slide mounted: phytoseiid species evaluated as a model. *Exp. Appl. Acarol.* **52**: 131–140.
- Jowett, T., (1986). Preparation of nucleic acids. In: Roberts, D.B., Ed., *Drosophila: A Practical Approach*. IRL Press, Oxford, UK, pp. 275–286.

- Kessler, C., and Manta, V., (1990). Specificity of restriction endonucleases and DNA modification methyltransferases—a review (Edition 3). *Gene* **92**: 1–248.
- Martin, P., Boulukos, K.E., and Pognonec, P., (2006). REtools: a laboratory program for restriction enzyme work: enzyme selection and reaction condition assistance. *BMC Bioinformatics* **7**: 98.
- Miller, J.H., (1992). *A Short Course in Bacterial Genetics. A Laboratory Manual and Handbook for Escherichia coli and Related Bacteria*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Nature Methods, 2011. Training the kit generation. *Nature Methods* **8**: 983.
- Neidhardt, F.C., Ed., (1999). *Escherichia coli and Salmonella: Cellular and Molecular Biology*, 2nd ed. ASM Press. Washington, DC.
- Pingoud, A., Alves, J., and Geiger, R., (1993). Restriction enzymes. In: Burrell, M.M., Ed., *Methods in Molecular Biology. Volume 16: Enzymes of Molecular Biology*. Humana Press, Totowa, NJ, pp. 107–200.
- Roberts, R.J., Vincze, T., Posfai, J., and Macelis, D., (2005). REBASE—restriction enzymes and DNA methyltransferases. *Nucleic Acids Res.* **33**: D230–D232.
- Rohland, N., Siedel, H., and Hofreiter, M., (2004). Nondestructive DNA extraction method for mitochondrial DNA analyses of museum specimens. *Bio. Techniques* **36**: 814–821.
- Rowley, D.L., Coddington, J.A., Gates, M.W., Norrbom, A.L., Ochoa, R.A., and Vandenberg, N.J., et al. (2007). Vouchering DNA-barcoded specimens: test of a nondestructive extraction protocol for terrestrial arthropods. *Mol. Ecol. Notes* **7**: 915–924.
- Sambrook, J., and Russell, D.W., (2001). *Molecular Cloning. A Laboratory Manual*, 3d ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Snyder, L., and Champness, W., (1997). *Molecular Genetics of Bacteria*. ASM Press, Washington, DC.
- Southern, E.M., (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* **98**: 503–517.
- Southern, E.M., (2000). Blotting at 25. *Trends Biochem. Sci.* **25**: 585–588.
- Wilfinger, W.W., Mackey, K., and Chomczynski, P., (1997). Effect of pH and ionic strength on the spectrophotometric assessment of nucleic-acid purity. *Bio. Techniques* **22**: 474–481.

# Some Additional Tools for the Molecular Biologist

## Chapter Outline

<b>6.1 Overview</b>	215
<b>6.2 Introduction</b>	216
<b>6.3 The Perfect Genomic Library</b>	218
6.3.1 Lambda ( $\lambda$ ) Phage as a Vector	219
6.3.2 Cloning with Cosmids	226
6.3.3 Cloning in the Filamentous Phage M13	228
6.3.4 Phagemids	230
6.3.5 BACs	230
<b>6.4 cDNA Cloning</b>	230
<b>6.5 Enzymes Used in Molecular Biology Experiments</b>	234
<b>6.6 Isolating a Specific Gene from a Library if Whole-Genome Sequencing is Not Done</b>	234
<b>6.7 Labeling Probes by a Variety of Methods</b>	240
6.7.1 Synthesis of Uniformly Labeled DNA Probes by Random Primers	241
6.7.2 Synthesis of Probes by Primer Extension	241
6.7.3 End-Labeled Probes	241
6.7.4 Single-Stranded Probes	241
6.7.5 Synthetic Probes	242
<b>6.8 Baculovirus Vectors Express Foreign Polypeptides in Insect Cells</b>	242
<b>6.9 Expression Microarray Analysis</b>	243
<b>General References</b>	247
<b>References Cited</b>	247

## 6.1 Overview

Cloning of DNA has five essential components: 1) a method for generating exogenous DNA fragments, 2) reactions to join exogenous DNA fragments to a vector, 3) a method to introduce the vector into a host cell where the vector ensures the exogenous DNA is replicated, 4) methods for selecting or identifying the vectors that contain the introduced DNA (recombinant molecules), and 5) methods for analyzing the cloned DNA. Genomic libraries can be constructed in  $\lambda$  phage (a bacteriophage that attacks *Escherichia coli*) and other vectors. Complementary DNA (cDNA) libraries can be cloned into various vectors,

including cosmids or phagemids. Cloning into single-stranded M13 phage results in single-stranded DNA suitable for sequencing.

Cloning is feasible because a diverse array of enzymes is available to synthesize, ligate, and modify the ends of DNA molecules. Nucleic-acid or antibody hybridizations, chromosome walking, and sequencing can be used to screen libraries for genes of interest.

Cloning provides the basis for identifying specific genes, producing DNA copies of mRNA (=cDNA), and, in some cases, producing gene products (proteins) in *E. coli*, yeast, or insect cells by incorporating the DNA into expression vectors. cDNA libraries are used to determine which genes are being transcribed in a particular tissue or cell at a particular time and to provide clues important in understanding the development of insects. Two insect baculoviruses have been genetically engineered to express proteins (expression vectors), and they are used to produce proteins in insect cells in tissue culture or in intact lepidopteran larvae. Microarrays are used to determine the expression of hundreds or thousands of genes at a time in expression microarrays.

## 6.2 Introduction

The term *cloning* has multiple meanings. For example, Dolly the sheep was a “clone” because a nucleus of a donor somatic cell was inserted into an egg from which the original nucleus was removed. That egg was implanted in a host ewe and produced a lamb, Dolly, that was genetically identical to the sheep donating the nucleus. In this chapter, *cloning* also means that a single vector molecule (plasmid or phage or engineered versions of these) containing exogenous DNA is multiplied in cells so that multiple identical copies (clones) are produced. A **vector** is the agent used to replicate, or multiply, the exogenous DNA. Vectors are segments of DNA with an **origin of replication** so that it can be replicated after it is introduced into a host cell (origins of replication are essential for cloning). Vectors can be plasmids, bacteriophage, baculoviruses, or hybrid engineered molecules called cosmids and phagemids.

Chapter 5 introduced the use of plasmid vectors. This chapter introduces vectors derived from the *Escherichia coli* bacteriophage  $\lambda$ ; the single-stranded DNA bacteriophage M13; and engineered, hybrid vectors combining components from bacterial plasmids and  $\lambda$  called cosmids. Phagemids are engineered hybrid molecules that combine elements of plasmids and M13 vectors. Bacterial artificial chromosomes (BACs) are used to clone very large segments of DNA. The most commonly used host cell for cloning is *E. coli*, but others are used including the bacterium *Bacillus subtilis* or the yeast *Saccharomyces cerevisiae*. Insect cells

in tissue culture also can be used as hosts for baculovirus expression vectors (see Section 6.8).

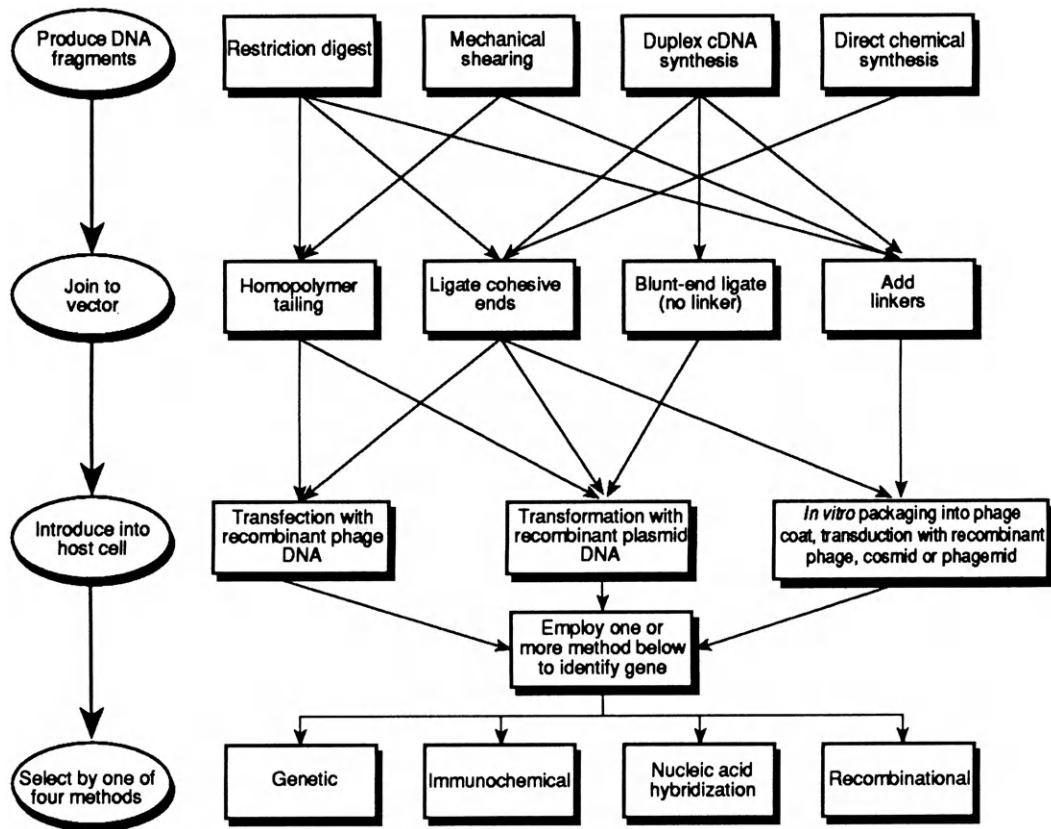
A multitude of vectors have been developed commercially for cloning. The choice of an appropriate vector depends upon the goal of your experiment. To develop a new vector requires an extensive knowledge of the biology and genetics of *E. coli* or other host cell, the plasmid or bacteriophage, and enzymology. It is impossible to describe all existing vectors and their uses; furthermore, such a description would become obsolete rather rapidly. Hundreds of vectors are commercially available, with new vectors engineered and advertised regularly.

Basically, a vector is a segment of DNA with an origin of replication that allows it to be maintained stably after it is introduced into its host cell. Most vectors contain unique restriction sites in a region of the vector that contains nonessential genetic information. The unique restriction sites are where exogenous DNA fragments can be inserted into the vector. When several cloning or restriction sites are combined into a single region, it is called a **polylinker** or multiple-cloning site.

One widely used plasmid vector series is a derivative of plasmid BR322 (pBR322). It was described and its structure illustrated in Chapter 5 (see Figure 5.2). Another plasmid vector series is the pUC group, and these plasmids contain a functional segment of the *E. coli lacZ* gene (see Figure 5.3). Thus, *E. coli* containing this plasmid produces blue colonies if provided with the appropriate substrate in the agar medium. If exogenous DNA is inserted into the cloning site, which is located within the *lacZ* gene, the *lacZ* gene sequence is disrupted, the gene product is no longer made, and the *E. coli* colonies are colorless, allowing the scientist to choose colonies that have been transformed with the vector that contains exogenous DNA. This is called blue-white selection. The pUC plasmids produce an increased copy number in *E. coli* that results in an increased yield of recombinant DNA molecules compared with the pBR322 plasmid series.

Cloning can be used to produce gene libraries, develop mutated genes for experiments, provide single-stranded DNA for sequencing, and permit eukaryotic genes to be translated in *E. coli*, insect tissue culture cells, or lepidopteran larvae. [Figure 6.1](#) identifies many of the steps and procedures involved in cloning, but a full description of all the techniques used is beyond the scope of this chapter.

Commercial companies will provide genomic or complementary DNA (cDNA) libraries for a fee if you provide the DNA or RNA. The availability of the



**Figure 6.1** Generalized scheme outlining the steps used in cloning DNA in *E. coli*. There are four major components: obtaining DNA fragments, joining them to the vector, introducing the recombinant molecule into an appropriate host cell, and identifying or selecting the recombinant DNA of interest. All of these steps can be achieved in several ways. Choosing an appropriate cloning scheme depends upon the goals of the experiment. (Revised from Old and Primrose 1989.)

polymerase chain reaction ([PCR], described in Chapter 8) makes the construction of libraries less important than formerly, especially if the goal is to isolate only one or a few genes. However, complete libraries are essential if the entire genome is to be sequenced. References at the end of this chapter provide additional information and protocols on constructing libraries.

### 6.3 The Perfect Genomic Library

No library is perfect. However, a perfect **genomic library** would contain all of the DNA sequences in the entire genome. The library would be stable and have a manageable number of overlapping clones. The clones would contain sufficiently large DNA segments that they could contain whole genes and their

flanking sequences. Ideally, the library could be amplified without loss or misrepresentation of sequences, and it could be stored for years without significant loss of information. Unfortunately, no single vector provides all of these desirable attributes.

Partially digested genomic DNA fragments can be cloned into bacteriophage  $\lambda$  relatively easily. The disadvantage is that the average  $\lambda$  library of an insect genome would contain  $>100,000$  clones, each with an insert that averages 15–20 kb. Cosmid vectors have an advantage in that the size of the inserted DNA can be two- to three-fold larger; therefore, fewer cosmids need to be evaluated to find the gene(s) of interest. Thus, cosmids can provide a significant advantage when it is important to work with an entire gene and its flanking sequences.

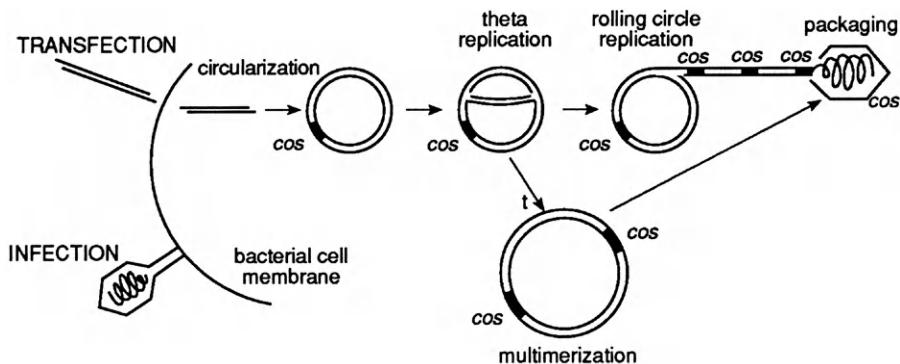
Genomic libraries constructed from random fragments of DNA are called “shotgun” libraries. To generate a genomic library for an insect may require  $10^6$  clones to ensure that all sequences in the genome are represented. The library only can be read if there is a key to open it. The key to libraries is some sort of probe (see Section 6.6).

### 6.3.1 Lambda ( $\lambda$ ) Phage as a Vector

Lambda ( $\lambda$ ) is a genetically complex, but well-studied, double-stranded-DNA bacteriophage of *E. coli*. The entire DNA sequence of the  $\lambda$  chromosome has been determined. Nearly 40% of the 48.5-kb chromosome is not essential for propagating the phage in its host. If this nonessential DNA is removed, approximately the same amount of exogenous DNA can be inserted. At each end of the linear DNA molecule there are short, single-stranded 5'-projections of 12 nucleotides, called **cos sites**, that are complementary to each other in sequence.

The cos sites enable the  $\lambda$  chromosome to circularize after the linear phage is injected into its *E. coli* host (Figure 6.2). After replication within the host cell, the  $\lambda$  DNA is in a linear form when it is packaged into a protein coat. The protein coat consists of an icosahedral head and a tail that ends in a tail fiber. The infective phage thus consists of the DNA molecule plus a protein head and tail. The protein coat allows phage particles to adsorb by the tip of their tail fiber to receptor sites on the outer membrane of their *E. coli* host cells. Adsorption is temperature independent, and is dependent upon the presence of magnesium ions.

A wild-type  $\lambda$  has two phases to its life cycle: temperate and lytic. Although the temperate or **lysogenic** phase is of little interest to the genetic engineer,  $\lambda$  which has been genetically modified to serve as a vector retains many characteristics of the second, or lytic, phase.

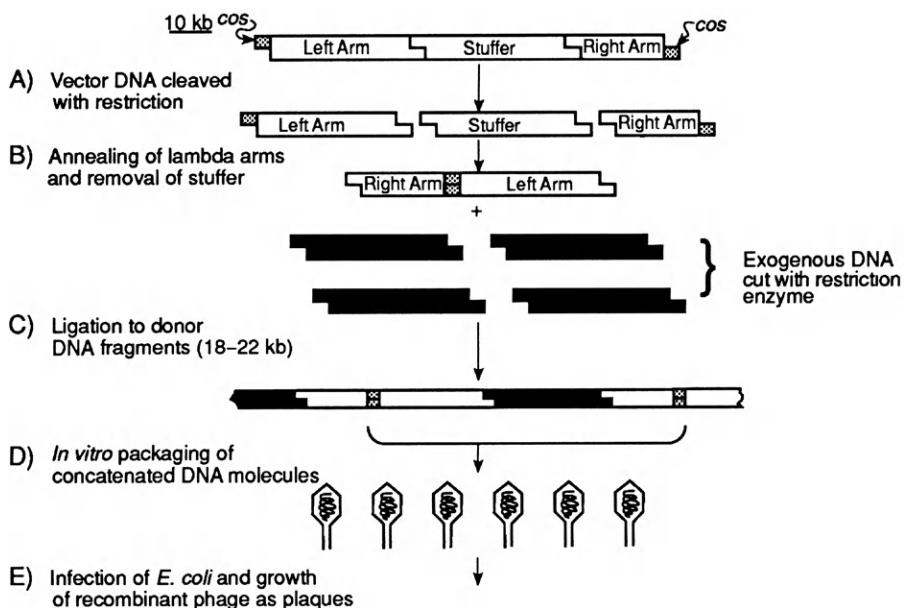


**Figure 6.2** The typical lytic cycle of bacteriophage  $\lambda$  begins when the phage adsorbs to an *E. coli* cell by the tail fiber. The linear DNA molecule is injected into the host cell where it circularizes by annealing at the *cos* sites. The  $\lambda$  DNA replicates by a “rolling circle” mechanism, producing long concatenated molecules with the individual chromosomes annealed at the *cos* sites. Later in the infection cycle, phage genes for the protein head and tail are turned on. When these components are produced, the replicated DNA is packaged into the icosahedral head after the concatenated DNA is cut at the *cos* sites. If the DNA is too long or too short, it will not be packaged in the head, but will be lost. Lysis of the host cell then occurs, typically releasing  $\approx 100$  progeny phage that enter new cells to replicate.

In the **lytic** phase, early DNA transcription establishes the lytic process, middle genes replicate and recombine the DNA, and late genes produce protein for packaging the DNA into mature phage particles. Phage DNA is replicated in a “rolling circle” mode (Figure 6.2). Multiple copies of the replicated DNA molecules are assembled in a linear tandem array, with the termini of each molecule joined at the *cos* sites. The *cos* sites form the recognition site of a specific endonuclease that cuts the DNA during the packaging process so that a single DNA molecule is inserted into the head of the protein coat. In summary, in a lytic infection, the phage takes over the host cell machinery: phage DNA is replicated, head and tail proteins are made, the replicated DNA is packaged, and the host cell is lysed to release  $\approx 100$  infective particles.

In the temperate or lysogenic phase, most phage functions are repressed and lysis is avoided. In lysogeny, the  $\lambda$  DNA is inserted into the host chromosome by site-specific recombination and the phage genome (called a prophage) is replicated as part of the *E. coli* chromosome. Nearly all  $\lambda$  vectors used in genetic engineering lack the ability to enter the lysogenic phase.

$\lambda$  has been genetically improved as a vector: 1) Genes in the central region of the chromosome that code for events associated with recombination and lysogeny have been deleted and can be replaced with exogenous DNA (Figure 6.3). 2)  $\lambda$  vectors have been engineered to contain cloning sites to facilitate the insertion of exogenous DNA. Engineered versions are of two major types: insertion



**Figure 6.3** Schematic outline of a bacteriophage  $\lambda$  replacement vector. A linear molecule contains the *cos* sites, a left (L) and right (R) arm, and a “stuffer” region with nonessential DNA. The vector is digested with an appropriate restriction enzyme, the stuffer fragment is removed, and the two arms anneal. Exogenous DNA that has been cleaved with an appropriate restriction enzyme is added, and the fragment is ligated in. Exogenous DNA fragments of 18 to 22 kb can be incorporated because these molecules can be successfully packaged by *in vitro* packaging. *Escherichia coli* is infected with the  $\lambda$ , and thousands of individual plaques are produced. Each plaque contains many thousands of replicas (clones) of a single phage containing exogenous DNA.

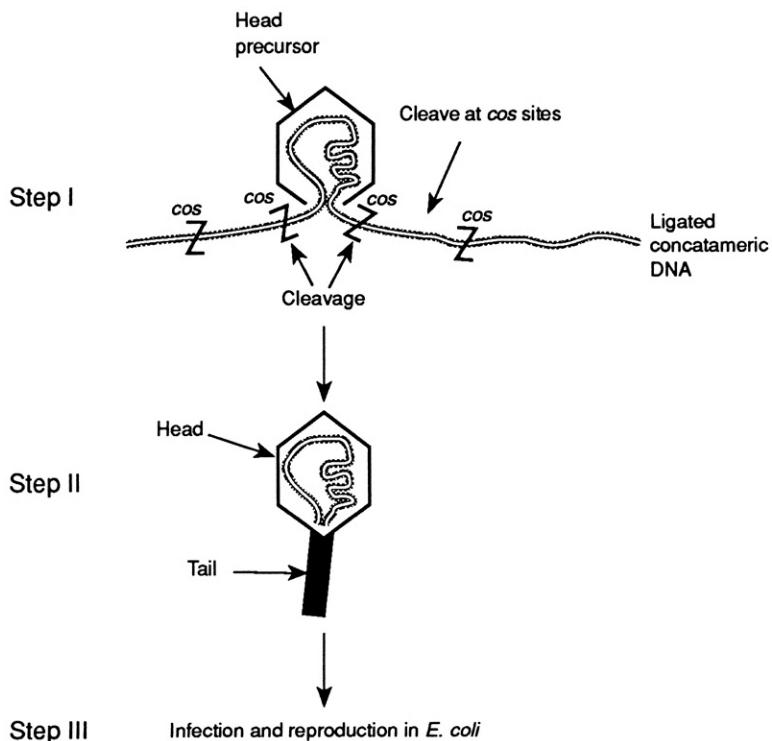
and replacement. **Insertion vectors** have a single target site at which foreign DNA can be inserted, whereas **replacement vectors** have a pair of sites defining a fragment that can be removed and replaced by foreign DNA (Figure 6.3).

Once exogenous DNA has been inserted into the  $\lambda$  vector, this molecule can be multiplied (cloned) by inserting it into host *E. coli* cells in one of two ways: transfection and *in vitro* packaging. Naked  $\lambda$  DNA (lacking a protein coat) can be introduced into *E. coli* cells in a process called transfection. **Transfection** is the infection of bacteria by viral nucleic acid alone. The efficiency of transfection is  $>10^4$  recombinant clones per microgram of donor DNA. This efficiency would suffice for the construction of genomic libraries from species with small genomes. However, larger genomes, such as those of insects, require a more efficient method of inserting the vector DNA into *E. coli*. The way to increase efficiency in introducing recombinant  $\lambda$  DNA molecules into *E. coli* is called ***in vitro* packaging**. By incorporating the recombinant DNA molecules into phage protein coats, *E. coli* can be infected much more readily, thereby increasing the

likelihood of producing complete genomic libraries. Efficiency of infection of *E. coli* with “packaged” DNA can be  $10^6$  recombinants per microgram of vector DNA, an increase in efficiency over transfection by nearly 2 orders of magnitude.

*In vitro* packaging involves several steps and specific conditions (Figure 6.4). One critical condition is the size of the inserted DNA. The amount of exogenous DNA inserted into the vector must be regulated carefully: the cos sites must be separated by DNA that is  $\approx 78\text{--}105\%$  of the length of the wild-type chromosome. In an insertion vector, only 14 kb of DNA, or less, can be cloned. In a replacement vector, up to 22 kb of DNA can be inserted. In replacement vectors, a pair of restriction sites flanks the nonessential central region of the phage DNA called the “stuffer region.” When the stuffer region is excised and the insert DNA is ligated into this region, a DNA molecule is produced that can be packaged efficiently.

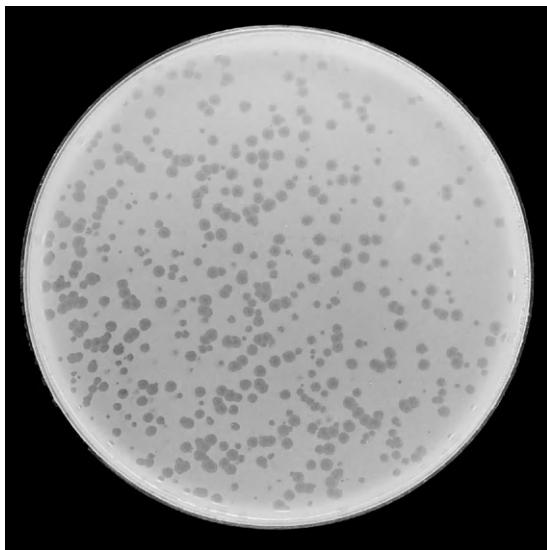
*In vitro* packaging requires the following components: 1) the DNA molecules to be packaged, 2) high concentrations of phage-head-precursor protein, 3) proteins that participate in the packaging process, and 4) phage tails. These



**Figure 6.4** *In vitro* packaging of cloned DNA in  $\lambda$  involves providing a protein head and tail precursor. The DNA is cut at the cos sites, and if the DNA is  $\approx 50\text{ kb}$ , it will be packaged in the head. The complete phage is then used to infect *E. coli* and amplify the recombinant molecule.

packaging ingredients are obtained by combining a very concentrated mixture of the lysate from two different  $\lambda$  strains that are lysogenic. One mutant  $\lambda$  strain can progress no further in the packaging process than the prehead stage because it carries a mutation in a gene (gene *D*) and therefore accumulates this precursor. The other mutant  $\lambda$  strain is prevented from forming any head structure by a mutation in a different gene (gene *E*), but it can produce the tail component. In the mixed lysate, both head and tail components become available so that a complete phage can be assembled that contains recombinant DNA.

Transfer of DNA into the *E. coli* host involves adsorption of phage to specific receptor sites on the outer membrane of the *E. coli*. Because phage will adsorb to dead cells and debris, only healthy bacterial cultures should be used to reduce loss of efficiency. Once the *E. coli* have taken up the phage, they are plated out on nutrient agar and allowed to grow at least overnight at 37 °C. Infected bacterial colonies grow, but clear areas (**plaques**) consisting of lysed cells will be seen surrounded by an opaque background of unlysed bacteria (Figure 6.5). Each plaque represents an original bacterial cell that was infected, ideally only by a single  $\lambda$ . Thus, each plaque should contain multiple copies of a *single kind* of recombinant DNA molecule. Even the smallest plaque is likely to contain sufficient phage DNA to be detectable by plaque hybridization, a probe technique described in Section 6.6.



**Figure 6.5** An agar plate with *E. coli* and plaques caused by the bacteriophage  $\lambda$ . Each clear area indicates where a single bacterium initially was infected with  $\lambda$ . After replication, the emerging  $\lambda$  attacked adjacent *E. coli* and lysed them, resulting in a clear “plaque” of killed *E. coli* and  $\lambda$  on the surface of the agar.

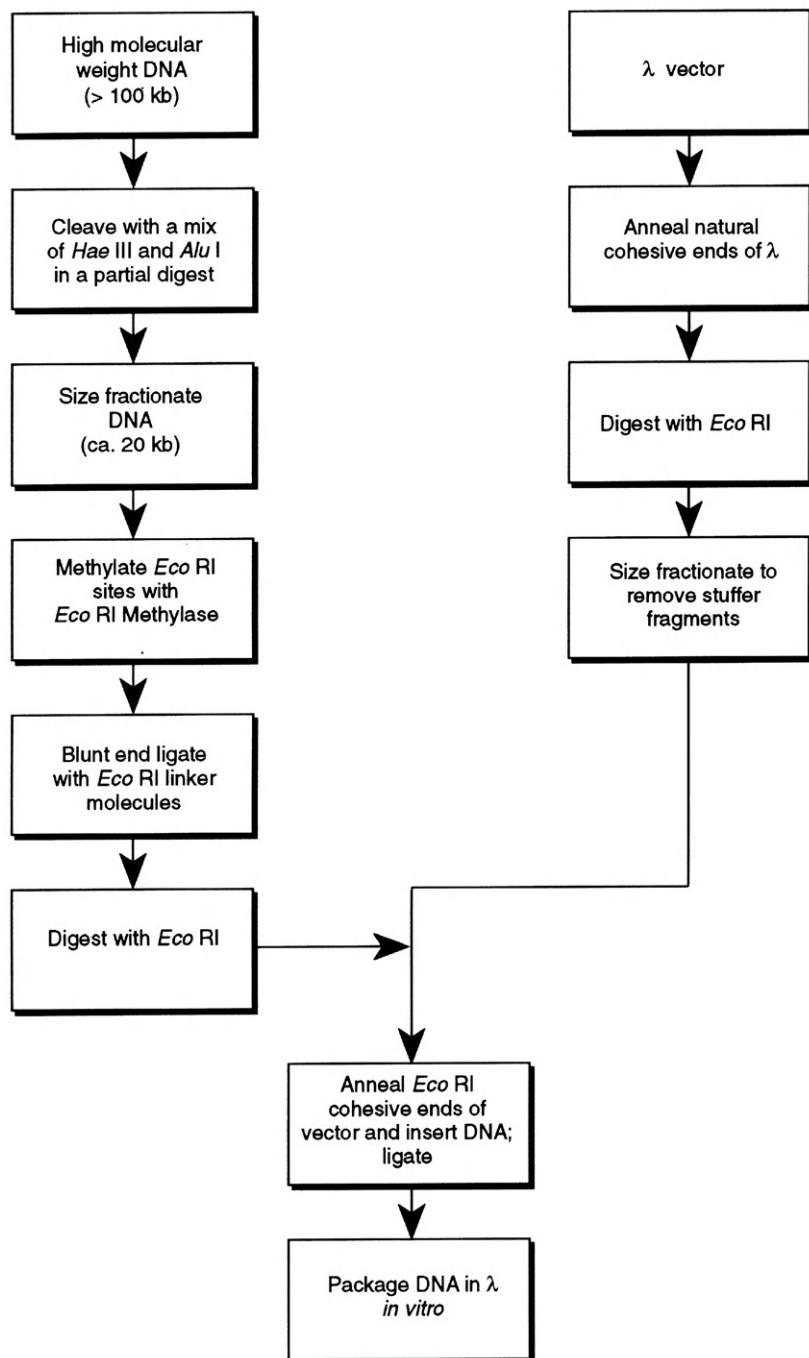
A visual method for identifying plaques containing  $\lambda$  with recombinant DNA involves the use of the *lacZ* gene. This gene codes for part of the  $\beta$ -galactosidase enzyme, an enzyme that cleaves lactose to produce glucose and galactose. Inserting exogenous DNA into this gene inactivates synthesis of  $\beta$ -galactosidase. To identify *E. coli* colonies containing recombinant phage, the agar is made up with a lactose analog called **X-Gal** (5-bromo-4-chloro-3-indolyl- $\beta$ -D-galactopyranoside). X-Gal is cleaved by  $\beta$ -galactosidase into a product that is bright blue. If exogenous DNA has inserted into and disrupted the  $\beta$ -galactosidase gene, plaques look white or colorless. Plaques containing  $\lambda$  without the exogenous DNA will produce the blue color.

Figure 6.6 outlines the steps involved in one strategy for producing a representative genomic library in a  $\lambda$  replacement vector. The genomic DNA and the vector DNA can be prepared simultaneously. In this example, the genomic DNA is cut with a mixture of two restriction endonucleases (*Hae*III and *A*lul) in such a manner that the DNA is only partially digested. The genomic DNA is then sized so that fragments of  $\approx$ 20kb are isolated. Meanwhile, the  $\lambda$  vector DNA is digested with the restriction enzyme *Eco*RI and purified so that the stuffer fragments are removed. The genomic DNA has linker molecules added to it before the annealing reaction. When the genomic DNA and the vector DNA are combined, they anneal at their complementary cohesive ends and are ligated together. The last step involves providing a protein coat for the DNA by *in vitro* packaging.

Commercial cloning kits simplify the procedures considerably because such kits provide vectors, enzymes, *in vitro* packaging materials, and detailed protocols. It is even simpler to supply genomic DNA to a company who will provide a complete genomic library for a fee. However, a few points should be made about constructing a genomic library.

The genomic DNA to be cloned must be of high molecular weight and not excessively sheared during its isolation from the insect. High-molecular-weight DNA is needed because the DNA will be partially digested with a restriction enzyme to generate a random collection of DNA fragments and the fragments need to be at least 20kb. The DNA to be digested actually must be longer than 20kb so that after digestion both ends of the fragment will have cohesive ends. DNA fragments with only one cohesive end (and one broken end) cannot be inserted into vectors. DNA shorter than 20kb won't be packaged into the phage and will be lost. Thus, DNA extraction should be carefully executed to avoid damaging or shearing the DNA.

The genomic DNA fragments ideally will be representative of the entire insect genome. If the restriction enzymes that are used cut relatively frequently



**Figure 6.6** One method for producing a representative genomic library in a  $\lambda$  replacement vector. Two parallel processes are carried out: preparation of the exogenous DNA and preparation of the vector DNA. These DNAs are then ligated together and packaged *in vitro*. The specific restriction enzymes and ligation method can be varied.

compared with the desired fragment size, a *partial* digestion will produce a set of overlapping fragments. Ideally, these fragments will be a nearly random array of the entire genome. However, it is possible that some regions of the genome will not be represented because they lack the appropriate cleavage sites for the enzyme used, or the DNA may not be cleaved with equal efficiency, particularly in heterochromatic regions containing repetitive elements. It is especially difficult to clone telomeres and centromeres, for example. Furthermore, some regions of the genome may be toxic to their *E. coli* host, so no clones will be produced.

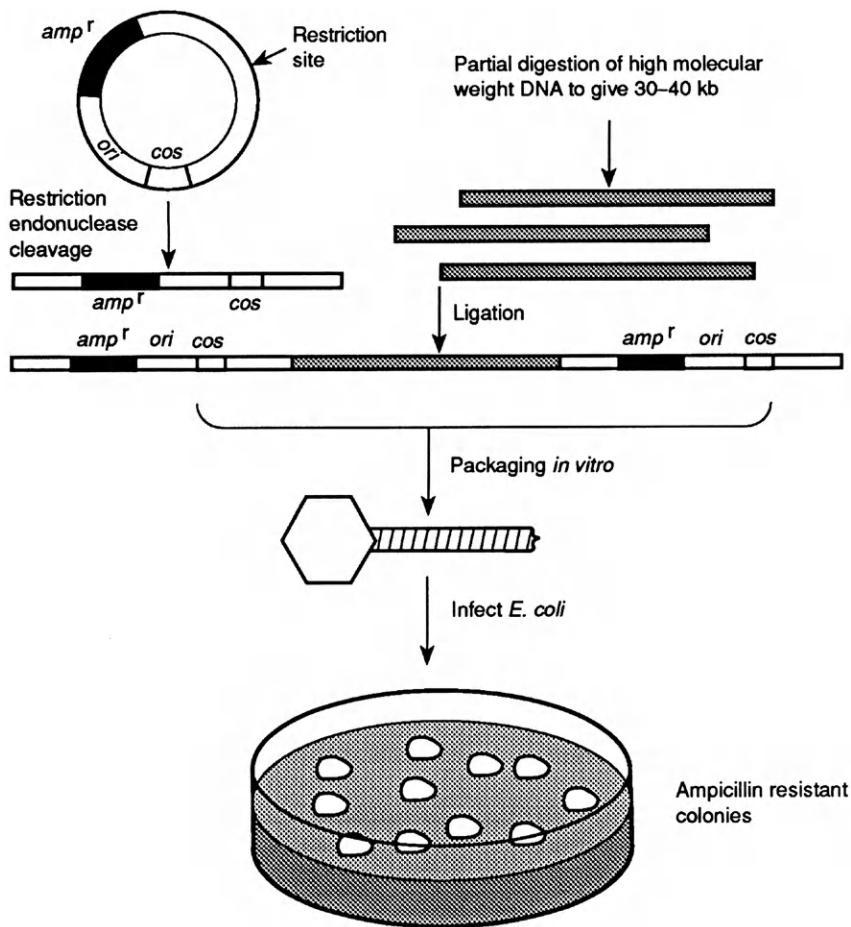
After the genomic DNA has been partially digested with an appropriate restriction endonuclease, the DNA fragments are size-fractionated by centrifugation or by gel electrophoresis. This separates out the DNA fragments greater than or smaller than 18–22 kb. Preparing a representative library requires high-quality vector DNA. Large-scale preparation of  $\lambda$  DNA should yield pure preparations that lack the central stuffer region, or it can reinsert back into the vector later on. Removal of the central stuffer region is carried out by centrifugation, elution, or electrophoretic separation. However, it is difficult to remove all of the stuffer fragments, so it is important to determine, by appropriate controls, how often the stuffer is reinserted back into the vector.

Optimizing concentrations of vector and exogenous will result in adequate ligations and thus a good library. Because a portion of the DNA molecules to be ligated will have damaged ends, the ratio of vector to insert DNA will probably have to be determined empirically in small reactions. It is desirable to produce long concatenated molecules that can be cut at the cos sites and packaged. Likewise, the appropriate ratios of ingredients used for *in vitro* packaging will have to be determined empirically. Once the DNA has been packaged, the phage can be stored at 4°C for years. Alternatively, the phage can be amplified by multiplication in *E. coli*.

Commonly used vectors derived from  $\lambda$  include the gt and EMBL series.  $\lambda$  gt10 was designed for cloning short DNA fragments, especially cDNA.  $\lambda$  gt11 is used to construct cDNA libraries, as described in Section 6.4. DNA properly aligned with the *lacZ* gene in  $\lambda$  gt11 will be expressed in *E. coli* as a **fusion protein**. EMBL vectors are a family of replacement vectors that provide a high level of reproduction in *E. coli*, polylinker cloning sites, and the ability to select for recombinant phage. EMBL3 and EMBL4 vectors, or their derivatives, are particularly useful for constructing genomic libraries.

### 6.3.2 Cloning with Cosmids

Cosmids are engineered vectors that combine characteristics of both plasmids and phage. They have been constructed to include a fragment of  $\lambda$  that includes



**Figure 6.7** Outline of procedures used in cloning with a cosmid vector. This vector contains a *cos* site, a restriction site for inserting exogenous DNA, and a gene for ampicillin resistance. Exogenous DNA is cut with an appropriate restriction enzyme, as is the vector. The vector and exogenous DNA are ligated together, producing a recombinant molecule of 37–52 kb that can be packaged in  $\lambda$  by *in vitro* packaging. The packaged vector infects *E. coli*, injecting its DNA into the host, where it circularizes and multiplies. *Escherichia coli* cells that receive the cosmid are distinguished from cells that are not infected by their ability to survive on media containing ampicillin.

the *cos* site (Figure 6.7). Cloning into cosmids is similar to cloning in  $\lambda$ . It involves digesting exogenous DNA with a restriction enzyme, cutting the cosmid vector with a compatible restriction enzyme, combining the two, and ligating them.

Once the exogenous DNA is inserted into the cosmids, cosmids are packaged in a manner similar to that used with  $\lambda$ . Packaging the cosmid recombinants into phage coats provides a useful method for selecting the size of the inserted DNA.

What is significant about cloning with cosmids is that larger DNA fragments, 32–47 kb, can be inserted into the vector and still be packaged.

After *in vitro* packaging, cosmids are used to infect a suitable *E. coli* strain. Infection of *E. coli* involves injection and circularization of the cosmid DNA, but no phage protein is produced. Transformed *E. coli* cells are identified on the basis of their resistance to a specific antibiotic.

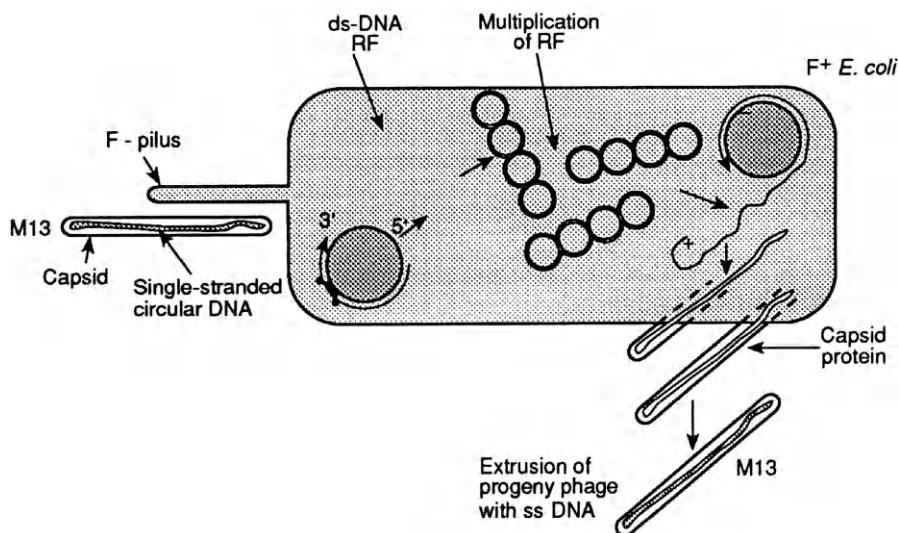
Although having a large capacity for DNA fragments is a benefit in cloning with cosmids, it can be a detriment. If, during a partial digestion with restriction enzymes, two or more genomic DNA fragments join together in the ligation reaction, a clone could be created with fragments that were not initially adjacent to each other. This situation could be a problem if the researcher is interested in the relationship between a gene of interest and the DNA surrounding it. The problem can be overcome by size fractionating the partial digest and dephosphorylating the foreign DNA fragments to prevent them from ligating together, but this process makes cosmid cloning very sensitive to the exact ratio of insert and vector DNAs. If the ratio is imbalanced, vector DNAs could ligate together without containing any exogenous DNA insert. This outcome is resolved by treating the vector to two separate digestions, thereby generating vector ends that are incapable of ligating to each other after phosphatasing.

Commonly used cosmid vectors include the pJB8 and the pcosEMBL family. The pcosEMBL family was designed to simplify isolation of specific recombinants from cosmid libraries and speed up isolating large regions of complex genomes in an ordered array of overlapping clones (i.e., chromosome walking, which is described in Section 6.6). The vectors in this family differ by having different cloning sites and different numbers of cos sites. Recombinant cosmids in our theoretical experiment can be identified by rearing *E. coli* in the presence of the antibiotic ampicillin ([Figure 6.7](#)).

### 6.3.3 Cloning in the Filamentous Phage M13

M13 is a filamentous phage of *E. coli* that contains a circular single-stranded DNA molecule that is 6407 nucleotides in length. M13 only infects strains of *E. coli* that have F pili because the site where this phage adsorbs seems to be at the end of the F pilus ([Figure 6.8](#)). Replication of M13 does not result in host-cell lysis. However, the infected cells grow and divide more slowly and extrude up to 1000 virus particles into the medium.

Replication of M13 phage (which is single-stranded) involves conversion of the DNA to a double-stranded (ds) or replicative form (RF). The ds RF multiplies



**Figure 6.8** Outline of the infection cycle of the bacteriophage M13. The single-stranded phage attaches to the F pilus of *E. coli*, injects its DNA into the host, and begins to produce  $\approx 100$  copies of double-stranded (RF) molecules. DNA replication then shifts to producing ss DNA molecules that are extruded through the host cell wall, during which time they are coated with a capsid protein coat. M13 has been engineered as a vector and is used to produce ss DNA molecules, a technology that is particularly useful in DNA sequencing reactions.

until  $\approx 100$  RF molecules are produced within the cell (Figure 6.8). The replication of the RF then becomes asymmetric due to the accumulation of a viral-encoded binding protein that is specific to single-stranded (ss) DNA. The binding protein binds to the M13 DNA and prevents synthesis of a complementary strand. Subsequently, only ss DNA is synthesized and extruded from the host cell. As the ss M13 DNA molecules move through the *E. coli* cell membrane, the DNA-binding protein is removed and the M13 DNA is coated with capsid protein.

M13 has many advantages as a vector. First, ss DNA is required in several applications, including the dideoxy DNA sequencing method (described in Chapter 7). Second, ss M13 vectors allow the genetic engineer to combine cloning, amplification, and strand separation of a ds DNA fragment in one operation. Third, because the phage DNA is replicated in a ds circular (RF) intermediate stage, the RF DNA can be purified and manipulated just like a plasmid. Fourth, both RF and ss DNA will transfect competent *E. coli* cells and yield either plaques or infected colonies. Fifth, it is possible to package DNA up to 6 times the length of the wild-type M13 DNA.

The M13 phage does not contain excess DNA that can be removed. However, there is a 507-bp region that contains the origins of replication for the viral DNA and its complementary strand. In most vectors derived from M13, foreign DNA is inserted at this site. Polylinkers and the *lacZ* gene have been inserted into M13 vectors so that white plaques are formed instead of blue if exogenous DNA is inserted into the *lacZ* cloning site.

#### 6.3.4 Phagemids

A phagemid is an engineered vector that contains plasmid and M13 components. Phagemids provide another method for obtaining ss DNA. Phagemids carry two replication origins, one a standard plasmid origin and the other derived from M13. The M13 origin is crucial for the synthesis of ss DNA. However, production of ss DNA requires enzymes and coat proteins coded by phage genes, which are lacking in phagemids. As a result, cells containing a phagemid vector must be coinfectected with a helper phage if ss DNA is desired. The helper phage converts the phagemids into ss DNA molecules that are then assembled into phage particles and secreted from the cell.

#### 6.3.5 BACs

**Bacterial artificial chromosomes** (BACs) were developed so that scientists could insert very large segments of DNA and have it cloned in *E. coli*. BACs can incorporate an average of 140 kb of exogenous DNA and are based on the *E. coli* F-factor plasmid; the *E. coli* F-factor plasmid maintains a low copy number in bacterial cells, thus minimizing the possibility of recombination that could result in the production of chimeric clones ([Shizuya et al. 1992](#)). Basic BACs do not contain a selection system or reporter genes suitable for expression in eukaryotic cell lines. However, BACs have been modified as expression vectors in eukaryotic cells ([Kim et al. 1998](#)).

BACs were extensively used in the Human Genome Project, a project that used Sanger sequencing methods (see Chapter 7). The advantages of the BAC vector are that it can maintain up to 300 kb of DNA, and the clones are highly stable in their host even after 100 generations. BACs thus facilitate the construction of DNA libraries of complex genomes because they allow fuller representation of all sequences.

### 6.4 cDNA Cloning

cDNA is DNA that is *complementary* to the mRNA that is transcribed from the gene. Because the mRNA is processed, it will lack introns and regulatory

elements usually found in arthropod genes. Thus, a cDNA clone will contain the DNA sequence of the protein of interest, but it will lack introns and probably not contain the control sequences that regulate gene expression. cDNA cloning is used to produce a cDNA library or to produce probes for screening genomic libraries.

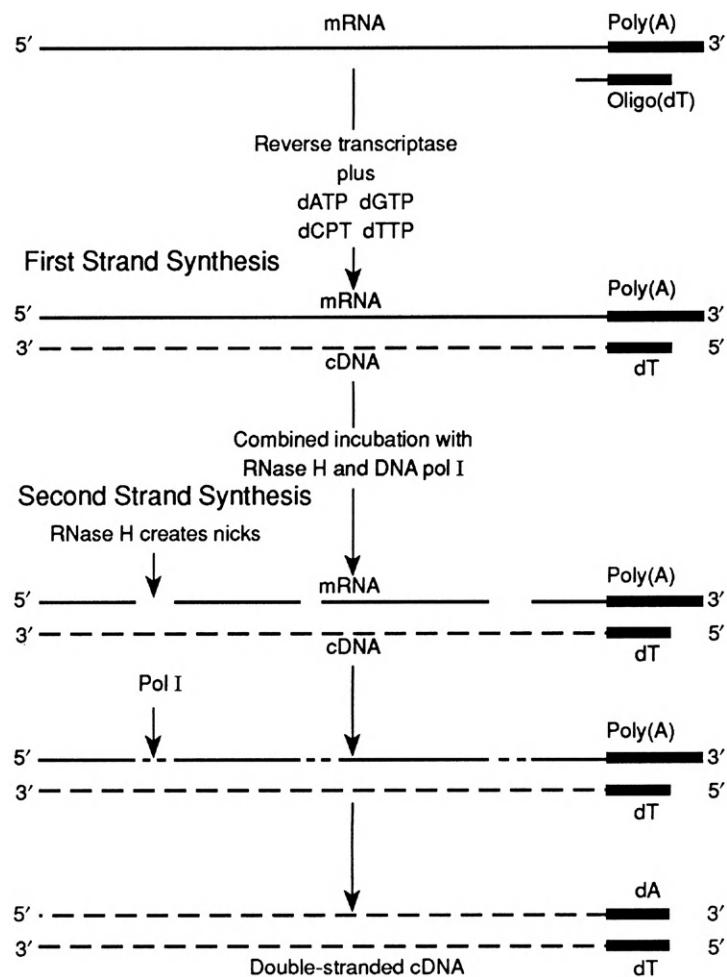
A **cDNA library** allows the genetic engineer to clone *only those genes that are active at a specific time or in specific tissues*. Genes that are not actively transcribed into mRNA will not be represented in the cDNA library. Thus, a cDNA library usually contains fewer clones than a genomic library. The gene of interest may occur in a frequency of one in  $10^3$  or one in  $10^4$  clones. By contrast, a single-copy gene may be present in a genomic library in a frequency of only one in  $10^5$  to one in  $10^6$  clones. Another benefit of a cDNA library is that it is possible, if an appropriate **expression vector** is used, to express a gene in a host such as *E. coli* or yeast. This enables the genetic engineer to produce large amounts of a specific gene product.

The quality of a cDNA library depends on the quality of the mRNA used as the template and the fidelity with which it can be **reverse transcribed** into cDNA. Messenger RNA, together with a suitable primer, and a supply of deoxyribonucleoside triphosphates can be converted into a ss DNA molecule with the enzyme **reverse transcriptase** (Figure 6.9).

The cDNA cloning process involves two steps: 1) the first strand of cDNA is produced by reverse transcription, and 2) a strand that is complementary to the first strand is then synthesized using DNA polymerase so that a ds cDNA molecule is produced. The primer used to synthesize the first DNA strand is usually an oligonucleotide consisting of deoxythymidine (dT) residues because it can hybridize to the 3' poly(A) tails of template mRNA and potentially give rise to full-length copies of the complementary DNA strand. Once the ds cDNA molecule has been synthesized, it can then be inserted into a plasmid or phage vector that is capable of replicating in *E. coli* and, in some cases, of being translated into a protein.

A key to producing cDNA is the enzyme **reverse transcriptase**. Reverse transcriptase is capable of two functions *in vitro*: a polymerase activity and a ribonuclease H activity. The polymerase activity requires 1) a template RNA molecule hybridized to a DNA primer with a 3'-OH group and 2) all four dNTPs to synthesize a DNA molecule that is a faithful *complement* of the mRNA.

Cloning a cDNA library is more complex than cloning a genomic library into  $\lambda$  or cosmids. Before beginning the process, the goals of the project must be



**Figure 6.9** cDNA cloning involves two steps. In the first step, an oligo(dT) anneals to the poly(A) region of mRNA. The enzyme reverse transcriptase and dATPs, dCTPs, dGTPs, and dTTPs are provided to produce the first cDNA strand. Synthesis of the second strand of the cDNA involves additional incubation with RNase H and DNA polymerase I. The double-stranded cDNA produced is a complement to the mRNA and thus lacks introns or regulatory sequences.

carefully considered and the basic approach chosen after deciding how the cDNA library will be screened to identify the gene(s) of interest (Kimmel and Berger 1987). For example, if antibodies will be used to identify clones capable of synthesizing specific peptides, the cDNA should be cloned into expression vectors to produce proteins.

The cloning techniques vary in the type of primer used, the method for second-strand synthesis, and methods for coupling the cDNA to the vector,

which can be either a plasmid or  $\lambda$ . Commercially available reverse transcriptases can synthesize copies of mRNA sequences that are  $>3$  kb. However, the transcripts often terminate prematurely, making clones containing the 5' end of the mRNA rare.

Figure 6.9 outlines the synthesis of double-stranded cDNA from mRNA. Messenger RNA can be prepared for cDNA cloning by affinity chromatography on oligo(dT) cellulose (or by commercially available kits). The reaction is preceded by a brief heat denaturation of the mRNA to eliminate its secondary structure, because reverse transcriptase is inhibited if the mRNA exhibits a secondary structure. The polyadenylated mRNA, the primer, and the reverse transcriptase are combined. The primer in this case is a short sequence of (dT) residues. The product of the first strand synthesis is a hybrid of mRNA and the synthesized cDNA. The first strand is used as a template-primer complex to make the second strand of DNA. The enzyme RNase H is used to introduce gaps in the mRNA strand. At the same time, DNA polymerase I uses the primer-template complexes formed by RNase H to synthesize a double-stranded DNA.

Once the double-stranded DNA is synthesized, it is inserted into a vector. To insert it into a vector, the synthesized molecule needs to have ends that can be ligated into the vector. One option is to make the cDNA blunt-ended by end filling with the Klenow fragment of DNA polymerase I and then ligating it into a vector that has been cut with a restriction enzyme that produces a blunt end. Another option involves the addition of cohesive ends to the cDNA so that it will ligate into a vector more easily. There are three methods to add cohesive ends to the ds DNA: 1) tailing with terminal transferase, 2) adding linkers, and 3) adding adaptors. The details of carrying out these procedures can be found in many cloning protocols.

There are  $\approx 10,000$  different mRNA molecules in an average insect cell. At least 200,000 cDNA clones should be generated to be sure that a representative cDNA library is constructed. If the desired clone is a single-copy gene, then it will be rare, so powerful screening methods are required to isolate the clone of interest.

Isolating RNA is more difficult than isolating DNA. Preparation of mRNA requires the absolute elimination of **ribonucleases** (RNases) from glassware, pipets, tips, and solutions. Anything that might contaminate the reactions with RNase must be eliminated, including hair, dust, and sneezes. Even fingerprints contain enough RNase to degrade your RNA. Furthermore, RNase is a very hardy enzyme and difficult to eliminate. Phenol extraction followed by ethanol precipitation was a common technique for isolating RNA, but various kits designed for

that purpose are available now. Once RNA has been isolated, it must be evaluated for quality, often by agarose gel electrophoresis.

## 6.5 Enzymes Used in Molecular Biology Experiments

Many enzymes used in genetic engineering have been mentioned in this and previous chapters. **Table 6.1** summarizes their names, principle activity(ies), sources, and functions in genetic manipulations. Enzymes used to synthesize DNA include terminal transferase, DNA polymerase I, and reverse transcriptases. Enzymes that modify DNA include S1 nuclease, exonuclease III, Bal31 nuclease, and DNase I. There are >1400 restriction endonucleases that can cleave DNA in a predictable manner. T4 and *E. coli* DNA ligases join DNA molecules. Calf intestinal phosphatase (CIP) and T4 polynucleotide kinase are used to modify the 5' ends of DNA molecules.

## 6.6 Isolating a Specific Gene from a Library if Whole-Genome Sequencing is Not Done

The production of a library is only a first step. Because the cost of sequencing whole genomes has decreased dramatically, it is now possible to screen libraries by sequencing them (see Chapter 7). However, if whole-genome sequencing is not done, the information in a library can only be obtained if the library can be screened. Screening identifies specific genes and provides information about genome organization, or information about gene regulation. The ability to screen a library is dependent upon the availability of a **probe**. As pointed out in Chapter 5, a probe is a molecule that can be labeled with radioactive isotopes or by nonradioactive methods.

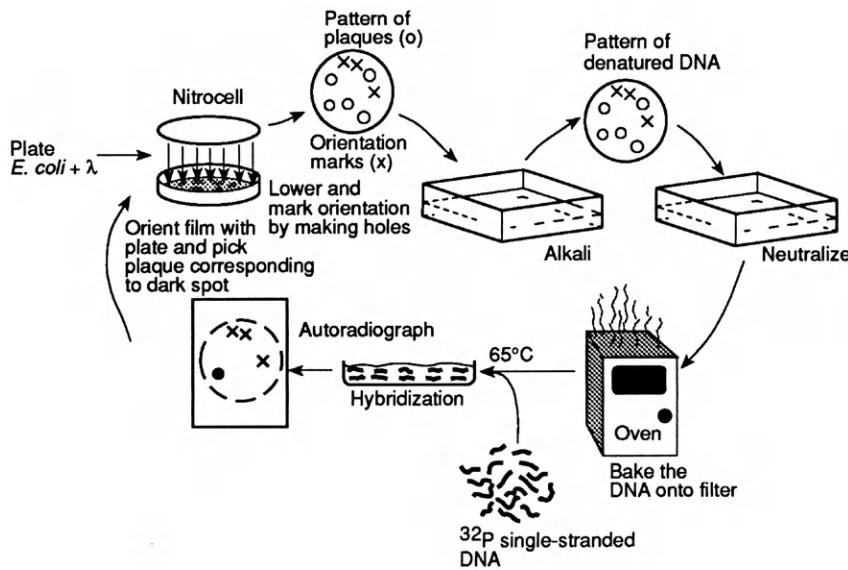
There are four ways to obtain a suitable probe for your library. 1) The amino-acid sequence of the protein is known for the species being studied, or a related species, and can be used to predict and synthesize the sequence of an oligonucleotide (oligo) hybridization probe. Because the genetic code is degenerate, the probe used may actually incorporate a mixture of oligos with optional bases, especially in the third site of the codon. 2) The gene of interest has already been cloned from a related organism, so that it can be used as a heterologous hybridization probe. For some genes, particularly housekeeping genes, conservation of functional domains in proteins has been extensive, so that probes from other species can be used effectively. 3) The protein is abundant in a particular tissue so the relevant clone can be identified by its relative abundance in a tissue-specific cDNA library. 4) If the protein has been purified, it can be used to generate an antibody against it. The antibody can be used to identify recombinant cells for the presence of the specific enzymes.

**Table 6.1: Enzymes Useful for DNA Manipulation.**

Enzyme type (name, source)	Functions in genetic engineering
<b>Enzymes that synthesize DNA</b>	
DNA polymerase I ( <i>E. coli</i> )	5' to 3' DNA synthesis of template DNA with a primer; exonuclease functions (5' to 3' and 3' to 5') used to correct errors in DNA synthesis <i>in vivo</i> ; generate labeled DNA probes by nick translation; synthesize cDNA
Klenow fragment of DNA polymerase I ( <i>E. coli</i> )	DNA synthesis without 5' to 3' exonuclease ability; makes ds DNA from ss DNA; used in dideoxy sequencing; DNA labeling by random priming or end filling; converts 5' overhangs of DNA cut with restriction enzymes to blunt ends; removes 3' overhangs to form blunt ends
T4 DNA polymerase (phage T4)	Exonuclease in 3' to 5' direction; fill in overhanging ends of DNA cut with restriction enzymes
T7 DNA polymerase (phage T7)	3' to 5' exonuclease activity used in DNA labeling; converts 3' overhangs to blunt ends
<i>Taq</i> DNA polymerase ( <i>Thermus aquaticus</i> )	DNA synthesis at 60–70 °C in PCR; several cloned versions are available, as are DNA polymerases from other microorganisms
<i>Pfu</i> DNA polymerase ( <i>Pyrococcus furiosus</i> )	This polymerase and others have 3' to 5' exonuclease activity that allows them to remove mismatches; used to amplify DNA fragments up to 40 kb by the PCR
Reverse transcriptases (from several RNA tumor viruses)	Synthesizes single strand of DNA from messenger RNA; used to produce cDNA libraries
Terminal transferase (mammalian thymus)	Adds residues to any free 3' terminus; used to add poly(dG) and poly(dC) to two DNA molecules to be joined
<b>Enzymes that degrade DNA</b>	
S1 nuclease ( <i>Aspergillus</i> )	Degradates ss DNA endonucleolytically; removes projecting 3' regions of ss DNA in cloning and S1 mapping
Exonuclease III ( <i>E. coli</i> )	Degradates one of two strands of ds DNA from 3' end of a blunt-ended double helix or from a projecting 5' end
Bal 31 nuclease ( <i>Alteromonas espejiana</i> )	Degradates both strands of ds DNA with blunt ends
DNase I (pancreas)	Introduces random nicks in ds DNA before labeling by nick translation; produces random fragments for shotgun cloning and sequencing in M13; study chromatin structure; study DNA-protein complexes
<b>Enzymes that join DNA</b>	
T4 DNA ligase	Seals ss nicks between adjacent nucleotides in ds DNA molecule, requires ATP; used to ligate two restriction fragments of DNA together in cloning
<i>E. coli</i> DNA ligase	Same as for T4 DNA ligase, but requires NAD <sup>+</sup>
<b>Enzymes that modify the 5' ends of DNA</b>	
Calf intestinal phosphatase	CIP removes 5'-phosphate groups to generate a OH terminus; prevents unwanted ligation of DNA fragments during cloning; used for end-labeling DNA probes
T4 polynucleotide kinase	Adds phosphates to 5'-OH ends; used in chemical cleavage method of DNA sequencing; used to add linkers or adapters in cloning
<b>Enzymes that cut DNA</b>	
Restriction endonucleases (many bacteria)	Type I, II, and III, >1400 types known; cleaves DNA, produces predictable termini, either blunt, 5' overhang, or 3' overhang

Once a probe is obtained and labeled, it is used in DNA hybridization experiments to identify those clones that contain the DNA of interest. **Hybridization** involves immobilizing DNA samples from different clones on a solid support (such as a nitrocellulose or nylon membrane) and then probing the unknown DNA with a labeled DNA or RNA sequence to identify the clones that contain the sequence of interest. Identification is possible because the labeled probe can base-pair with the desired DNA and then be detected by autoradiography or other methods. There are several different DNA hybridization techniques, including the Southern-blot analysis described in Chapter 5. Another DNA hybridization technique is plaque screening.

**Figure 6.10** illustrates **plaque screening** of *E. coli*. First, *E. coli* cells are infected with  $\lambda$  that contain exogenous DNA and allowed to grow on an agar substrate. A nitrocellulose filter is laid onto the *E. coli* lawn and plaques. The precise orientation of the filter is marked. Some of the phage in the plaques containing the phage are adsorbed onto the filter, where they release their DNA. The DNA is



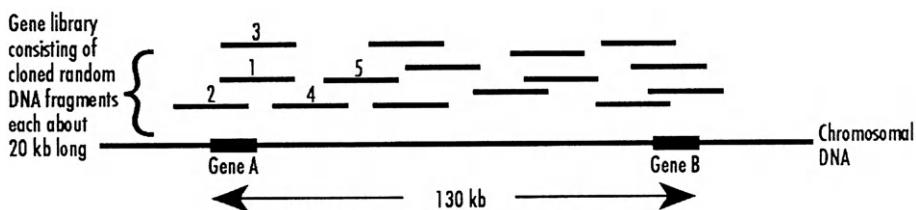
**Figure 6.10** Plaque screening can locate specific genes. Plaque screening involves *in situ* hybridization of *E. coli* that have been transformed with a  $\lambda$  vector. *Escherichia coli* infected with recombinant phage are plated out. A nitrocellulose filter is laid on top of the bacterial lawn and plaques. Some of the phage in a plaque adsorb to the filter. The filter is treated with an alkali to denature the phage DNA, neutralized, baked in an oven to immobilize the DNA, and placed in a solution with a radiolabeled DNA probe. The probe base pairs with sequences in the DNA that are complementary and identifies plaques that may contain the gene of interest. It is then possible to go back to the plate, pick a few phage from those plaques, and multiply them in *E. coli*.

denatured by an alkali treatment and then brought to a neutral pH. After denaturation, the now single-stranded DNA on the filter is incubated with a probe. The probe base-pairs with the specific nucleotide sequence from the gene of interest, but not with DNA from plaques containing other genes.

The position of the probe that is hybridized to the immobilized DNA on the filter can be located by autoradiography. The filter and the original agar substrate are then compared using the marker, and the corresponding plaque is located on the original agar substrate. A few phage can be picked from each plaque that yielded a spot on the X-ray film. The phage from that plaque is used to infect individual new *E. coli* cultures to produce multiple copies of that phage.

Plaque hybridization allows several hundred thousand plaques to be screened at once, so a single-copy gene can be isolated from thousands of clones. Because the DNA that was inserted into the  $\lambda$  vector was cut at random, it is likely that more than one clone (plaque) will contain the DNA of interest. Ideally, at least one clone isolated by the probe will contain the complete gene, but this outcome can only be determined after the DNA has been sequenced, a technique that is described in Chapter 7.

Another technique used to identify specific DNA sequences is called **chromosome walking** (Figure 6.11). It is particularly useful with *Drosophila* and less useful with insects for which less genetic information is available. Chromosome walking is used to isolate a gene of interest for which no probe is available. The gene of interest *must* be linked to a marker gene that has been identified and cloned. This marker gene is used as a probe to screen a genomic library. All fragments



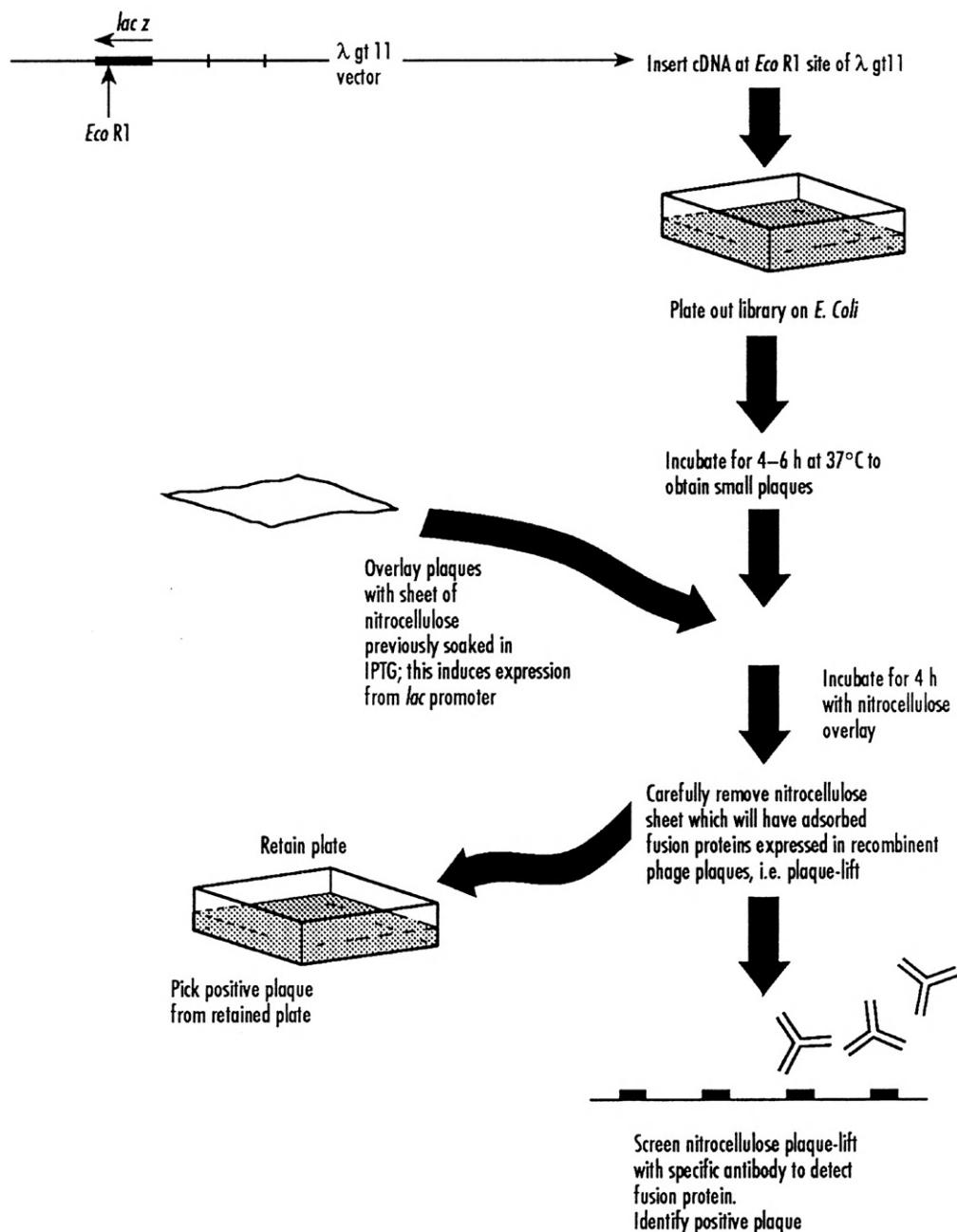
**Figure 6.11** Chromosome walking is used to identify a gene of interest when a probe is not available. It can only be carried out when it is known that the target gene is linked to another gene that has been cloned and sequenced. First a clone that contains gene A is isolated in fragment 1. This fragment is sequenced and new probes are synthesized that contain sequences from each end of the fragment. The new probes are used to identify overlapping DNA clones in the library on each side of fragment 1, i.e., clones 2, 3, and 4. Clone 4 can, in turn, be sequenced, and used as a probe to identify clone 5, and so on until gene B is reached.

containing the marker gene are selected and sequenced. The fragments are then aligned, and those cloned segments farthest from the marker gene in both directions are subcloned for the next step. The subclones are used as probes to screen the genomic library again to identify new clones containing DNA with overlapping sequences. As the process is repeated, the nucleotide sequences of areas farther and farther away from the marker gene are identified, and eventually the gene of interest will be found. As shown in [Figure 6.11](#), the goal is to identify gene B, for which no probe is available. However, sequences of a nearby gene (A) are available in cloned fragment 1. In a large and random genomic DNA library, many overlapping cloned fragments will be present. Thus, clone 1 can be used as a probe to identify overlapping clones 2, 3, and 4. Clone 4 subsequently can be used as a probe to identify clone 5 until gene B is reached.

Once a gene has been identified in a genomic library, its DNA sequence can be determined. However, a DNA sequence by itself is of limited value. If you do not know something about the gene product, it may be difficult to determine unambiguously which sequences are the coding regions and which are introns. Intron boundaries may be established based on similarities to sequences of known introns (consensus sequences). If the gene product is unknown, it may be possible to identify the sequenced gene's function by comparing the DNA sequence with other sequences in DNA databases, although a fully convincing match is not always found. Thus, going from clone to DNA sequence to gene product may be a challenge, particularly if genes are being studied for which there are no known gene products.

One solution is to attempt to express the gene to obtain a gene product. To be expressed, genes require a promoter and, often, upstream control sequences. A variety of expression vectors have been developed to express cloned genes in *E. coli*. Such vectors require *E. coli* promoters if the eukaryote sequence is to be expressed. A detailed description of *E. coli* expression vectors is provided by [Pouwels \(1991\)](#). Of particular interest to entomologists is the use of baculovirus expression vectors to express insect genes, as well as other eukaryotic genes, as described in Section 6.8.

cDNA libraries can be screened by hybridization screening or by expression screening. If a protein of interest has been purified and part or all of the protein sequence has been obtained, then it is possible to predict the sequence of synthetic oligonucleotides that can be used as a hybridization probe to detect the appropriate cDNA clone(s). Alternatively, if an antibody to the protein is available, it can be used to identify the clone(s) of interest if the cDNA library is cloned into an expression vector ([Figure 6.12](#)). This vector contains the *lacZ* gene



**Figure 6.12** Identifying specific genes also can be done by immunochemical screening. A  $\lambda$  gt11 library in *E. coli* is screened in a manner similar to plaque hybridization except that the gene(s) of interest are identified by binding of a specific antibody to a fusion protein.

of *E. coli*, and it has a unique *EcoRI* restriction site near the end of the gene. A cDNA cloned into this site in the correct orientation and reading frame will produce a fusion protein. Upon lysis of the *E. coli* cells, the protein is released and picked up on nitrocellulose in just the same way as in plaque screening. The plaque containing the interesting cDNA clone can be detected by incubating the filter with a specific antibody.

It is possible to determine the difference in abundance between two different mRNA populations. Thus, mRNA produced from different tissues from the same organism or mRNA produced from a tissue before and after a specific induction signal can be compared by differential screening. A cDNA library is prepared from one of the two mRNA populations, and the two copies are immobilized on filters. The filters are then screened twice, once with highly labeled cDNA prepared from one of the two mRNA populations and once with a probe from the other mRNA population. By comparing the signals produced on the two filters probed with the different probes, it is possible to determine whether mRNA sequences are present in one population but absent or rare in the alternative mRNA population. However, this method is used less often now that gene chip methods are available (see Section 6.9).

## 6.7 Labeling Probes by a Variety of Methods

**Nucleic-acid hybridization** is used for many different purposes in molecular genetics. Formerly, nearly all phases of cloning and characterization or analysis of DNA involve hybridizing one strand of nucleic acid to another by complementary-base pairing. Nucleic-acid hybridization relies on the fact that two single-stranded nucleic-acid molecules with complementary bases (DNA with DNA and DNA with RNA) are able to pair via hydrogen bonds. The length of the homologous sequences determines the strength of the hybridization, as do the experimental conditions (temperature and pH) and the degree of sequence homology. The Southern blot, a nucleic-acid hybridization method, and one method for labeling probes (nick translation) were described in Chapter 5. Other nucleic-acid hybridization techniques include colony or plaque hybridization, and Northern-blot analysis, in which the immobilized nucleic acid is RNA instead of DNA.

Several labeling methods other than nick translation are available and are outlined here. The success of nucleic-acid hybridization often relies on methods to introduce a label into cloned segments of DNA or RNA. Each labeling technique has optimal sizes, efficiency, and different amounts of nucleic-acid template that are required. One measure of the efficiency of radiolabeling is the

specific activity of the label. Specific activity refers to the amount of radioactivity per milligram of probe DNA. The specific goals of the project will determine which labeling technique is used. Detailed protocols are available in a variety of laboratory manuals and kits. The safe use of radioactivity for labeling requires that the user obtain specific training in handling procedures and disposal.

Nonradioactive probes using biotin and chemiluminescent labels are available in kits and are safer for novices to use. These labels may be sufficiently sensitive for most applications. However, if very small amounts of DNA or RNA are being screened, radioactive probes may be most sensitive.

#### ***6.7.1 Synthesis of Uniformly Labeled DNA Probes by Random Primers***

Short oligonucleotides can serve as primers for DNA synthesis by DNA polymerases on single-stranded templates. If the primers used are random in sequence, they will form hybrids at many different locations along the template strand so that the strand being synthesized will incorporate a labeled dNTP randomly along its length. If reverse transcriptase is used for synthesis, the template can be RNA. If DNA is the template, then the Klenow fragment of DNA polymerase I is used.

#### ***6.7.2 Synthesis of Probes by Primer Extension***

Primer extension is used to synthesize probes from denatured ds DNA. It can be used to produce probes from denatured, closed circular DNA or from denatured, linear ds DNA. The purified DNA is mixed with random primers, denatured by boiling, and labeled dNTPs and the Klenow fragment of DNA polymerase I are added to carry out synthesis of the probe. Random primers anneal to the denatured DNA and the product is synthesized by primer extension. Probes prepared by random priming are usually 400–600 nt.

#### ***6.7.3 End-Labeled Probes***

A variety of methods are available to introduce labels at either the 3' or 5' ends of linear DNA. Usually, only a single label is introduced at one end of the molecule. Both DNA and RNA can be end-labeled. The advantages to end labeling are that the location of the labeled group is known and very small fragments of DNA can be labeled, including restriction-digest fragments.

#### ***6.7.4 Single-Stranded Probes***

Single-stranded DNA, cDNA, or RNA probes have an advantage over ds probes because more probe is available to hybridize with the nucleic acid of interest.

Single-stranded probes should not anneal to themselves so that hybrids composed of reannealed probes cannot be made. RNA probes do not need to be denatured before being used because they are already single-stranded. Double-stranded DNA probes must be denatured before using, which produces two strands. If only one of the DNA strands has been labeled, the unlabeled strand can dilute the reaction mixture.

Single-stranded probes are prepared from DNA templates by synthesizing labeled DNA that is complementary to sequences cloned in a bacteriophage vector such as M13 or a phagemid. RNA probes can be produced by transcription of ds DNA in a vector with a powerful promoter derived from *E. coli* bacteriophages T7 and T3 by a bacteriophage DNA-dependent RNA polymerase. The labeled transcript produced is complementary to one of the two template strands. The probe can therefore be used as strand-specific probes in hybridization reactions. cDNA probes are used to isolate cDNA clones of genes that are expressed in specific cells or tissues.

#### **6.7.5 Synthetic Probes**

The knowledge of the sequence of a few amino acids in a protein will allow a specific gene to be isolated with a synthetic probe. Automated machines can synthesize short segments of ss DNA in which the sequences are defined precisely. Probes also can be produced that consist of alternative sequences, as determined by the degeneracy of the genetic code. Because there are 64 possible codons and only 20 amino acids, most amino acids are coded for by more than one codon. Thus, a probe that consists of a mixture of degenerate sequences can be used to screen libraries.

### **6.8 Baculovirus Vectors Express Foreign Polypeptides in Insect Cells**

*Escherichia coli*, infected with plasmid or phage expression vectors, has been used to express foreign eukaryotic genes. However, biologically active proteins are difficult to produce in *E. coli* because *E. coli* cannot make posttranslational changes such as **glycosylation** and **phosphorylation**, processes that are essential to effective protein function in eukaryotes. As a result, eukaryotic expression vectors have been developed for use in yeast and insect cells.

The most effective expression vectors used in insect cells were engineered from baculoviruses (Luckow and Summers 1988, Maiorella et al. 1988, Jarvis et al. 1990, O'Reilly et al. 1992, van Oers 2011). **Baculoviruses** are viruses with ds, circular DNA genomes contained within a rod-shaped protein coat. The Baculoviridae are divided into three subgroups: nuclear polyhedrosis viruses (NPVs), granulosis

viruses, and nonoccluded viruses. Their host range is limited to insects, so the viruses are considered safe for vertebrates. Likewise, the lepidopteran cells used are free of human pathogens (van Oers 2011). The proteins produced are used for vaccine preparations or diagnostics or functional studies. The first vaccine commercially produced in the baculovirus system protects against classical swine fever (or hog cholera). The first vaccine for human use protects against cervical cancer, and another product is used to treat prostate cancer.

Most NPVs primarily infect lepidopterans, where they produce nuclear inclusion bodies in which progeny virus particles are embedded. Polyhedrin is the protein component of the crystalline matrix that protects the viral particles when they are outside their insect host. Several NPVs have been used as biological pesticides in pest-management programs, including baculoviruses for the control of codling moth *Cydia pomonella* on apple, the velvet bean caterpillar *Anticarsia gemmatalis* in soybean, and the cotton bollworm *Helicoverpa armigera* (van Oers 2011). Perhaps the most extensively studied baculovirus is *Autographa californica* NPV. It has a relatively broad host range, and the life cycle of wild-type baculovirus begins when caterpillars eat the protein matrix (polyhedrin), thereby releasing the virus particles. Virus replication occurs within host cells, but the protein matrix is not produced early in the infection. However, when the caterpillar is near death, the virus resumes polyhedrin production until  $\approx 20\%$  of the insect cell proteins consist of polyhedrin.

The *A. californica* nuclear polyhedrosis virus (AcNPV) and the *Bombyx mori* nuclear polyhedrosis virus (BmNPV) have been exploited as vectors to carry exogenous DNA into insect cells in cell culture or into living silk moth larvae to produce novel proteins (Maeda 1989). The productivity of baculovirus expression vectors is based on the extremely high efficiency of its polyhedrin gene promoter. The polyhedrin promoter enables very large amounts of the desired polypeptides to be produced. The level of expression of foreign gene products varies, depending on the specific protein being produced.

In principle, any foreign gene could be expressed in the baculovirus–insect cell expression system. However, cytoplasmic proteins are expressed at higher levels than secreted glycoproteins, and transmembrane proteins are even more difficult to produce.

## 6.9 Expression Microarray Analysis

Microarray analysis of gene expression was adopted and used extensively by human geneticists who wanted to identify genes important in disease or to determine what drugs are effective (Wu et al. 2005; MAQC Consortium 2006,

2010; Gohlmann and Talloen 2009). Entomologists also use microarrays to evaluate gene expression and function. Microarrays are being used to study gene function in insects, with the goal of understanding development and responses to environmental cues such as pathogens, temperature, and parasitism (Gibson 2002, Johnson et al. 2009). The responses of plants to herbivory by insects can be studied using microarrays (Maleck et al. 2000, Schenk et al. 2000, Reymond 2001, Kempema et al. 2007).

Each insect cell contains the same DNA, yet each cell and tissue type is different, presumably due to differences in gene activity. Expression microarrays allow geneticists to determine which genes are on and which are off in specific cells, tissues, or organisms treated to different environmental conditions. To produce microarrays, sequences of genes are needed and microarrays can hold the sequences of an entire genome. Microarrays containing DNA, cDNAs, oligonucleotides (oligos), or proteins allow geneticists to determine the expression level of a large number of genes in a single experiment. Some microarrays containing insect genes are commercially available because the community of scientists studying the insect is large and the complete genome has been sequenced and annotated (for example, *Drosophila*). Companies and core facilities at universities also produce specific microarrays for fees, upon request. For example, Indiana University produces a gene expression microarray for analysis of the parasitoid *Nasonia vitripennis*, a wasp that has had its genome completely sequenced (Werren et al. 2010, Center for Genomics and Bioinformatics 2012). It also is possible to make your own microarrays (Stewart 2000), although there is concern that the data obtained from these are less reproducible (Beisvag et al. 2011). Microarrays are expensive, but the cost is declining and, relative to the amount of data that can be obtained from a single experiment, the investment may be cost-effective.

DNA microarrays are similar to Southern blots in that they rely on complementary-base pairing between complementary nucleotides for specificity of the signal (except for protein microarrays). Each spot on a microarray contains multiple identical strands of DNA (or oligos) and the sequence on each spot is unique. Each spot should represent one gene, and thousands of spots are arranged in orderly rows and columns on a solid surface in a precise and unique manner. Glass, silicon chips, or nylon membranes provide the solid surface and are approximately the size of a microscope slide (Lockhart and Winzeler 2000). The location and sequence of each spot are recorded. DNA chips or microarrays can be made with oligonucleotides that are synthesized *in situ* or by conventional synthesis, followed by on-chip immobilization. The array then is exposed to labeled sample DNA, hybridized, and complementary sequences can be

determined (Ramsay 1998). The technology of depositing nucleic acids on glass slides at very high densities has allowed the miniaturization of nucleic-acid arrays with dramatic increases in experimental efficiency and information content. Arrays with >250,000 different cDNAs per square centimeter can be produced (Lockhart and Winzeler 2000).

To conduct an experiment in which gene activity of two different tissue types (or the two sexes, or different developmental stages) are compared, cells are broken open and then the mRNAs present are extracted and converted into complementary DNA (cDNA) copies. The cDNAs are labeled with different dyes (fluorophores) and applied to the microarray. The cDNAs bind to (hybridize with) the DNA template on the microarray based on complementary-base pairing. The colors of the spots are analyzed. If, for example, one cDNA sample has a red label and the other cDNA sample has a green label, there will be four color options on the microarray: red spots, green spots, no color (where no cDNAs bound to the spot), and yellow (where both cDNAs bound to the spot). The microarray is laser scanned, and data are recorded and analyzed with computer software that can estimate the intensity of the colors. Because the location of each spot and its sequence are known, scientists can determine which genes are expressed at a specific stage of development or in specific tissues.

However, scientists are concerned that measuring mRNA levels (as cDNAs) may not be the best way to measure gene expression. It is not clear how closely levels of mRNAs relate to levels of their corresponding proteins, indicating that studies will have to be done to determine how to accurately quantify protein levels and their differential stability in the organism (White et al. 1999, Andrews et al. 2000).

The use of gene chips has provided some surprises and enormous amounts of data (Andrews et al. 2000, Quackenbush 2001). There is almost too much data produced, and data analysis methods are still under development. For example, the question needs to be answered, "Does gene expression indicate function?" At present, the conclusion is that there is a correlation between distinct expression profiles and function, but expression should not be taken as sufficient evidence for function. For example, not all genes involved in a function such as DNA replication are expressed periodically during the cell cycle and some genes that do not need to be cell-cycle regulated are transcribed in a periodic manner (Lockhart and Winzeler 2000).

Gene chip or microarray data require careful analysis because experimental evidence shows there is a disparity between the relative expression levels of mRNA and their corresponding proteins. Thus, expression information from both mRNA and proteins is required to understand a gene network (Dutt and

Lee 2000). Unfortunately, errors in gene chips have been found that also could lead unwary biologists to erroneous conclusions (Knight 2001). Despite the fact that microarray production is heavily automated, errors may creep in because bacterial cultures used to amplify the plasmids with the cDNAs can become contaminated. Technicians can make errors such as loading plates into the robots the wrong way around or taking samples from the wrong well for sequencing. Estimates suggest between 1 and 5% of the clones in even the best-maintained microarray sets do not contain the sequence they are supposed to contain. Even microarrays based on oligos can contain errors if the sequences in the databases are wrong or the wrong strand from the DNA helix is used (the noncoding strand) (Knight 2001). Other errors can occur when inadequate experimental controls are used or replications are not conducted. Erroneous results will continue to be published until the faulty chips and experimental design methods are corrected. At least, the sequences of the spot concerned should be verified by sequencing and by comparing the result using alternative methods of monitoring gene expression (Knight 2001).

Quality-control methods are being developed for experimental design and analysis of data (Taylor et al. 2008, Kauffmann and Huber 2010). Careful sample collection, data collection, and experimental design are essential to a successful experiment with microarrays. Experiments on global gene expression may yield data for thousands of genes, forcing the experimenter to consider processes, functions, and mechanisms about which we know very little. More sophisticated systems are needed to represent the data and incorporate sequence, genetics, gene expression, homology, regulation, function, and phenotype information in an organized and usable form (Lockhart and Winzeler 2000). At present, different researchers may use different arrays and methods of analyzing the data, which makes it difficult to compare the results from different laboratories or from different microarray platforms (Ionnidis et al. 2009). Efforts to resolve the problems include projects such as Empowering the Microarray-based European Research Area to Take a Lead in Development and Exploitation (EMERALD), a project that has developed workshops, tutorials, and symposia, as well as quality-control tools (Beisvag et al. 2011). One result has been Minimum Information About a Microarray Experiment (MIAME) in which scientists are asked to provide details about the experiment in a uniform format (Brazma 2009). In addition, scientists are expected to deposit their data into public repositories so that others can attempt to replicate the experiments (Edgar et al. 2002, Brazma et al. 2003). Many publicly available software programs were produced to improve quality control (Beisvag et al. 2011). The MAQC consortium also has developed quality control data (MAQC 2006, 2010).

Some have speculated that the advent of inexpensive next-generation sequencing of cDNAs developed from mRNAs could replace the use of microarrays in the future (Ansorge 2009). In Chapter 7, we learn how to sequence DNA and how to use these sequences.

## General References

- Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.M., Seldman, J.G., and Smith, J.A., et al. (1987). *Current Protocols in Molecular Biology*. Greene Publishing Associates, New York. These two ring-bound volumes offer detailed research protocols. The publisher has a subscription update service so that new pages can be inserted to replace outdated material or expand on existing information.
- Berger, S.L., and Kimmel, A.R., Eds., (1987). *Methods in Enzymology, Volume 152: Guide To Molecular Cloning Techniques*, Academic Press, San Diego, CA.
- Brown, T.A., Ed., (1991). *Essential Molecular Biology. A Practical Approach*, Vols. I and II. IRL Press, Oxford, UK.
- Glover, D.M. Ed., (1985). *DNA Cloning Volume I—A Practical Approach*. IRL Press, Oxford, UK.
- Old, R.W., and Primrose, S.B., (1989). *Principles of Gene Manipulation. An Introduction to Genetic Engineering*, 4th ed. Blackwell Scientific, Oxford, UK.
- O'Reilly, D.R., Miller, L.K., and Luckow, V.A., (1992). *Baculovirus Expression Vectors. A Laboratory Manual*. W. H. Freeman, New York.
- Ptashne, M., (1986). *A Genetic Switch: Gene Control and Phage Lambda*. Blackwell Scientific, Oxford, UK.
- Rio, D.C., Ares, M., Hannon, G.J., and Nilsen, T.W., (2011). *RNA: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Sambrook, J., Fritsch, E.F., and Maniatis, T., (1989). *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Sambrook, J., and Russell, D.W., (2001). *Molecular Cloning: A Laboratory Manual*, 3rd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY. Volumes 1–3 provide detailed protocols for a large number of techniques.
- Williams, J.G., and Patent, R.K., (1988). *Genetic Engineering*. IRL Press, Oxford, UK.

## References Cited

- Andrews, J., Bouffard, G.G., Cheadle, C., Lu, J., Becker, K.G., and Oliver, B., (2000). Gene discovery using computational and microarray analysis of transcription in the *Drosophila melanogaster* testis. *Genome Res.* **10**: 2030–2043.
- Ansorge, W.J., (2009). Next-generation DNA sequencing techniques. *Nature Biotechnol* **25**: 195–203.
- Beisvag, V., Kauffmann, A., Malone, J., Foy, C., Salit, M., and Schimmel, H., et al. (2011). Contributions of the EMERALD project to assessing and improving microarray data quality. *BioTechniques* **50**: 27–31.
- Brazma, A., (2009). Minimum information about a microarray experiment (MIAME)—successes, failures, challenges. *Scientific World Journal* **9**: 420–423.
- Brazma, A., Parkinson, H., Sarkans, U., Shojatalab, M., Vilo, J., and Abeygunawardena, N., et al. (2003). ArrayExpress—a public repository for microarray gene expression data at the EBI. *Nucleic Acids Res.* **31**: 68–71.

- Center for Genomics and Bioinformatics (2012). Indiana University, Bloomington, IN. <<http://cgb.indiana.edu>>.
- Edgar, R., Domrachev, M., and Lash, A.E., (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* **30**: 207–210.
- Gibson, G., (2002). Microarrays in ecology and evolution: a preview. *Mol. Ecol.* **11**: 17–24.
- Gohlmann, H., and Talloen, W., (2009). *Gene Expression Studies Using Affymetrix Microarrays*. CRC Press, Boca Raton, FL.
- Ionnidis, J.P.A., Alison, D.B., Ball, C.A., Coulibay, I., Cui, X., and Culhane, A.C., et al. (2009). Repeatability of published microarray gene expression analyses. *Nat. Genet.* **41**: 149–155.
- Jarvis, D.L., Fleming, J.G.W., Kovacs, G.R., Summers, M.D., and Guarino, L.A., (1990). Use of early baculovirus promoters for continuous expression and efficient processing of foreign gene products in stably transformed lepidopteran cells. *Biotechnology* **8**: 950–955.
- Johnson, R.M., Evans, J.D., Robinson, G.E., and Berenbaum, M.R., (2009). Changes in transcript abundance relating to colony collapse disorder in honey bees (*Apis mellifera*). *Proc. Natl. Acad. Sci. USA* **106**: 14790–14795.
- Kauffmann, A., and Huber, W., (2010). Microarray data quality control improves the detection of differentially expressed genes. *Genomics* **95**: 138–142.
- Kempema, L.A., Cui, X., Holzer, F.M., and Walling, L.L., (2007). *Arabidopsis* transcriptome changes in response to phloem-feeding silverleaf whitefly nymphs. Similarities and distinctions in responses to aphids. *Plant Physiol.* **143**: 849–865.
- Kim, S.Y., Horrigan, S.K., Altenhofen, J.L., Arbieva, Z.H., Hoffman, R., and Westbrook, C.A., (1998). Modifications of bacterial artificial chromosome clones using Cre recombinase: introduction of selectable markers for expression in eukaryotic cells. *Genome Res.* **8**: 404–412.
- Kimmel, A.R., and Berger, S.L., (1987). Preparation of cDNA and the generation of cDNA libraries: overview, pp. 307–316. In: Berger, S.L. and Kimmel, A.R., Eds., *Methods in Enzymology, Volume 152: Guide to Molecular Cloning Techniques*, Academic Press, San Diego, CA.
- Knight, J., (2001). When the chips are down. *Nature* **410**: 860–861.
- Lockhart, D.J., and Winzeler, E.A., (2000). Genomics, gene expression and DNA arrays. *Nature* **405**: 827–836.
- Luckow, V.A., and Summers, M.D., (1988). Trends in the development of baculovirus expression vectors. *Biotechnology* **6**: 47–55.
- Maeda, S., (1989). Expression of foreign genes in insects using baculovirus vectors. *Annu. Rev. Entomol.* **34**: 351–372.
- Maiorella, B., Inlow, D., Shauger, A., and Harano, D., (1988). Large-scale insect cell-culture for recombinant protein production. *Biotechnology* **6**: 1406–1410.
- Maleck, K., Levine, A., Eulgem, T., Morgan, A., Schmid, J., and Lawton, K.A., et al. (2000). The transcriptome of *Arabidopsis* during systemic acquired resistance. *Nat. Genet.* **26**: 403–410.
- MAQC Consortium, (2006). The MicroArray Quality Control (MAQC) project shows that inter- and intraplatform reproducibility of gene expression measurements. *Biotechnology* **24**: 1151–1161.
- MAQC Consortium, (2010). The MicroArray Quality Control (MAQC)-II study of common practices for the development and validation of microarray-based predictive models. *Biotechnology* **28**: 827–838.
- Pouwels, P.H., (1991). Survey of cloning vectors for *Escherichia coli* In: Brown, T.A. (Ed.), *Essential Molecular Biology. A Practical Approach*, Vol. I. IRL Press. Oxford, UK, pp. 179–239.
- Quackenbush, J., (2001). Computational analysis of microarray data. *Nat. Rev. Genet.* **2**: 418–427.
- Ramsay, G., (1998). DNA chips: state-of the art. *Nat. Biotechnol.* **16**: 40–44.
- Reymond, P., (2001). DNA microarrays and plant defense. *Plant Physiol. Biochem.* **39**: 313–321.
- Schenk, P.M., Kazan, K., Wilson, I., Anderson, J.P., Richmond, T., and Somerville, S.C., et al. (2000). Coordinated plant defense responses in *Arabidopsis* revealed by microarray analysis. *Proc. Natl. Acad. Sci. USA* **97**: 11655–11660.

- Shizuya, H., Birren, B., Kim, U.J., Mancino, V., Slepak, T., and Tachiri, Y., et al. (1992). Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. *Proc. Natl. Acad. Sci. USA* **89**: 8794–8797.
- Stewart, D.J., (2000). Making and using DNA microarrays: a short course at Cold Spring Harbor Laboratory. *Genome Res.* **10**: 1–3.
- Taylor, C.F., Field, D., Sansone, S.A., Aerts, J., Apweiler, R., and Ashburner, M., et al. (2008). Promoting coherent minimum reporting guidelines for biological and biomedical investigations: the MIBBI project. *Biotechnology* **26**: 889–896.
- van Oers, M.M., (2011). Opportunities and challenges for the baculovirus expression system. *J. Invertebr. Pathol.* **107** 53–S15.
- Werren, J.H., Richards, S., Desjardins, C.A., Hiehuis, O., Gadau, J., and Colbourne, J.K., et al. (2010). Functional and evolutionary insights from the genomes of three parasitoid *Nasonia* species. *Science* **327**: 343–348.
- White, K.P., Rifkin, S.A., Hurban, P., and Hogness, D.S., (1999). Microarray analysis of *Drosophila* development during metamorphosis. *Science* **286**: 2179–2184.
- Wu, L., Williams, P.M., and Koch, W., (2005). Clinical applications of microarray-based diagnostic tests. *BioTechniques* **39**: 5577–5582.

# DNA Sequencing and the Evolution of the “-Omics”

## Chapter Outline

- 7.1 Overview 252
- 7.2 Introduction 253
- 7.3 The Dideoxy or Chain-Termination (Sanger) Method 255
- 7.4 The Maxam and Gilbert Sequencing Method 261
- 7.5 Shotgun Strategies for Genomes 262
- 7.6 Sequencing DNA by the Polymerase Chain Reaction (PCR) 263
- 7.7 Automated Sanger Sequencers 264
  - 7.7.1 Decreasing Costs of Sanger Sequencing 264
- 7.8 Analyzing DNA Sequence Data 265
- 7.9 DNA-Sequence Data Banks 267
- 7.10 A Brief History of the *Drosophila* Genome Project 268
  - 7.10.1 The Original *Drosophila* Genome Project 269
  - 7.10.2 The Actual *Drosophila* Genome Project 269
  - 7.10.3 *Drosophila* Genome Analysis 270
  - 7.10.4 Surprises in the *Drosophila* Genome 271
- 7.11 Next-Generation Sequencing Methods and Beyond 272
  - 7.11.1 Next-Generation (NextGen or Second-Generation) Sequencing 272
    - 7.11.1.1 Roche (454) GS FLX Sequencer 273
    - 7.11.1.2 Illumina (or Solexa) Genome Analyzer 274
    - 7.11.1.3 Applied Biosystems SOLiD Sequencer 275
    - 7.11.1.4 Standards 275
  - 7.11.2 Third-Generation Sequencing 276
- 7.12 Bioinformatics 277
  - 7.12.1 Gene Ontology 279
- 7.13 Genome Analyses of Other Arthropods 279
  - 7.13.1 Interesting Findings from Completed Genomes 281
  - 7.13.2 What Do You Need to Do to Sequence Your Favorite Insect’s Genome? 285
- 7.14 Transposable Elements (TEs) as Agents of Genome Evolution 287
- 7.15 Transcriptomics 288
  - 7.15.1 Tiling Microarrays 290

7.16 Metagenomics	290
7.17 Proteomics: Another “-Omic”	291
7.18 Functional Genomics	292
7.19 Structural Genomics—Another New Horizon?	293
7.20 Comparative Genomics	294
7.21 Interactomes or Reactomes	295
7.22 The Post-Genomic Era: Systems Genetics	295
General References	297
References Cited	297

## 7.1 Overview

Sequencing, resolving the order of the bases in DNA, is carried out on genomic DNA, cDNA, or mitochondrial DNA. Sequencing is a first step in evaluating regulatory sequences as well as coding and noncoding regions. DNA sequences are used to reconstruct phylogenies, identify taxonomic groups, evaluate the evolution of genomes, and conduct research on population ecology and genetics.

There are two original methods for sequencing DNA (first-generation methods). For both, sequencing involves four procedures: 1) cloning/preparing template DNA, 2) performing the sequencing reactions, 3) electrophoresis of the samples, and 4) compiling and interpreting the data.

The most commonly used first-generation sequencing method, the Sanger or dideoxy chain-terminating protocol, was developed in 1975. The Sanger method involves synthesis of DNA *in vitro* on a single-stranded template by using a primer, a mixture of labeled deoxynucleotide triphosphates (dNTPs), and dideoxynucleotide triphosphates (ddNTPs). The reaction terminates at the position at which the ddNTP, instead of the dNTP, incorporates into the growing DNA chain. Four different reactions are carried out, one for each base. The DNA fragments from the four reactions originally were separated by acrylamide gel electrophoresis with the base sequence being identified by autoradiography of the four banding patterns.

In the other first-generation method, a “chemical” sequencing method developed by Maxam and Gilbert in 1977, single-stranded DNA derived from double-stranded DNA and labeled at the 5' end is subjected to cleavage protocols that selectively make breaks on one side of a specific base. Fragments from the reactions were then separated according to size by acrylamide gel electrophoresis, and the sequences are identified by autoradiography.

DNA sequencing using the first-generation Sanger method now is highly automated, using bases labeled with fluorescent dyes. Scanners read sequences directly and sequence analysis is automated, thereby reducing costs and increasing speed. Automated sequencing methods were used for the very large-scale sequencing required to sequence the entire *Drosophila melanogaster* genome. The information obtained from the *Drosophila* Genome Project has revolutionized both fundamental and applied aspects of insect genetics. Subsequently, several next-generation (NextGen) sequencing methods were developed that are more rapid, less expensive, and provide massive amounts of data. These NextGen methods make it possible to resequence genomes, do *de novo* sequencing of whole genomes, and sequence cDNAs (transcriptomes) inexpensively and rapidly.

As a by-product of genome projects, new scientific disciplines, called “-Omics,” have been developed. Genomics is providing insight into development, speciation, protein interactions, and evolution. Proteomics involves understanding the structure and function of the proteins encoded by the genes. Transcriptomics studies which genes are transcribed during development in specific tissues. Comparative genomics allow scientists to compare whole genomes of multiple species to understand the evolution of genomes. Because genomics and proteomics produce huge amounts of data, biologists need to use computers and other information-management tools, generating a new discipline called bioinformatics. The next few years should see the completion of thousands of genome projects of arthropods, projects that should provide insight into arthropod biology and evolution never before feasible.

## 7.2 Introduction

DNA sequencing is an important component of many molecular genetics projects. Sequencing often is a necessary component of a project, whereas in other cases it is the desired end point and the sequences are used in taxonomic, ecological, or evolutionary studies. Advances in technology have made it feasible to sequence entire genomes. Such sequencing is revolutionizing both basic and applied knowledge of gene structure, gene function, and evolution. In identifying the sequences of promoters, protein-coding sequences, and noncoding regions of DNA in genomic or mitochondrial DNA, it is possible to deduce relationships between organisms and to reconstruct their evolutionary history.

The development of extensive computerized databases of DNA and protein sequences allows hypotheses to be constructed and tested regarding the structure and function of proteins and their secondary structures. All of these

opportunities became possible only after DNA sequencing methods were developed in the late 1970s.

Two basic DNA sequencing methods were developed about the same time: the chemical, or Maxam-Gilbert method (1977, 1980), and the chain-termination method of [Sanger et al. \(1977\)](#). Both used the same basic approaches: 1) cloning or preparing the DNA templates, 2) performing the sequencing reactions on the DNA templates, 3) electrophoresis of the samples, and 4) compiling and interpreting the data.

DNA to be sequenced can be genomic DNA, mitochondrial DNA, or cDNA. Because cDNA lacks introns and regulatory elements, sequencing of cDNA provides less information than genomic DNA. Sequencing only cDNA probably would miss some genes that are expressed at very low levels or in a tissue- or time-dependent manner.

Effective computer tools are needed to discover the sequences that actually code for a gene, because up to 90% of genomic sequences can be noncoding DNA (that do not code for proteins). Some noncoding DNA sequences are associated with centromeres or telomeres, and others have no known function. Different computer programs have been developed to search DNA sequence data and identify possible regulatory sequences, potential start or stop codons, open reading frames (ORFs), and sequences that may indicate the location of intervening sequences or introns. Unfortunately, no computer program is 100% accurate in identifying genes ([Bork 2000](#)).

The most reliable way to find genes currently is to identify them because they are similar to known proteins from the same or other organisms or similar to cDNAs from the same or a closely related organism ([Stormo 2000](#)). However, many genes have no significant similarity with other known sequences (**orphan genes**). Furthermore, just because a DNA sequence is similar to the sequence of a known gene in *D. melanogaster* does not necessarily indicate that this gene functions in the same manner.

The length of DNA that can be sequenced by a single Sanger reaction varies from 200 to  $\approx$ 1000 bases, depending upon the method used ([Sambrook and Russell 2001](#)). Vectors can contain DNA inserts ranging in size from 100 to 1,000,000 bp. For example, yeast artificial chromosomes (YACs) can contain inserts up to 1 million base pairs, and cosmids can contain inserts of 30,000–45,000 bp. Thus, cloned DNA typically is converted into smaller segments or **sub-clones** that are then inserted into vectors that are specialized for sequencing, such as M13 or plasmid sequencing vectors.

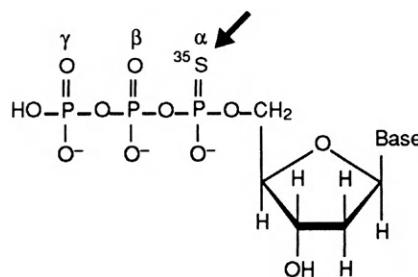
### 7.3 The Dideoxy or Chain-Termination (Sanger) Method

In brief, the **dideoxy** or **chain-termination** method developed by Sanger involves *de novo* synthesis of a series of labeled DNA fragments from a single-stranded (ss) DNA template. The ss DNA segment to be sequenced serves as the template for the synthesis, by base pairing, of a new labeled strand of DNA. Labeling initially was achieved by labeling with  $^{32}\text{P}$  or  $^{35}\text{S}$ .

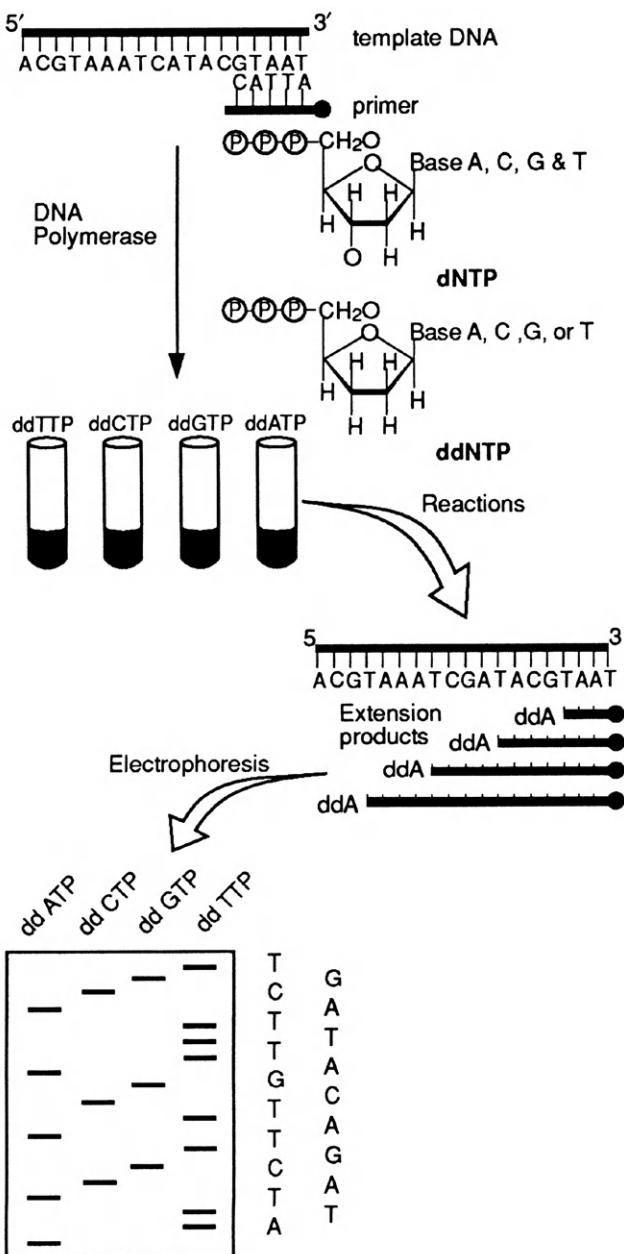
Because DNA synthesis is used, the sequencing reaction requires a DNA polymerase, labeled deoxyribonucleotides (dNTPs), a primer, and dideoxyribonucleotides (ddNTPs) (Figure 7.1). Several different DNA polymerases could be used (Klenow fragment, Sequenase 2, or thermophilic DNA polymerases such as *Taq*), with different protocols (Sambrook and Russell 2001). The dNTPs can be labeled either with  $^{32}\text{P}$  or  $^{35}\text{S}$ , but  $^{35}\text{S}$  labeling produces sharper bands and improves the resolution of the autoradiogram. The structure of a dNTP that has been labeled with  $^{35}\text{S}$  is shown in Figure 7.1.

DNA sequencing kits that contain the necessary enzymes and components can be purchased. Each has specific protocols provided and is ideal for first-time sequencers, although they are too expensive for large-scale projects (Sambrook and Russell 2001).

The dideoxy or chain-terminating reaction begins by adding a short oligonucleotide primer that is complementary to a region of DNA adjacent to the DNA segment to be sequenced (Table 7.1, Figure 7.2). The primer is normally 15–30 nucleotides and is annealed to the template in a preincubation step. Four separate reactions are set up to determine the position, respectively, of the A, T, G, and C bases in the template DNA. Each reaction requires a mixture of DNA polymerase, primers, dNTPs, ddNTPs, and the template DNA. The ddNTPs are



**Figure 7.1** Structure of  $\alpha^{35}\text{S}$ -deoxynucleoside triphosphate. Labeling with  $^{35}\text{S}$  results in sharp bands, which increases the resolution of the sequencing gels.



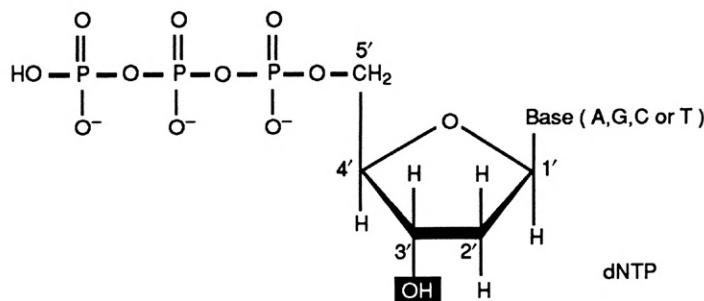
**Figure 7.2** Diagram of the steps involved in the Sanger dideoxy chain-termination method of DNA sequencing. Four reactions are carried out. (Modified from Hunkapiller et al. 1991.)

**Table 7.1: An Example of a Dideoxy (Sanger) Sequencing Protocol Using Modified Bacteriophage T7 DNA Polymerase (Sequenase™).**

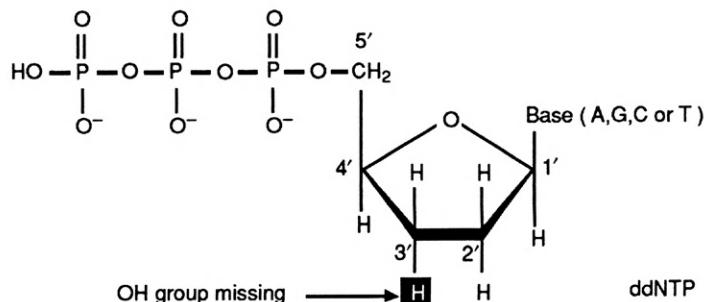
Sequencing reagents								
Annealing buffer (5× concentrate) 200 mM Tris-HCl (pH 7.5) 100 mM MgCl <sub>2</sub> 250 mM NaCl Dithiothreitol (DDT), 0.1 M Labeling nucleotide mixture (for use with radiolabeled dATP) 1.5 mM dGTP, 1.5 mM dCTP, and 1.5 mM dTTP Termination nucleotide mixtures (one for each dideoxynucleotide) Each mixture contains 80 mM dGTP, 80 mM dATP, 80 mM dTTP, 80 mM dCTP, and 50 mM NaCl. In addition, the “G” mixture contains 8 mM dideoxy-dGTP; the “A” mix, 8 mM ddATP; the “T,” 8 mM ddTTP; and the “C,” 8 mM ddCTP. <i>Stop solution:</i> 95% formamide, 20 mM EDTA, 0.05% bromophenol blue, and 0.05% xylene cyanol FF								
Procedures								
<b>Annealing template and primer</b> 1. For each template, a single annealing (and subsequent labeling) reaction is used. Combine the following: <table> <tr> <td>Primer</td> <td>0.5 pmol</td> </tr> <tr> <td>DNA</td> <td>0.5–1.0 pmol (1 mg of M13)</td> </tr> <tr> <td>Annealing buffer</td> <td>2 ml</td> </tr> <tr> <td>Water to bring the total volume to 10 ml</td> <td></td> </tr> </table> Warm the capped tube to 65°C for 2 minutes, allow the mixture to cool slowly to room temperature over a period of ≈30 minutes.	Primer	0.5 pmol	DNA	0.5–1.0 pmol (1 mg of M13)	Annealing buffer	2 ml	Water to bring the total volume to 10 ml	
Primer	0.5 pmol							
DNA	0.5–1.0 pmol (1 mg of M13)							
Annealing buffer	2 ml							
Water to bring the total volume to 10 ml								
Labeling reaction								
<b>2. To the annealed template-primer add the following:</b> <table> <tr> <td>DTT (0.1 M)</td> <td>1 ml</td> </tr> <tr> <td>Labeling nucleotide mix</td> <td>2 ml</td> </tr> <tr> <td>α-<sup>35</sup>S or α-<sup>32</sup>P dATP</td> <td>5 μCi (typically 0.5 ml)</td> </tr> <tr> <td>Sequenase™</td> <td>3 units</td> </tr> </table> Total volume should be ≈15 ml; mix thoroughly and incubate for 2–5 minutes at room temperature.	DTT (0.1 M)	1 ml	Labeling nucleotide mix	2 ml	α- <sup>35</sup> S or α- <sup>32</sup> P dATP	5 μCi (typically 0.5 ml)	Sequenase™	3 units
DTT (0.1 M)	1 ml							
Labeling nucleotide mix	2 ml							
α- <sup>35</sup> S or α- <sup>32</sup> P dATP	5 μCi (typically 0.5 ml)							
Sequenase™	3 units							
Termination reactions								
<b>3. Label four tubes “G,” “A,” “T,” and “C.” Fill each with 2.5 ml of the appropriate dideoxy termination mixture.</b> When the labeling reaction is complete, transfer 3.5 ml of it to the tube prewarmed to 37°C, labeled “G.” Do the same for each of the other three tubes (A, T, and C). After 2–5 minutes at 37°C, add 4 ml of Stop Solution to each termination reaction, mix, and store on ice. To load the gel, heat the samples to 75–80°C for 2 minutes or more and load 2–3 ml in each lane.								
4. Polyacrylamide gel electrophoresis								
5. Analysis of sequences								

(Modified from Tabor and Richardson 1987.)

## Normal deoxynucleoside triphosphate ( i.e. 2'-deoxynucleotide )



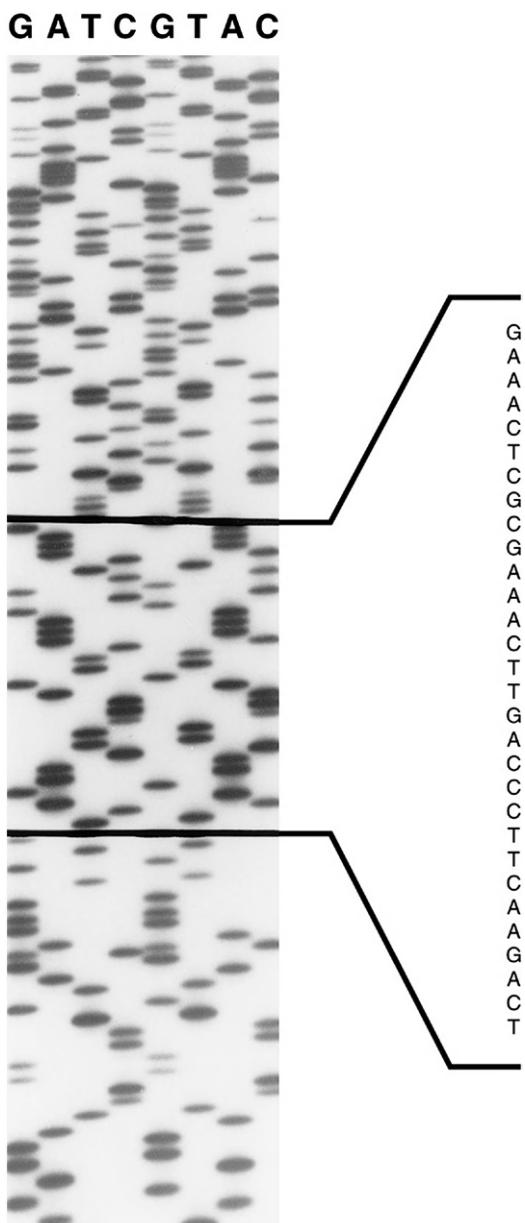
## Dideoxynucleoside triphosphate ( i.e. 2', 3' dideoxynucleotide )



**Figure 7.3** Dideoxynucleoside triphosphates (ddNTPs) (bottom) act as chain terminators because they lack a 3'-OH group found on normal deoxynucleotides (dNTPs) (top).

derivatives of dNTPs that do not contain a hydroxyl group at the 3' position of the deoxyribose ring (Figure 7.3).

When ddNTPs are incorporated into the growing DNA chains instead of dNTPs, that particular DNA molecule is terminated at that point. All four dNTPs are present in each tube with each ddNTP, but the ratio is adjusted so that ddNTPs are less abundant than dNTPs. This makes the incorporation of a ddNTP a random event. The newly synthesized DNA molecules in a specific reaction tube therefore are a *mixture* of DNA fragments of different lengths, each with a fixed *starting point* (determined by the primer) but with variable ending points. Thus, for example, in the reaction in which the chain is terminated when thymines (T) are incorporated, a ddNTP does not always get incorporated into the first site where a T occurs. Nor does a ddNTP necessarily get incorporated into the chain where the second T occurs. However, over the length of the DNA being sequenced, each site where a T is incorporated will have ddNTPs incorporated, so that a series of DNA molecules of different length are produced (Figure 7.4). Thus, *populations* of synthesized molecules are produced in which



**Figure 7.4** Autoradiograph of a sequencing gel obtained by the dideoxy chain-termination method. Four different reactions are carried out in which DNA synthesis of a complementary DNA chain is terminated by incorporating a radiolabeled dideoxy base (ddATP, ddCTP, ddGTP, or ddTTP) rather than a deoxy base (dATP, dCTP, dGTP, dTTP). The resulting fragments of synthesized DNA are visualized by acrylamide gel electrophoresis and autoradiography. The sequence is read by determining which lane contains each succeeding DNA segment, as determined by a band. Thus, reading *from the bottom*, the shortest fragment has a band in the T lane (just above the line). The next band is a C, then the next is an A, and so on. Bands at the very top and bottom of the gel are not read. This sequence is a portion of the *mariner* transposable element cloned from the predatory mite *Metaseiulus occidentalis* (Jeyaprakash and Hoy 1995).

the chain is terminated at each site where Ts occur. These DNA molecule populations can be visualized by gel electrophoresis.

The base sequence is visualized by running the radiolabeled DNA fragments from the four reactions on an acrylamide gel in four adjacent lanes. Each reaction tube will produce a series of bands, with each band representing a population of molecules at which the DNA molecule was terminated by incorporating a ddNTP. The banding pattern in the four lanes was visualized on an X-ray film ([Figure 7.4](#)). The two strands should be sequenced independently to reduce errors generated by artifacts in the sequence reactions or inadequate resolution of regions of the sequence on the gel.

There are several different protocols for sequencing DNA by the dideoxy chain-termination method ([Ambrose and Pless 1987](#), [Barnes 1987](#), [Mierendorf and Pfeffer 1987](#), [Tabor and Richardson 1987](#), [Howe and Ward 1989](#), [Sambrook et al. 1989](#)). Variables include such factors as whether the DNA is double- or single-stranded and whether the DNA is sequenced directly from double-stranded plasmid DNA or after subcloning into a single-stranded phage such as M13. Today, the ddNTPs are labeled by one of four different fluorescent dyes. The results are shown as a chromatogram with a series of peaks of A, T, C, and G labeled with different colors.

Both Sanger and Maxam and Gilbert sequencing methods generate sets of DNA molecules that share a common end that is determined by the primers but that vary in their length at the other end. Both methods also originally used radioactive labeling to visualize the results. Once the DNA was electrophoresed, the gel was dried onto paper and exposed to X-ray film. The autoradiogram produced displayed a ladder of bands representing the migration position of the different radiolabeled DNA segments. For example, the sequence of 20 nucleotides can be read to the right of the four lanes in [Figure 7.4](#).

Both the Sanger and Maxam and Gilbert sequencing methods originally used very thin polyacrylamide gels to discriminate between nucleic-acid molecules that may differ in length by only one nucleotide. A sequencing gel is a high-resolution gel containing 6–20% polyacrylamide and 7 M urea. The DNA to be analyzed is denatured before electrophoresis by heating it to 80°C in a buffer containing formamide, which lowers the melting temperature of ds DNA molecules. The urea minimizes DNA secondary structure, which could affect mobility of the DNA through the gel. The gel is run using sufficient power so that it is heated to about 50°C, which also minimizes DNA secondary structure. The length of the DNA molecule determines its rate of migration. The shorter

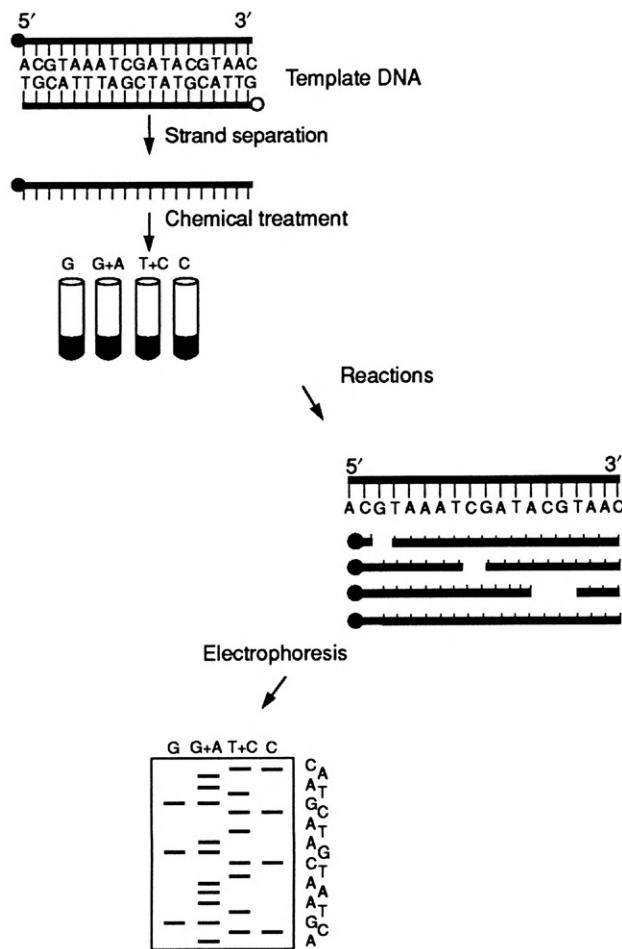
fragments migrate more quickly than the longer, thus ending up nearer the bottom of the gel (Davies 1982, Sealey and Southern 1982, Howe and Ward 1989).

If the template DNA is a subclone that was inserted into the multiple-cloning site of a vector, then the primer almost always is an oligonucleotide complementary to sequences flanking the multiple-cloning site. This allows any fragment cloned into the multiple-cloning sites to be sequenced using the same primer. Most vectors have the *lacZ* sequences flanking the multiple cloning sites. Thus, an oligonucleotide primer directed to this sequence, which is 16 or 17 nucleotides long, is commonly used and called the **M13 universal primer**. This primer is designed to be complementary to the strand of DNA packaged into M13 or into any plasmid vector containing an M13 origin of replication. Methods also have been developed for direct sequencing from denatured plasmid DNA (Mierendorf and Pfeffer 1987), eliminating the need to isolate or subclone DNA fragments.

## 7.4 The Maxam and Gilbert Sequencing Method

**Maxam and Gilbert DNA sequencing** is called the chemical-cleavage method (Maxam and Gilbert 1977, 1980). It uses chemical reagents to generate base-specific cleavages of the DNA to be sequenced (Figure 7.5). It was less-often used, in part because the chemicals are toxic and the methods are labor-intensive. The primary advantage of this method is that DNA sequences are obtained from the original DNA molecule and not from a synthesized copy. Thus, one can analyze DNA modifications such as methylation and study DNA secondary structure and the interaction of proteins with DNA.

To start, one needs pure DNA that has been cut by restriction endonucleases to generate DNA of specific length and with known sequences at one end. Each DNA fragment then can be radioactively labeled at one end with a  $^{32}\text{P}$ -phosphate group in sufficient quantity that at least four different chemical reactions can be carried out. Next, specific bases in the DNA fragment are altered in four separate chemical reactions. For example, guanine (G) is methylated by dimethylsulfate. Each reaction is carried out in a manner that limits the reaction so that, for example, only one G is modified per several hundred Gs. The altered G is then removed, in a subsequent step, by cleavage at the modification points with piperidine. The result is a set of end-labeled fragments of different lengths that will show up as a ladder of bands on the gel because the reaction was limited and not all the Gs were altered in the reaction. Four different reaction samples (one for A + G, G, T + C, and C) are then run



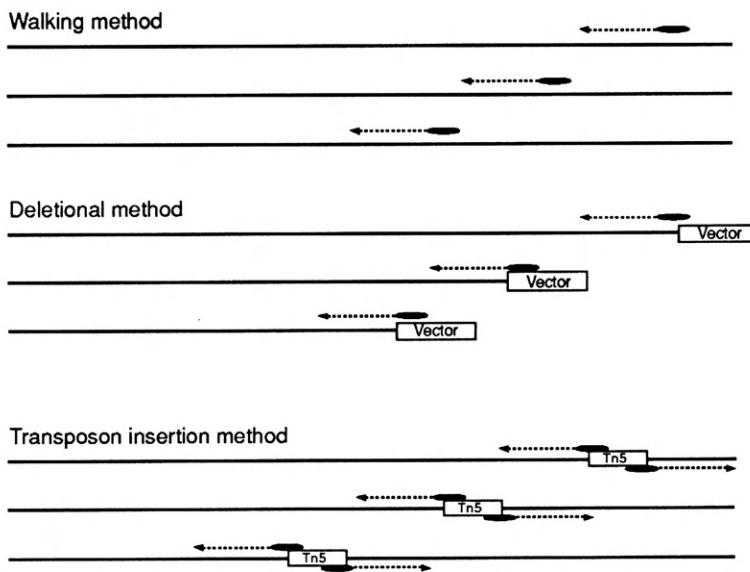
**Figure 7.5** Diagram of the steps involved in the Maxam and Gilbert chemical cleavage method of DNA sequencing (Modified from Hunkapiller et al. 1991.)

side by side on a sequencing gel, and the results can be visualized by autoradiography (Figure 7.5).

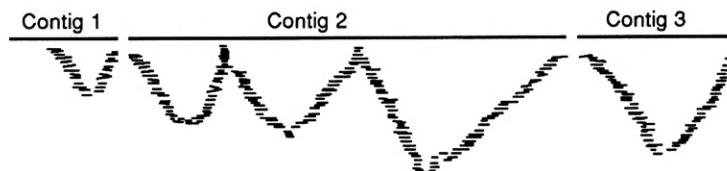
## 7.5 Shotgun Strategies for Genomes

In “shotgun” sequencing strategies, the DNA is digested with a restriction endonuclease and subfragments are cloned and sequenced (Figure 7.6). The nucleotide sequence of the various inserts is obtained, and a computer is used to determine how the fragments (=contigs) overlap and to establish the entire sequence of the original DNA used to generate the subclones. Disadvantages to the shotgun method are that it may underrepresent some fragments,

### A) Directed strategy



### B) Random strategy



**Figure 7.6** Strategies for DNA sequencing of long DNA segments involve either directed (A) or random (B) methods. Directed methods include walking, deletional sequencing, or transposon insertion. Analyzing sequences from many contigs requires large-scale computer alignments.

sequencing must be redundant to ensure that the entire sequence has been included in one or more subclones, and there is no way to identify specific fragments. But, as you will discover soon, the shotgun approach has been adopted widely due to innovations that allow assembly of the sequences into a contiguous scaffold.

## 7.6 Sequencing DNA by the Polymerase Chain Reaction (PCR)

The **polymerase chain reaction** (PCR) is a procedure by which a specific segment of DNA can be amplified by 1 million-fold, or more, using a DNA polymerase (see Chapter 8). DNA fragments can be amplified by the PCR directly from

genomic or cloned DNA, eliminating the need to prepare large amounts of DNA from tissues for sequencing and subcloning steps.

Conventional PCR requires primers of known sequence that flank the region to be amplified. Several techniques have been developed for sequencing ds DNA produced by the PCR. Cycle sequencing is advantageous because it requires only very small amounts of template DNA, and it can work with ds templates as well as ss templates, eliminating the need to subclone the DNA (Murray 1989, Bevan et al. 1992). However, problems that need to be considered include confirming that you have amplified the correct fragment, removing all residual PCR primers and unincorporated nucleotides before sequencing, and providing good quality primers for sequencing. Recombinant DNA sequences can be generated by the PCR, so care must be taken especially when studying multigene families (Bradley and Hillis 1997).

## 7.7 Automated Sanger Sequencers

The invention of automated, fluorescent Sanger sequencers made large-scale genome projects feasible. Instruments can automate nearly every step of the large-scale sequencing process. Integrated machines can isolate DNA, clone or amplify DNA, prepare sequencing reactions, purify DNA, and separate and detect DNA fragments containing fluorescent labels to obtain the DNA sequence (Meldrum 2000a). Automated sequencers first used horizontal or vertical slab gels, but more recent commercial systems use capillary sequencers (Meldrum 2000b). A commonly used sequencer, the ABI PRISM 377 DNA Sequencer, uses multicolor fluorescence labeling and a four-dye, one-lane detection system. Two hundred bases per sample per hour can be analyzed, and 18, 36, 64 or 96 samples can be analyzed simultaneously.

Large-scale sequencing facilities used a random shotgun phase combined with a directed finishing phase to complete analysis of the difficult regions of the early genomes sequenced. Others used a whole-genome shotgun approach in which random fragments of total genomic DNA were subcloned, and high-throughput sequencing was used to generate sequences that provide at least 10-fold coverage of the genome. These sequences were ordered and put into a linear sequence with the aid of high-speed computers (Meldrum 2000a).

### 7.7.1 Decreasing Costs of Sanger Sequencing

Sanger sequencing is now automated (Men et al. 2008). DNA is isolated, fragmented, and cloned into a vector for DNA amplification in bacterial cells.

Millions of individual bacterial colonies are produced and individually placed in multiwall plates by robotics to isolated individual DNA clones. This DNA then goes through a sequencing reaction and the sequenced DNA undergoes capillary electrophoresis where labeled nucleotides are collected and scanned by a laser to produce sequencing reads. The raw data are then converted into computer files showing the sequence and the quality of each base. The final data are stored and released to public databases such as GenBank.

As a result of the industrialization of DNA sequencing by the Sanger method, the cost decreased to  $\approx$ US\$0.20–0.30 per base to sequence the *Drosophila* genome when the accuracy was held to less than one error in 10,000 bases. During 2000, approximately one complete bacterial genome was obtained each month. The *Drosophila* Genome Project was completed in fall 2000 (Adams et al. 2000), and on June 26, 2000, a working draft of the human genome was completed at a cost of  $\approx$ US\$1 billion (Bentley 2006). The year 2000 was called “The Year of the Genome,” but was only the start of the genomics revolution. In 2006, the cost to sequence the human genome using the Sanger method would take  $\approx$ 30 instruments for 1 year to complete and would cost \$US10 million (Bentley 2006). As discussed in Section 7.11, newer sequencing methods have reduced the costs and time to completion, and revolutionized genome sequencing, although they also made analysis more complex and time-consuming.

## 7.8 Analyzing DNA Sequence Data

Even small-scale DNA sequencing projects generate substantial amounts of data and require computer assistance for their analysis (Reese et al. 2000, Stein 2001, Mount 2004, Hodgman et al. 2009, Pevsner 2009). Software packages are available for laboratory computer systems and, depending on the size of the computer, can analyze the sequences in greater or lesser detail. DNA sequences obtained from automated sequencing machines are provided online or on computer disk.

Computer programs can compare reads from several sequencing runs, search for and identify overlaps, compare results from sequencing the complementary strands of the DNA, and identify possible errors. Once the sequences have been entered into the computer, the next step is to analyze the data.

In a shotgun genome-sequencing project, the DNA is broken into fragments that are cloned and sequenced. The relationships between the cloned fragments are determined by comparing their sequences. DNA segments related to one another by a partial overlap are called **contigs**. If a sequence overlaps with another, then the two contigs can be joined. The process of comparing

sequences and aligning them is continued until it is possible to produce a continuous DNA sequence for the DNA of interest.

A variety of questions can be asked after the sequences are obtained. Sequences can be searched for all known restriction endonuclease target sites, and the computer can generate a comprehensive and precise restriction map. The sequences can be searched for interesting structures such as **tandem repeats** and **inverted repeats**, which would indicate the insertion of transposable elements (TEs). The sequences can predict which proteins are coded for based on the sequences in the possible open reading frame (ORF) on each of the two strands. An ORF is a segment of DNA that does not include a termination codon and that may indicate that a polypeptide-coding region exists. Both strands must be interpreted because it is not known in advance which is the coding and which is the noncoding strand.

The DNA sequence, or the deduced polypeptide sequence, may be compared with sequences in data banks. Often, because of the degeneracy of the DNA code, similarities are found when two polypeptide sequences are compared; these sequence similarities might not have been apparent if the comparison had been carried out only at the DNA-sequence level. DNA-sequence similarities may be present because of **convergent evolution** or through homology. Convergent evolution implies that the two sequences did not have a common ancestral sequence but that selection for a particular function in two different lineages has converged on a particular structure or related structures.

The term *homology* has become controversial because it has multiple definitions. Traditionally, similar structures in different organisms have been called homologous if the organisms have descended from a common ancestor. However, some molecular biologists have used the term “per cent homology” when they mean there is *similarity* in DNA sequence, which may not be due to descent from a common ancestor.

A search of the DNA or protein sequence banks for **similarities** with any newly discovered sequence might turn up amazing degrees of similarity. For example, **homeotic** genes, genes that direct cells in different segments to develop in particular patterns, have been cloned from *D. melanogaster*. The **homeobox**, a segment of  $\approx 180$  bp, is characteristic of the homeotic class, and probes using the homeo-box sequence have been used to isolate previously unknown homeotic genes in other insects. Sequences homologous to the homeobox have been isolated from mice and humans, indicating that similar genetic mechanisms control aspects of development in higher organisms. The high degree of conservation between the homeobox sequences of *Drosophila*, frogs, mice, and humans

indicates that these sequences have been conserved for >500 million years, when invertebrates and vertebrates diverged.

Sequenced genomes often are searched for transposable elements. TEs are ancient and are present in nearly all organisms, are very diverse, and may constitute up to 80% of the genomic DNA. There are thousands of different TE families. Identification and annotation of TEs found in genomes is a challenge due to this diversity. [Wicker et al. \(2007\)](#) proposed a unified hierarchical classification system based on the transposition method, sequence similarities, and structural relationships. [Kapitonov and Jurka \(2008\)](#) also produced a database of eukaryotic repetitive and TEs and called it Repbase. Repbase categorizes eukaryotic TEs as either retrotransposons or DNA transposons and places them into five major classes: long terminal repeat (LTR-retrotransposons, non-LTR retrotransposons, cut-and-paste DNA transposons, rolling-circle DNA transposons (*Helitrons*), and self-synthesizing DNA transposons (*Polintons*) ([Kapitonov and Jurka 2008](#)). Each class is composed of superfamilies and each superfamily consists of numerous families.

## 7.9 DNA-Sequence Data Banks

DNA-sequence data banks are expanding rapidly and are important resources for the research community. There are three major DNA sequence databases: the DNA Data Bank of Japan (DDBJ), the European Molecular Biology Laboratory Nucleotide Sequence Data Library (EMBL), and the GenBank Genetic Sequence Data Bank (GenBank). Subsets of these databases have been organized. For example, there is a database of metazoan mitochondrial DNA sequences ([Lanave et al. 2000](#)), a eukaryotic promoter database ([Perier et al. 2000](#)), database of restriction enzymes and methylases ([Roberts and Macelis 2000](#)), a database for intron sequence and evolution ([Schisler and Palmer 2000](#)), a database for homeo domains ([Banerjee-Basu et al. 2000](#)), and many others ([Wheeler et al. 2000](#)). The InSatDb is microsatellite database of fully sequenced insect genomes ([Archak et al. 2007](#)). The journal *Nucleic Acids Research* publishes an annual list of hundreds of online databases covering significant aspects of molecular biology, many of which are interconnected or hyperlinked ([Galperin and Fernandez-Suarez 2011](#)). VectorBase is a data resource for invertebrate vector genomics and includes the genomes of mosquitoes, ticks, body lice, and other arthropod vectors of disease agents ([Lawson et al. 2009](#)). SilkDB is a repository for silk moth genetic and genomic data ([Wang et al. 2005](#)). There is a Hymenoptera genome database ([Munoz-Torres et al. 2011](#)), a butterfly database ([Papanicolaou et al. 2008](#)), an aphid database ([Legeal et al. 2010](#)), and a locust database ([Zongyuan et al. 2006](#)). No doubt others will be added as more arthropod genome sequences are obtained.

The data banks can be searched over the web, but some caution is required because errors apparently are common in both the data banks and scientific journals. Errors, particularly in noncoding regions, may arise from sequencing or clerical errors. Submission of a sequence to a database in a machine-readable form is a prerequisite for publishing in many journals (Cinkosky et al. 1991). Methods of DNA sequence analysis using computer programs such as BLAST and PAUP are described briefly in Chapter 12 on molecular systematics and are described by Mount (2004).

There are so many databases available on the web, each with useful information on various aspects of molecular biology, that there are now databases of databases (Baxevanis 2001).

## 7.10 A Brief History of the *Drosophila* Genome Project

The Human Genome Project was the reason the *Drosophila* Genome Project was initiated. The Human Genome Project was one of the largest initiatives in the history of biology and was one of the most controversial. To prepare for this enormous undertaking, the genomes of the yeast *Saccharomyces cerevisiae*, the nematode *Caenorhabditis elegans*, the fruit fly *D. melanogaster*, and the laboratory mouse *Mus musculus* were targeted for sequencing to develop and improve methods before undertaking the larger human genome. Even these proposed subprojects generated considerable controversy.

Controversy arose as to whether it was appropriate to spend resources to sequence the *Drosophila* genome. Knowledge of the structure and function of the *D. melanogaster* genome already was far greater than that for any other multicellular organism (Kafatos et al. 1991), and some believed a *Drosophila Genome Project* was unnecessary. The *D. melanogaster* genome is  $\approx$ 180 Mb of DNA, a third of which is heterochromatin. The 120 Mb of euchromatin is on the two large autosomes and the X; the fourth chromosome is mostly heterochromatin, with only  $\approx$ 1 Mb of euchromatin. By 1991,  $\approx$ 3800 different genes, approximately a fourth of the total, already had been mapped by recombination studies. Many had been associated cytogenetically with one of the 5000 bands of the polytene-gland chromosomes. Approximately 3000 “transcription units” had been placed on the cytogenetic map by localizing the DNA on specific polytene chromosomes by *in situ* hybridization. Nearly 10% of the total genes, 1300 genes, in *D. melanogaster* already had been cloned and sequenced by individual laboratories (Rubin and Lewis 2000). However, after the project was concluded it was discovered that approximately one-third of the genes in *Drosophila* do not have obvious phenotypes when mutated, making the sequencing project a useful gene-discovery method.

### 7.10.1 The Original *Drosophila* Genome Project

The original *Drosophila* Genome Project had the following aims:

1. Develop a high-resolution **physical map** that would serve as a basis for DNA sequencing and detailed functional studies. A physical map is a series of overlapping clones for which information is available on the sequences at their ends and knowledge of their physical location on the chromosomes. The physical map would be integrated in a database with cross-references to the genetic information already available for *D. melanogaster*.
2. Conduct feasibility studies for large-scale DNA sequencing projects, especially for regions containing DNA of great biological interest. Large-scale studies were defined as those that attempted to determine three megabases (Mbp) of contiguous DNA sequence within 3 years using the Sanger method.
3. Develop new bioinformatic techniques to identify coding sequences in genomic DNA and to obtain high-quality cDNA libraries that were representative of the complete coding information of the genomic DNA (Merriam et al. 1991).

### 7.10.2 The Actual *Drosophila* Genome Project

The *Drosophila* Genome Project actually was completed much more quickly and by a different strategy than originally planned (Adams et al. 2000, Pennisi 2000a). *Drosophila melanogaster* became only the second multicellular organism (after the nematode *C. elegans*) to have its entire genome sequenced.

The initial *Drosophila* sequencing effort was initiated in 1990 and was only partially completed when Venter et al. (1996) proposed using a “shotgun strategy.” This was a novel approach to sequencing such a large genome and involved breaking the entire genome into small pieces, sequencing them with an array of very fast and expensive new Sanger sequencing machines, and then using powerful supercomputers to assemble the sequenced fragments into the correct order. A collaboration was undertaken by a company founded by Craig Venter (Celera), the Berkeley *Drosophila* Genome Project, and its European counterpart to guide the work and interpret the data.

Shotgun cloning had never been attempted previously with such a complex genome. The complexity is due to the presence of hundreds to thousands of repeated sequences that are scattered throughout the genome and cause problems in assembly of the sequence data. The solution was to obtain sequences from *both ends* of fragments (paired ends) that were  $\approx$ 2, 10, and 150 kb. These oriented bits of sequence were assembled into increasingly dense and inter-linked scaffolds that generated long continuous stretches of DNA sequence

with few gaps (Adams et al. 2000). The success of the shotgun approach with *Drosophila* encouraged a similar approach with the human genome (Venter et al. 2001). The actual sequencing of the *Drosophila* genome began in May 1999 at Celera, and by late fall 1999 the sequencing was completed and the computers had assembled the sequences! The sequences were published in *Science* in March 2000 (Adams et al. 2000). The publication of the *Drosophila* genome represented a major milestone for insect molecular genetics (Hawley and Walker 2000). The entire *Drosophila* sequence is available in GenBank and at FlyBase on the web (FlyBase Consortium 1997). FlyBase is a database of genetic and molecular data and includes genes, alleles, phenotypes, aberrations, transposons, clones, stock lists, locations of *Drosophila* workers, and bibliographic references (Misra et al. 2000).

The genomes of *Escherichia coli*, *Saccharomyces*, *C. elegans*, *D. melanogaster*, and *Homo sapiens* had been completed ahead of schedule and less expensively than expected by the end of 2000. On February 15, 2001, the sequences of the human genome were published (International Human Genome Sequencing Consortium 2001, Venter et al. 2001). Since these model organisms had their genomes sequenced (at great expense), inexpensive and rapid sequencing methods were developed and used to sequence other genomes (including those of other arthropods). These next-generation sequencing methods revolutionized genome sequencing, as described in Section 7.11.

### 7.10.3 Drosophila Genome Analysis

Obtaining the DNA sequence is only a first step (Stein 2001). Analyses of the sequence data must be conducted and that is becoming increasingly more difficult and time-consuming. An early analysis involved “annotating” the *Drosophila* genome; as many genes as possible were identified and the function of the proteins/gene products was predicted (Adams et al. 2000, Pennisi 2000b, Reese et al. 2000). The Genome Annotation Assessment Project (GASP) assessed the accuracy of the annotation. GASP focused on analysis of a well-known region of the *Drosophila* genome, and the 12 groups carrying out the analysis did best in identifying the coding regions, with a success rate averaging >95%. The correct intron/exon structures were predicted for >40% of the genes. Almost half the genes in the region were recognized and assigned functions by homology with known genes. However, less than one-third of the promoters in the region were found by the GASP group (Reese et al. 2000). Subsequent annotations and evaluations were required to refine and improve on these initial annotations.

Kaminker et al. (2002) subsequently analyzed the *Drosophila* genome and identified 85 known and eight novel families of TEs, with copy numbers varying from 1 to 146. Most of the TEs were partial elements, and many were localized in the regions near the centromeres. Most TEs were found outside genes, and some were found nested within other elements of the same or different classes. Depra et al. (2009) found that *hobo* (class II) transposons and *hobo*-related elements are expressed as early developmental genes in *Drosophila* embryos and in the nervous system because regulatory sequences similar to those in developmental genes exist in *hobo* sequences. Thus, TEs may provide potentially beneficial developmental and evolutionary changes by acting as regulatory elements. As noted in Chapter 4, the telomeres of *Drosophila* chromosomes are formed by non-long terminal repeat retrotransposons. These retrotransposons are named *Het-A*, *TART*, and *TAHRE*, and they are found in many species of *Drosophila*. These elements transpose onto chromosome ends to form long arrays that extend the chromosome and compensate for loss (George et al. 2010).

#### 7.10.4 Surprises in the *Drosophila* Genome

Several unexpected results were found. First, early analyses of the *Drosophila* genome suggested there are  $\approx$ 13,600 genes, slightly fewer than the number found in the nematode *C. elegans* (Adams et al. 2000). *Drosophila melanogaster* was expected to have  $\approx$ 30,000 genes. *Drosophila* probably has  $>$ 13,600 genes because it has a relatively large number of overlapping genes (Ashburner 2000), and many protein-coding genes undergo alternative splicing.

Immediately after obtaining the *D. melanogaster* genome sequences, a comparison was made to the genomes of *C. elegans* and *S. cerevisiae* in the context of cellular, developmental, and evolutionary processes (Rubin et al. 2000). The comparisons indicated *Drosophila* had many undiscovered genes, despite having been the subject of extensive genetic analyses. Annotation of the *Drosophila* genome also indicated this insect is surprisingly relevant to the study of genes and metabolic pathways involved in tumor formation and development in humans (Potter et al. 2000, Chintapalli et al. 2007). Many of the well-studied signal pathways in tumor development in humans are conserved between flies and humans, and at least 76 *Drosophila* genes are homologs to mammalian cancer genes and are under intensive study. Furthermore, 178 (62%) of 287 known human-disease genes seem to be conserved in *Drosophila*, including genes causing neurological (Alzheimer’s disease, Huntington’s disease, Duchenne muscular dystrophy, and juvenile-onset Parkinson’s disease), renal, cardiovascular, metabolic, and immune diseases and malformation syndromes (Fortini et al. 2000).

Liu et al. (2012) found that the microRNA miR-34 is important in aging and neurodegeneration in *Drosophila*. This microRNA is highly conserved, with orthologs in *C. elegans*, mouse, and humans. miR-34 shortened life span and resulted in late-onset brain deterioration. The ortholog in humans is highly expressed in the brains of adults; expression can increase in age and can be misregulated in humans with degenerative disease.

There is now a *D. melanogaster* genetic reference panel, based on the sequencing of 168 highly inbred lines derived from a wild population using next-generation (Illumina and 454) sequencing methods (Mackay et al. 2012). The sequences obtained were compared with the *D. melanogaster* reference genome to compare molecular-genetic variation and variation in quantitative traits. The new genome data can be used to map quantitative trait loci (QTLs) and to conduct systems-genetics analyses of the relationship between molecular variation and genetic variation for complex traits.

## 7.11 Next-Generation Sequencing Methods and Beyond

The development of newer sequencing technology has revolutionized all of genetics (Janitz 2008; Mardis 2008a,b; Shendure and Ji 2008; Metzker 2010; Pareek et al. 2011). The second generation of sequencing platforms produces millions of DNA sequence reads in a single run and provides inexpensive, genome-wide sequence data for many uses, including mutation mapping, polymorphism discovery, and discovery of noncoding RNAs. These new methods allow high-throughput functional genomics research, including gene-expression profiling, genome annotation, small-noncoding RNA discovery, and profiling and detection of aberrant transcriptions that previously were available only through analysis by microarrays (Morozova and Marra 2008). Other applications include study of DNA methylation, posttranslational modification of histones, and nucleosome position on a genome-wide scale. Sanger sequencing remains useful for certain projects, especially those in the kilobase-to-megabase range, but genome sequencing by next-generation methods (NextGen) has revolutionized genome analysis of arthropods (Bentley 2006).

### 7.11.1 Next-Generation (NextGen or Second-Generation) Sequencing

The general process of sequencing DNA using NextGen methods includes several steps. First DNA is extracted and fragmented. The fragments are then ligated to adaptors. The ligated fragments are then processed in specific protocols, depending on the specific platform used, to develop an array of millions of immobilized PCR colonies (also known as **polonies**). Each polony consists of

**Table 7.2: Features of Several Next-Generation Sequencing Platforms.**

Platform	PCR type	Sequencing by synthesis	Read length (bp) <sup>a</sup>
Roche 454	Emulsion	Pyrosequencing	500–1000
Illumina (Solexa)	Bridge	Reversible terminators	20–40
SOLiD	Emulsion	Ligase	35
Polonator	Emulsion	Ligase	13
HeliScope	Single molecule	Polymerase	30

<sup>a</sup>Read length and cost change as the companies involved improve their platforms.

many copies of a single DNA fragment. All polonies are attached to the array, and reagents are then applied to amplify all polonies in the array in parallel. Image-based detection of fluorescent labels incorporated with each PCR extension is used to obtain sequence data in parallel. Successive iterations of PCR cycles and imaging are used to develop a contiguous sequencing read for each component of the array. 454 sequencing (Roche Applied Science), Illumina (Solexa technology), the SOLiD platform (Applied Biosystems), the Polonator (Dover/Harvard), and the HeliScope Single Molecule Sequencer (Helicos) all use this general scheme. The biochemistry of each platform is different; however, the workflows are similar. For details about each platform, go to the companies' websites for diagrams and additional references. The following descriptions provide a general overview, as does **Table 7.2**, of these sequencing systems.

#### 7.11.1.1 Roche (454) GS FLX Sequencer

This was the first NextGen platform available commercially and involves production of adaptor-flanked libraries of DNA fragments that are 400–600 bp ([Dressman et al. 2003](#), [Margulies et al. 2005](#)). The double-stranded DNA fragments are separated into single strands, each with a special adaptor. These fragments are loaded onto micrometer ( $\mu\text{m}$ )-sized beads using a proprietary process, each bead containing one fragment. Emulsion PCR (PCR that occurs within small water droplets contained in a synthetic oil) generates millions of copies on each capture bead. After breaking the emulsion, the beads are cleaned and beads lacking DNA are eliminated, whereas beads holding more than one type of DNA fragment are filtered out during the signal processing. Sequencing occurs by synthesis: a single-stranded DNA fragment is copied, making the fragment double-stranded. Starting at one end, the enzyme sequentially adds a single nucleotide that is the match of the nucleotide on the single strand. After the amplification of DNA strands, the DNA-capture beads are placed into tiny wells on a plate for sequencing. The wells are 44  $\mu\text{m}$  in diameter, and each well can hold only one bead. Each plate contains 1.6 million wells. After the wells are

filled with the capture beads, small enzyme-containing beads are added, and the plate is placed into the sequencing instrument. The instrument can wash the plate with A, C, G, and T nucleotides. The nucleotides are flowed sequentially in four washes over the plate; when these nucleotides are incorporated into the DNA strands, the enzymes on the enzyme-containing beads convert the chemicals generated during the nucleotide incorporation into light in a chemiluminescent reaction similar to that of a firefly. The method is called **pyrosequencing**, because it involves incubating the DNA-bearing beads with *Bacillus stearothermophilus* (*Bst*) DNA polymerase, single-stranded binding protein, ATP sulfurylase, and luciferase. When incorporation of a nucleotide occurs, pyrophosphate is released, resulting in a burst of light that is detected by the recording device. The signal strength is proportional to the number of nucleotides incorporated.

The 454 platform has difficulty in sequencing DNA regions with long strings of the same bases because the length of such sequences is inferred from the intensity of the signal and that can be a source of error. However, the 454 system can generate  $\approx$ 400,000 reads with lengths of 200–1000 bp each, the longest read lengths of the NextGen systems, but the most expensive. [Hayden \(2009\)](#) reported that this system sequences 1000 bp for US\$0.05.

#### 7.11.1.2 Illumina (or Solexa) Genome Analyzer

The Illumina analyzer, also known as the Solexa, uses adaptor-flanked DNA fragments up to several hundred base pairs and bridge PCR to amplify these fragments ([Bentley 2006](#), [Fedurco et al. 2006](#)). First, the genomic DNA is fragmented, and adaptors are ligated to both ends. The DNA is attached randomly to the surface of the flow-cell channels that contain a dense lawn of primers. Next, unlabeled nucleotides and enzyme are added to initiate bridge amplification. In bridge PCR, both forward and reverse PCR primers are attached to a solid substrate by a flexible linker so that all amplicons arising from a single template molecule during amplification remain immobilized and attached in a single location on the array. Bridge PCR relies on alternating cycles of extension with *Bst* polymerase and denaturation with formamide, resulting in clusters of  $\approx$ 1000 clonal amplicons. Several million clusters can be amplified at locations within each of eight lanes on a single-flow cell, which means that eight different libraries can be sequenced in parallel during a single run. Each cycle of sequence analysis consists of single-base extension with a modified DNA polymerase and a mixture of four nucleotides that are modified so that only a single base is incorporated in each cycle and one of four fluorescent labels is used to identify each nucleotide. Read lengths are  $\approx$ 20–100 nt. [Hayden \(2009\)](#) reported that this platform produces 1000 bp of sequence for US\$0.002.

#### 7.11.1.3 Applied Biosystems SOLiD Sequencer

Applied Biosystems SOLiD sequencing uses ligase to synthesize the DNA amplicons on the surface of tiny 1- $\mu\text{m}$  paramagnetic beads (Shendure et al. 2005) and has been commercially available since 2008. First, a library of the DNA is prepared, and emulsion PCR is performed in microreactors (with template DNA, PCR reaction components, beads, and primers). After the PCR, the templates are denatured and the beads are separated that contain extended templates. The template on the beads then undergoes a modification to attach it to the slide. Sequencing is by ligation. Primers hybridize to the adapter sequence on the beads. A set of four fluorescently labeled two-base probes compete for ligation to the sequencing primer. Multiple cycles of ligation, detection, and cleavage are performed with the number of cycles determining the read length. After a series of cycles, the extension product is removed and the template is reset with a primer complementary to the n-1 position for a second round of ligation cycles. Five rounds of primer reset are completed for each sequence tag, so that every base is evaluated in two independent ligation reactions by two different primers. After ligation, the images are obtained in four channels, thus obtaining data for the same base positions across all beads. Details of this process are available at the Applied Biosystems website. This approach typically produces reads of 35 bp, and Hayden (2009) reported the SOLiD system could produce 1000 bp of sequence for US\$0.002.

#### 7.11.1.4 Standards

Field et al. (2009) proposed that sharing of “Omics” data should be adopted as a standard component of such research and that databases should be developed to assist in this. Standard methods for describing, formatting, submitting, and exchanging data need to be developed for the enormous amounts of genetic data already developed and to be developed. Data release and exchange need to become part of the scientific culture, even though such requirements add costs to the project. Collecting, holding, and exchanging data requires long-term funding and a well-designed system, which is not fully available at this time.

One possible solution to the problem of storing vast amounts of data involves cloud computing (*Nature Biotechnology* 2010). There are at least a hundred research centers around the world generating thousands of gigabases of DNA sequence data every week. For example, a single Illumina machine can generate up to 90 billion bases per run, which results in large amounts of raw data to analyze and store. Data storage is possible for large sequencing facilities, but even their space is becoming an issue. For smaller laboratories with limited data-storage capacity, the option may be to handle and store data by cloud computing.

Cloud computing allows a user to rent processing time on a computer cluster; the user can load software and access to the cluster by disk or by the internet. Software is being developed that makes it possible to analyze an entire genome using cloud computing and a laptop (*Nature Biotechnology* 2010).

### 7.11.2 Third-Generation Sequencing

Although NextGen sequencing has provided scientific advances using high-throughput methods at a greatly reduced cost compared with Sanger sequencing, even NextGen sequencing methods cannot provide all the answers geneticists want (Harris et al. 2008, Eid et al. 2009, Pettersson et al. 2009, Schadt et al. 2010, Ghose 2012). Several third-generation sequencing technologies are being developed that will further reduce costs and the time required to sequence genomes (Bennett et al. 2005).

Second-generation sequencing techniques involve sequencing by synthesis, relying on PCR to grow clusters of a given DNA template, attaching the clusters of DNA templates to a solid surface that is then imaged as the clusters are sequenced by synthesis in a phased approach. By contrast, third-generation technologies are attempting to sequence *single molecules* of DNA so that biases introduced by PCR amplification are eliminated (Harris et al. 2008). The goal is to increase read length from tens of base pairs to thousands of base pairs per read and to reduce the time from days to hours or minutes (Schadt et al. 2010). Thus, the goal is higher throughput, faster turnaround, longer read lengths to enhance assembly, higher accuracy rates, the need for smaller amounts of starting material, and lower cost. Using such methods, the human genome, in theory, could be sequenced for <US\$100 (Schadt et al. 2010).

One approach to third-generation sequencing involves ion sequencing, in which each well lies above an ion-sensitive metallic oxide layer coupled to an electronic sensor that registers tiny and transient pH changes in each well as determined by the base sequence (Robison 2011).

The Ion Torrent Personal Genome Machine (PGM) costs ≈US\$50,000 and relies on detection of the release of hydrogen ions when DNA polymerase adds a nucleotide during DNA synthesis. The DNA is present in up to a million or more microwells that are flooded with each of the four nucleotides in succession. If the nucleotide complements that of the DNA template, it is incorporated, resulting in a hydrogen ion that is detected by a pH sensor and translated into a voltage change recorded by a semiconductor sensor. In January 2012, Ion Torrent released a more powerful machine, Ion Proton, that can “churn out a human

genome in a single day for the long-sought-after price of \$1000” ([Ghose 2012](#)). It is, however, not able to accurately sequence long stretches of the repetitive DNA.

Pacific Biosciences released the Single Molecule Real Time (SMRT) sequencer in April 2011. It relies on fluorescently labeled nucleotides, but it sequences just one molecule at a time by trapping a single DNA molecule, along with a DNA polymerase, in one of 150,000 tiny holes, and then floods the surface with all four fluorophore-labeled nucleotides. As DNA polymerase attaches each nucleotide to the sequence, laser beams of two different wavelengths illuminate each hole, and a camera can detect the particular nucleotide being incorporated into each chain. This machine can generate long sequencing reads (>1000 bp and up to 10,000 bp), but it is expensive at US\$700,000 and has a fairly high error rate compared with the Illumina platform.

Another approach to third-generation sequencing, nanopore sequencing, involves reading DNA sequences by resolving changes in ionic current that correspond to a known DNA sequence by passing a single DNA molecule through a protein pore of known size at a specific rate so that the DNA base can be identified accurately ([Derrington et al. 2010](#), [Cherf et al. 2012](#), [Manrao et al. 2012](#)). [Podolak \(2010\)](#) noted that third-generation sequencing methods could “change the landscape yet again,” and the National Institutes of Health (NIH) is supporting the race to achieve an effective third-generation technology with their US\$1000 human genome ([Podolak 2010](#)).

## 7.12 Bioinformatics

As the GASP project for *Drosophila* indicated, it is not always easy to find genes hidden amongst the thousands of nucleotide sequences produced by a genome-sequencing project. *Drosophila* analysis methods averaged a 70% accuracy rate in predicting structural and functional features ([Bork 2000](#)). Part of the problem is defining a “gene” ([Attwood 2000](#)). Is a gene a heritable unit corresponding to an observable phenotype? Is it genetic information that encodes a protein or proteins? Is it the genetic information that encodes RNA? Must the gene be translated? Is DNA a gene if the DNA is transcribed but not expressed? There are multiple definitions of a gene; hence, the estimates of the total number of genes in a sequenced genome can vary.

A variety of approaches have been taken to improve the process of finding genes in eukaryotes ([Stormo 2000](#)). For example, the *Drosophila* genome has **isochores**, long >300-kb DNA segments that are compositionally homogeneous on the basis of GC frequencies. As is found in humans, *Drosophila* isochores

are rich in coding sequences compared to genome segments lacking high GC frequencies ([Jabbari and Bernardi 2000](#)). The effectiveness of gene-finding programs is based on the type of information used by the program and the algorithm used to combine that information into a coherent prediction. Three types of information are used to predict the location of genes: 1) "signals in the sequence" such as splice-sites, 2) "content" statistics such as codon bias, and 3) similarity to known genes ([Stormo 2000](#)). Start and stop codons can be useful in predicting exons. Unfortunately, they can be uninformative if the reading frame is unknown. Some programs look for sites associated with promoters such as TATA boxes, transcription-factor binding sites, and CpG islands. Poly(A) signals are used to aid in identifying the carboxyl terminus of the gene. As the number of known coding sequences increases, the accuracy of gene-prediction programs improves because the larger sample size of known genes will allow for more reliable statistical measures, as well as a much greater likelihood of encountering a gene that is related to one that has been identified previously.

Large genomic projects can only be analyzed computationally; continued improvements in analysis and annotation methods are needed ([Ashburner 2000](#), [Pop and Salzberg 2008](#), [Alkan et al. 2011](#)). Advances have been made in identifying DNA sequences as coding or noncoding. Although current methods leave uncertainties, having the exact coding prediction is unnecessary. Even partially correct predictions can focus experiments to determine the true gene structure faster than would be possible if these predictions were unavailable. Continued advances in computational and experimental methods for identifying genes, their regulatory elements, and function are expected ([Stormo 2000](#), [Baxevanis and Ouellette 2001](#), [Chen and Tompa 2010](#), [Miller et al. 2010](#)).

[Yandell and Ence \(2012\)](#) state, "Sequencing costs have fallen so dramatically that a single laboratory can now afford to sequence large, even human-sized genomes. Ironically, although sequencing has become easy, in many ways, genome annotation has become more challenging." The reasons for this include the shorter read lengths of the NextGen sequencing platforms, so assemblies rarely are able to obtain sufficient contigs to include the whole genome. Also, as nonmodel organisms are sequenced, it becomes harder to identify genes in these novel species, especially when as many as one-third of the genes could be new, so-called orphan genes. Another issue is that today's annotation projects now involve scientists with little bioinformatics and computational biology training. Unfortunately, genome annotation is not yet a "point-and-click" process. However, [Yandell and Ence \(2012\)](#) provide a beginner's guide to genome annotation that is very helpful.

The amount of data produced from genomics is stressing the system (Kahn 2011). A single week-long sequencing run can produce as much data as did entire genome centers a few years ago. Thus, the availability of large genomic data sets raises concerns over access to the data and its security. “A doubling of sequencing output every nine months has outpaced and overtaken performance improvements within the disk storage and high-performance computation fields” (Kahn 2011).

Improvements in analysis methods are ongoing. The modENCODE Consortium (2010) involves projects to map transcripts, histone modifications, chromosomal proteins, transcription factors, replication proteins, and nucleosome properties across a developmental time course and in multiple cell lines, tripling the annotated portion of the *Drosophila* genome. Fan and Li (2012) reported the results of comparisons of eight genome assemblers. Zerbino et al. (2012) report on the development of models from population genetics and phylogenetics and their relationship with graph theory, statistics, signal processing, and computer science to “provide a rich quantitative foundation for genomics that can only be realized with the aid of a computer.”

### 7.12.1 Gene Ontology

Gene Ontology (GO) is a bioinformatics project with the goal of standardizing the representation of gene and gene-product attributes across species and databases. The goal is to provide a specific vocabulary of terms for describing gene product and gene-product annotation data, as well as to provide tools to access and process the data ([www.geneontology.org](http://www.geneontology.org)). The goal is to “unify” biology, so that different scientists use the same terms and references to function (Ashburner et al. 2000, Gene Ontology Consortium 2001). The ontology is set up with three sets of terms to describe molecular function, biological process, and cellular component. The model organisms (human, yeast and *Drosophila*) were annotated using GO terms, and this has become the standard for other genomes. The GO website provides access to the annotated gene-product data sets. It should be remembered, however, that *sequence similarity does not necessarily equate to similar functions*. Until the function of a gene has been evaluated in a particular species, the assigned function is “putative.”

## 7.13 Genome Analyses of Other Arthropods

The ability to sequence the *Drosophila* genome led to proposals to sequence the genomes of other arthropods, especially species that are of significant economic importance such as mosquitoes (*Anopheles gambiae*, the vector of malaria) and

**Table 7.3: Arthropod Genomes Completely Sequenced by July 2012.**

Species (common name)	Selected references
<i>Acromyrmex echinatior</i> (leaf-cutter ant)	Nygaard et al. 2011
<i>Acrythosiphon pisum</i> (pea aphid)	International Aphid Genomics Consortium 2010, Godfray 2010, Huerta-Cepas et al. 2010
<i>Aedes aegypti</i> (yellow fever mosquito)	Nene et al. 2007, Waterhouse et al. 2008
<i>Anopheles gambiae</i> (malaria mosquito)	Holt et al. 2002
<i>Apis mellifera</i> (honey bee)	Honeybee Genome Sequencing Consortium 2006
<i>Atta cephalotes</i> (leaf-cutter ant)	Suen et al. 2011
<i>Bombyx mori</i> (silk moth)	Mita et al. 2004, Xia et al. 2004, Wang et al. 2005, Goldsmith and Marec 2010
<i>Culex quinquefasciatus</i> (southern house mosquito)	Arensburger et al. 2010
<i>Danaus plexippus</i> (monarch butterfly)	Zhan et al. 2011
<i>Drosophila melanogaster</i>	Drosophila 12 Genomes Consortium 2007, Stark et al. 2007
<i>pseudoobscura, sechellia, simulans, persimilis, pseudoobscura, yakuba, erecta, ananassae, willistoni, virilis, mojavensis, grimshawi</i>	
<i>Glossina</i> species (tsetse flies)	VectorBase, <a href="http://www.vectorbase.org/sections/Forum/viewtopic.php?f=6&amp;t=760">www.vectorbase.org/sections/Forum/ viewtopic.php?f=6&amp;t=760</a>
<i>Heliconius melpomene</i>	Heliconius Genome Consortium 2012
<i>Ixodes scapularis</i> (black-legged tick)	Van Zee et al. 2007, Hill et al., unpubl.
<i>Linepithema humile</i> (Argentine ant)	Smith et al. 2011a
<i>Metaseiulus occidentalis</i> (western orchard predatory mite)	Hoy 2009, Hoy et al., unpubl.
<i>Nasonia vitripennis, giraulti, longicornis</i> (jewell wasps [parasitoids])	The Nasonia Working Group 2010
<i>Pediculus humanus humanus</i> (human body louse)	Kirkness et al. 2010
<i>Pogonomyrmex barbatus</i> (red harvester ant)	Smith et al. 2011
<i>Rhodnius prolixus</i> (kissing bug)	Gourbiere et al. 2012
<i>Solenopsis invicta</i> (red imported fire ant)	Wurm et al. 2011
<i>Tetranychus urticae</i> (twospotted spider mite)	Grbic et al. 2011
<i>Tribolium castaneum</i> (red flour beetle)	Tribolium Genome Sequencing Consortium 2008
<i>Varroa destructor</i> (ectoparasite of <i>A. mellifera</i> )	Cornman et al. 2010

ticks. A proposal was made to sequence the genome of *A. gambiae* in five years with a cost of between US\$50 and \$90 million using Sanger sequencing ([Balter 1999](#)). In 2001, plans were finalized for sequencing the *An. gambiae* genome at a cost of  $\approx$ \$10 million ([Balter 2001](#)). **Table 7.3** lists the arthropod species that have had their genomes sequenced by July 2012, with most of the recent genomes having been sequenced by next-generation (NextGen) sequencing methods.

**Table 7.3** will become outdated very rapidly because the use of newer sequencing techniques have reduced the costs and time required to sequence

the genome of nearly every arthropod of interest to both basic and applied scientists. In 2011, a group proposed to sequence the genomes of 5000 species of insects (and other arthropods) during the next 5 years (Robinson et al. 2011). The initiative, called i5K, is meant to transform our understanding of insect biology and improve our ability to manage pests.

### 7.13.1 Interesting Findings from Completed Genomes

Mining the genomes of the various arthropods sequenced to date will provide new information for years to come, especially when functional analyses are conducted in the laboratory to confirm proposed gene functions. Examples of key discoveries as the result of whole-genome sequencing projects to date are provided below; these examples represent selected highlights and the full papers should be read for additional insight and details. Much more will be learned as additional studies are conducted using these genome sequences.

The genomes of three mosquitoes were sequenced: *Anopheles gambiae*, *Aedes aegypti*, and *Culex quinquefasciatus* (Holt et al. 2002, Nene et al. 2007, Arensburger et al. 2010). Severson and Behura (2012) compared the results of these efforts, in which *Cx. quinquefasciatus* seems to have the most genes (18,883), *Ae. aegypti* has the next most (15,419), and *An. gambiae* has the fewest (12,457). Genome sizes vary dramatically, with *An. gambiae* having a 289-Mbp genome, *C. quinquefasciatus* having a 579-Mbp genome, and *Ae. aegypti* having the largest genome with 1380 Mbp. The larger genome size of *Ae. aegypti* is due to the fact that TEs comprise 50% of the genome. VectorBase (<http://www.vectorbase.org/>) contains the genomic data for these, and other insect or tick, vectors of human pathogens. Arensburger et al. (2010) noted that *Cx. quinquefasciatus* has a significant expansion in olfactory and gustatory receptors, salivary gland genes, and genes associated with detoxification.

Mita et al. (2004) sequenced the genome of the long-domesticated *Bombyx mori* and compared the sex-determination system to that of *Drosophila*, finding that the two systems differ greatly. Xia et al. (2004) provided a draft genome sequence, as well, and found an estimated 18,510 genes, greater than the number found in *D. melanogaster*. Approximately 1874 genes found are related to silk production.

The *Apis mellifera* genome was sequenced in 2006 (The Honeybee Genome Sequencing Consortium 2006). It has few transposons compared with the other genomes sequenced at the time, primarily consisting of *mariner* elements, and evolved more slowly than *D. melanogaster* and the mosquito *An. gambiae*. *Apis* has fewer genes than *D. melanogaster* or *Anopheles* for innate immunity, as well as fewer detoxification enzymes, cuticle-forming proteins, and gustatory receptors, perhaps reflecting their social behavior within the protected hive.

Many aspects of the complex eusocial life style of the honey bee are being deciphered with ongoing functional analyses (Adams et al. 2008, Chan et al. 2008).

The beetle *Tribolium castaneum* had its 160-Mbp genome completed in 2008 (*Tribolium Genome Sequencing Consortium* 2008). The beetle is a pest of stored products and has a large expansion in odorant and gustatory receptors. Its expansion of P450 and other detoxification enzymes may make it prone to developing resistance to pesticides used for its control. Its development is “more representative of other insects than is *Drosophila*” in both gene content and function. It exhibits short-germ band development and also has a systemic RNA interference (RNAi) capacity (which *D. melanogaster* lacks), although its system is different from that in *C. elegans*. One third of the genome consists of repetitive DNA.

The first parasitoid genome sequenced was that of *Nasonia vitripennis* (*Nasonia Genome Working Group* 2010). This tiny wasp is a gregarious endoparasitoid of dipteran pupae. The larvae feed gregariously within fly pupae, emerging as adults from the puparium. Genome sequences confirm that substantial amounts of *Wolbachia* DNA were transferred into its genome. The *Nasonia* genome also includes a strong presence of odorant-binding, gustatory-receptor, odorant-receptor, and venom-protein gene families, which is to be expected in a species that uses chemical cues to locate and inject venoms into hosts during oviposition. Interestingly, 60% of the *Nasonia* genes have a human ortholog, 18% are arthropod-specific, and 2.4% seem to be specific to the Hymenoptera. Approximately 12% seem to be specific to *Nasonia*. Because *Nasonia* feeds on an amino-acid-rich diet as larvae and adults, it seems to have lost some amino-acid metabolic pathways, reflecting its parasitic biology. The analysis of venom genes was of particular interest, and 79 candidate venom proteins were identified, some of which were unrelated to any known insect venoms.

The human body louse, *Pediculus humanus humanus*, genome was completed in 2010 (Kirkness et al. 2010). *Pediculus* is an obligatory parasite of humans and a vector of human pathogens. The louse has a primary endosymbiont *Candidatus Riesia pediculicola*, and its genome also was sequenced. The body louse has a small genome (108Mbp), with 10,773 protein-coding genes. The genome contains relatively fewer genes associated with environmental sensing and response, including odorant and gustatory receptors and detoxifying enzymes, perhaps relevant to its relatively protected parasitic life style. One unusual finding was that the louse has 18 minicircular chromosomes in the mitochondrion (compared with one chromosome in most arthropods). The symbiont

has <600 genes on a short, linear chromosome and a circular plasmid, which contains genes for synthesizing pantothenate, an essential vitamin.

The twospotted spider mite, *Tetranychus urticae* (Arthropoda: Chelicera: Acari: Tetranychidae), is not an insect at all, but an important chelicerate agricultural pest, feeding on >250 plant species. It produces silk, hence the name “spider mite.” Its genome was sequenced and provides another example (other than the pea aphid) of a genome of a plant-feeding pest (Grbic et al. 2011). This tiny genome (90 Mbp) contains novel genes for silk production, and expanded detoxification-gene families that are likely associated with feeding on so many different plants. Interestingly, carotenoid biosynthesis genes were found in *T. urticae*’s genome (Altincicek et al. 2012), as they were in the pea aphid (Moran and Jarvik 2010), suggesting that the carotenoid-synthesis genes were horizontally transferred from fungi to the ancestor of both aphids and the twospotted spider mite. The carotenoids found in *T. urticae* were thought to have been obtained from their host plants and are important in diapause induction. The spider mite, which has a greatly reduced segmentation, also has only eight of the 10 expected *Hox* genes, with *Hox3<sup>+</sup>* and *abdominalA<sup>+</sup>* missing. Silk is produced by paired glands connected to the pedipalps (mouthparts) of spider mites, and it is stronger and thinner than the silk produced by the spider *Nephila clavipes*.

The pea aphid, *Acyrtosiphon pisum*, genome was sequenced and found to be very large (464 Mbp) (Godfray 2010, International Aphid Genomics Consortium 2010). The genome has extensive gene duplications in >2000 gene families, and it has lost some evolutionarily conserved genes. Some gene duplications seem to belong to TEs and encode reverse transcriptases and transposases. Other duplications involve sugar-transporter proteins that may help the aphid feed on phloem, which is rich in carbohydrates but low in nitrogenous nutrients. Comparisons of all the 34,600 proteins of the aphid with the proteins of the 13 other fully sequenced arthropods indicated aphids have had multiple gene expansions specific to aphids, with the highest protein-gene content of all species sequenced so far (Huerta-Cepas et al. 2010). Examples of gene families that have greatly expanded include amino-acid transport, anti-apoptosis, response to oxidative stress, sensory perception of smell, carbohydrate transport, and olfactory behavior. A number of gene losses have occurred (defense response, immune response, detection of bacterium, antimicrobial humoral response, and sensory perception of taste). Moran and Jarvik (2010) found that the pea aphid has multiple enzymes involved in carotenoid biosynthesis that were derived from fungal genes that transferred horizontally and became integrated into the aphid genome and duplicated. The pea aphid was the first animal known to make its own carotenoids.

The monarch butterfly, *Danaus plexippus*, genome was sequenced by [Zhan et al. \(2011\)](#) and found to have 16,866 protein-coding genes. The monarch has a full repertoire of the circadian clock, and an extensive expansion of chemo-receptors that may be important for its long-distance migration behavior. The monarch also contains a sodium-potassium pump that may be important in chemical defense against cardenolide glycosides, compounds that are sequestered from their host plants for defense against predation. The authors detected differences in small RNAs in populations that undergo migration compared with those that do not.

The [Heliconius Genome Consortium \(2012\)](#) sequenced the genome of *H. melpomene* and compared it with the genomes of three related species. They found that this genus of tropical butterflies hybridizes and its three mimics (*H. timareta*, *elevatus*, and *melpomene*) exchange genes that provide protective color patterns. [Nadeau et al. \(2012\)](#) compared specific regions of the sequenced genomes in the species and races and found differences between races across a hybrid zone, indicating these genomic regions are under strong divergent selection.

Ant genomics is well advanced, with seven complete genome sequences published for the leaf-cutter ant *Atta cephalotes* ([Suen et al. 2011](#)), the leaf-cutter ant *Acromyrmex echinatior* ([Nygaard et al. 2011](#)), the Argentine ant *Linepithema humile* ([Smith et al. 2011a](#)), the red harvester ant *Pogonomyrmex barbatus* ([Smith et al. 2011b](#)), Jerdon's jumping ant *Harpegnathos saltator*, the Florida carpenter ant *Camponotus floridanus* ([Bonasio et al. 2010](#)), and the red imported fire ant *Solenopsis invicta* ([Wurm et al. 2011](#)). Sequencing of other ant genomes is in progress ([Gadau et al. 2012](#)). All 14,000 described ant species are eusocial, have diverse life histories, and are a dominant component of terrestrial habitats. Eusociality apparently originated once in the ant lineage, but there are different types of social organization. Genome sizes among the seven species range from 250 to 753 Mbp, with differences in size due primarily to differences in repetitive elements. *Solenopsis invicta* has the largest genome and seems to have 64% of its genome as repetitive elements. All ant genomes, as has the honey bee genome, have a reduced number of innate-immunity genes compared to solitary insects. A reduced innate-immunity system is thought to be due to the fact that social insects conduct hygienic behaviors in the nest, removing dead or diseased individuals, as well as grooming themselves and each other. By contrast the hymenopteran parasitoid *Nasonia*, which is not eusocial, has a normal complement of innate-immunity genes. Studies on the control of castes and other important aspects of social behavior now can be studied.

The genome of the vector of Chagas disease of humans, *Rhodnius prolixus*, is being sequenced ([Gourbiere et al. 2012](#)), and the data obtained are expected to

allow the design of improved vector control methods. Likewise, the *Ixodes scapularis* genome project (Arthropoda: Chelicerata: Acari: Ixodidae) should provide important data on another major vector of human disease (Lyme disease) (Van Zee et al. 2007). Its genome is very large and full of repetitive elements, so a full analysis has not yet been published.

Once *D. melanogaster*’s genome was sequenced, 12 other species of *Drosophila* have had their genomes sequenced and this provided an opportunity to compare functional elements in them (Stark et al. 2007).

### **7.13.2 What Do You Need to Do to Sequence Your Favorite Insect’s Genome?**

If you wish to sequence the genome of your favorite insect using NextGen sequencing, several steps will be involved. Ideally, you will know the size of the target species’ genome. That is important to the sequencing project because it is necessary to know how many reads will be needed to obtain the desired coverage of the genome. Also, if possible, it helps to have an inbred line from which to extract DNA. Inbreeding will reduce variability in the genome, which will be helpful in making the assembly of the short reads into scaffolds. Unfortunately, by inbreeding you will lose some of the genetic variability in your species, but resequencing will discover that. Next, you need an effective DNA extraction protocol that will provide clean DNA that has a low level of fragmentation so that the libraries can be prepared for NextGen sequencing. Once that is done, a skilled University core facility or commercial facility will prepare the libraries, conduct the sequencing reactions, and assemble the data using programs appropriate to the sequencing method.

Once the DNA has been assembled, the genome will need to be annotated, and annotation typically is done using both automated and manual methods. As described by Yandell and Ence (2012), genome annotation is a complex and lengthy process that can involve the use of multiple software packages and analysis methods. The full discussion of genome annotation is beyond the scope of this chapter, but Yandell and Ence (2012) provide “A beginner’s guide to eukaryotic genome annotation” that is helpful in explaining the terminology and steps involved, as well as listing and referencing the many software sources used for different purposes. Chain et al. (2009) discuss the different criteria by which genomes can be compared for their quality (Box 7.1).

Analysis of the genomes of nonmodel organisms is best done if protein or RefSeq or expressed sequence tag (EST) data are available to aid in the delimitation of exons, introns, and splice variants, and involves searching databases containing related species’ genomes to determine whether the sequences match

**Box 7.1 Not All Genomes Are Equal: Categories of Genome-Analysis Quality**

Type	Description
Standard Draft	Contains minimally or unfiltered data, which are assembled into contigs. This is the minimum standard for a submission to the public databases. Sequence of this quality will likely harbor many regions of poor quality and can be relatively incomplete, but is the least expensive to produce and contains useful information.
High-quality Draft	Includes at least 90% of the genome or target regions, excluding contaminating sequences. It is still a draft assembly with little or no manual review of the product.
Improved High-quality Draft	Additional work by either manual or automated methods, should contain no discernible misassemblies and should undergo some type of gap resolution to reduce the number of contigs and scaffolds. Low-quality regions and potential base errors may be present, but the quality is adequate for comparison with other genomes.
Annotation-directed Improvement Quality Draft	This may overlap with previous standards, but emphasizes the verification and correction of anomalies within coding regions. Gene models and annotation should fit the biology. Problems should be noted in the submission. Repeat regions are not resolved.
Noncontiguous Finished	High-quality assembly subjected to automated and manual improvement, and gaps, misassemblies, and low-quality regions have been resolved. Some repetitive or other areas, including heterochromatin may not be resolved. For eukaryotes, this used to be called “finished.”
Finished	Refers to the gold standard of less than 1 error per 100,000 bp and each replicon is assembled into a single contiguous sequence. All sequences are complete, reviewed and edited. Repetitive sequences have been assembled. Achieved primarily with small microbial genomes.

Modified from Chain et al. (2009).

to other genes. This provides a preliminary set of annotations, but these are provisional and it is likely that up to one-third of your sequences will have no homology to any in the databases. These so-called orphan genes may be of particular interest because these may be providing your species with the traits that are unique to its biology. A community-based annotation phase comes next, in which cooperating scientists assist in comparing gene families. Functional gene analyses also need to be done to confirm the function of candidate genes that may support conclusions about the function of specific genes. In addition to annotating protein-coding genes, the identification of transposons, regulatory regions, pseudogenes, and noncoding RNAs can be conducted, although these are more difficult. The resulting information can be published, but it provides only a starting point for understanding the genome of your species. Genome sequences can be used as the basis of experiments for years, using a variety

of methods to resolve many biological questions. For example, once whole-genome sequences are available, microarrays can be produced to analyze gene expression in different tissues and life stages. Metabolic pathways can be identified, or the whole genome can be compared with other whole genomes of other arthropods. Tiling microarray experiments can be done to assay transcription at regular intervals of the genome, using regularly spaced probes; these allow the discovery of genes with rare transcripts (Johnson et al. 2005). Finally, as more is learned about the genome, updates of the annotation need to be conducted so that accurate data are available for use by others.

## 7.14 Transposable Elements (TEs) as Agents of Genome Evolution

Our perception of the role of TEs in genome evolution has undergone a rapid change as more complete genomes are compared (Kidwell and Lisch 2001, Levin and Moran 2011). TEs now may be thought of as “natural genetic engineering systems” that act to provide genetic variability and other functions (Shapiro 1999). Their designation as “selfish DNA” or “junk DNA” is “... either inaccurate or misleading and ... a more enlightened view of the transposable element–host relationship encompasses a continuum from extreme parasitism to mutualism” (Kidwell and Lisch 2001).

TEs do carry costs; they require host-cell functions to replicate and proliferate, and their activity poses a risk to the host because their integration into new sites in the genome often results in deleterious mutations. In *D. melanogaster*, retrotransposons are responsible for as much as 80% of all spontaneous mutations (Miller et al. 1997). Despite these negative aspects, TEs are abundant and ancient components of eukaryotic genomes and their long coexistence within eukaryotic genomes suggests there has been some form of host–TE coevolution. We know that TEs can acquire a functional role in the host genome; in *D. melanogaster* retrotransposons are the telomeres (Eickbush 1997). TEs cause inversions in *Drosophila* species (Caceres et al. 1999), which can tie up beneficial gene combinations so that they are not scrambled during recombination.

The fact that TEs can be activated by environmental and population factors indicates that TEs have a positive role by creating new genetic variability that could be useful under conditions that reduce the fitness of an organism (Capy et al. 2000). One hypothesis proposes that the activation of host-defense genes during stress and the activation of TEs are similar processes. Alternatively, stresses could induce destabilization of the genome, leading to the malfunction of genetic systems that would lead to the increased activity of TEs as a secondary, rather than a direct, effect of stress (Capy et al. 2000). Over evolutionary

time, TEs have provided novel regulatory regions to preexisting host genes and TE-derived components have undergone a molecular transition into novel host genes through a process called “molecular domestication” (Miller et al. 1997).

TEs may be more than just agents for local mutations; TEs might provide coordinated changes in the genome by inserting into a series of genes whose products already function together (Shapiro 1999). Under this scenario, different insertions could recruit new proteins into the system. During periods of extensive genome reorganization, TEs could interact with cell signals to confer a far higher probability of evolving useful new multilocus systems. Thus, the relationship of TEs to their host could resemble “more symbiosis than parasitasis” (Brosius 1999).

The association of TEs with their hosts over evolutionary time could lead to three different outcomes: 1) the coevolution of TE-derived mechanisms to minimize the negative effects of TEs on their hosts (such as transposon self-regulation, tissue specificity, targeting and genome partitioning); 2) the evolution of host-defense mechanisms, which include host suppressors; and 3) the evolution of new and altered functions of TEs in hosts (regulatory functions, structural functions, enzymatic functions, and new coding sequences) (Kidwell and Lisch 2001). Thus, TEs could enable genomes to enhance their own evolution and serve as a major source of tools for generating the necessary diversity to respond to changes in the environment. Analysis of the *D. melanogaster* genome yielded an estimate of approximately one horizontal transfer event per *TE family* every 20 million years (Schaack et al. 2010).

## 7.15 Transcriptomics

Transcriptome analysis is often done to discover genes and to annotate both coding and noncoding regions of a sequenced genome. Sequencing of transcriptomes of arthropods also is often done before sequencing the genomes of nonmodel organisms because it provides important information as to “what is a gene” (Iyer and Chinaiyan 2011). For example, Mita et al. (2003) analyzed the cDNAs from multiple life stages and tissues to discover as many genes as possible of the silk moth *Bombyx mori*, and Nunes et al. (2004) developed expressed sequence tags (ESTs) from the honey bee *Apis mellifera* before the sequencing of their genomes. Transcriptomic analyses also allow scientists to determine what genes are being transcribed in what tissues and in which developmental stage. A **transcriptome** is an analysis of the transcripts of genes, including large and small RNAs. It can provide information on novel transcripts from unannotated genes, and information on splicing variants.

Sanger sequencing of mRNAs (actually cDNAs derived from RNAs) is expensive and provides relatively few ESTs because it often detects only the more abundant transcripts (Nagaraj et al. 2006). The use of microarrays also offers a relatively limited ability to identify and quantify the diverse RNA molecules that are expressed in genomes (Ozsolak and Milos 2011). Next-generation sequencing techniques (Illumina, SOLiD, and 454) allow rapid and high-throughput RNA sequencing without the construction of clone libraries, allowing the sequencing of cDNA fragments in a few days and at a low cost (Haas and Zody 2010, Martin and Wang 2011, Ozsolak and Milos 2011).

High-throughput mRNA sequencing (RNA-Seq) typically provides 100–1000 reads per base pair of a transcript, which can provide a nearly complete transcriptome, including rare transcripts with regulatory roles such as microRNAs, PiwPi-interacting RNAs, small nucleolar RNAs, and small interfering RNAs (siRNAs). In fact, RNA-Seq studies have found that non-protein-coding RNAs make up most of the transcriptome, with protein-coding RNAs less common (Gingeras 2009, Hass and Zody 2010). RNA-Seq studies using NextGen sequencing provides relatively short sequences (35–500 bp), so bioinformatic analysis is more complex than with Sanger sequencing (Martin and Wang 2011, Ozsolak and Milos 2011). Some of the transcripts are not co-linear, containing sequences from distant parts of the genome, produced by posttranscriptional events (chimeric RNA) (Gingeras 2009).

One example of transcriptomic studies using RNA-Seq includes an analysis by Burke and Moran (2011) of the pea aphid and its facultative symbiont *Serratia symbiotica*. Infection by the symbiont elicited small changes in expression of only 28 genes, and no change in expression of innate-immunity genes. This suggested that, although *S. symbiotica* has a major influence on the host's metabolome and resistance to heat, this is likely due to the metabolism of the symbiont itself or to posttranscriptional modification of aphid-gene expression and not due to changes in gene transcription by the aphid. Another example involved the analysis of the transcriptome of the Asian tiger mosquito *Aedes albopictus* to identify candidate transcripts for diapause preparation (Poelchau et al. 2011). The authors used 454 sequencing to obtain ESTs that were compared with the fully sequenced genome of *Ae. aegypti*. The data suggested 57 genes with higher expression and 257 with lower expression were active under diapause-inducing conditions. Quantitative PCR was conducted to confirm the results of the 454 sequence data. The combined data provide new opportunities to analyze the genetic basis of insect diapause in a mosquito of medical importance.

Transcriptome analysis also can be done to compare the activity of genes of insects under different physiological conditions. For example, Doroszuk et al. (2012) compared transcriptomes of a long-lived strain of *Drosophila* with flies with a normal life span and discovered significant differences in stress-related genes and in reproduction. Long-lived flies expressed reproduction genes at a lower level and did not down-regulate them with age, and several candidate markers of aging and life span extension were identified.

### 7.15.1 Tiling Microarrays

Tiling microarrays can be used to analyze and measure gene expression levels (Mockler and Eckler 2005, Oliver 2006, Liu 2007). These microarrays incorporate sequences from the whole genome and can be used to analyze DNA methylation status, genome-wide-binding locations of transcription factors and other DNA-binding proteins (this is called ChIP-chip, in which chromatin immunoprecipitation is detected on a chip). In addition, it is possible to localize nucleosomes, which determines which genes are able to function in the genome.

## 7.16 Metagenomics

Metagenomics involves sampling the genome sequences of a community of microorganisms inhabiting a common environment (Marco 2011). It may be defined as any type of analysis of DNA obtained directly from the environment and typically involves analysis of microbial genomes (Hugenholtz and Tyson 2008). It is a culture-independent method of identifying microbes (or their genes) in a specific environment. Because so few (probably <5%) microorganisms can be cultured, it is the only way to investigate the activity of microorganisms in many environments, including the guts of termites or other insects. Obviously, analysis of the data is easier if the environment sampled is relatively simple, but metagenomic studies have been applied to complex environments such as the human microbiome, the enormous complex of microorganisms associated with the human body. Sometimes whole microbial genomes can be reconstructed from an environmental sample, especially if the environment is relatively simple. Often only genes or physiological pathways can be identified. So far, analysis of insect microbial communities attempts to eliminate the genome of the insect to reduce sequencing costs, but if sequencing costs are reduced significantly further, it could be tractable to sequence and analyze both insect and symbiont genomes. Metagenomics could answer some interesting pest-management problems. For example, Cox-Foster et al. (2007) conducted a metagenomic survey of microorganisms in bee colonies affected and unaffected by colony collapse disorder. Metagenomics also can be conducted to determine what genes are involved in digestion in termite guts.

Metagenomics analysis can be done using several methods. One method involves extracting genomic DNA from organisms from a sample from a specific environment. The DNA is then cut into uniform lengths and inserted into a vector, replicated in *E. coli*, and sequenced using traditional Sanger sequencing. The clones are then sequenced and analyzed. More recently, sequencing is by “shotgun sequencing” of randomly sampled DNA, which is less expensive, using next-generation sequencing methods.

Data analysis involves assembly of short overlapping sequence reads into a consensus sequence and predicting which genes or partial genes are present. In complex communities with many microbial species, the reads may not assemble because overlapping reads may not be produced. The amount of an individual microbe’s genome that is sequenced depends on how abundant it is in the environment and how many reads are produced. It also is possible to reverse-transcribe RNAs from the environment into DNA and sequence these, which provides information on metabolic pathways.

### 7.17 Proteomics: Another “-Omic”

Once we know the complete DNA sequences of organisms (genomics), the next goal is to understand how the genes are translated in living cells (proteomics). What proteins function to provide structure and function in the living organism? Proteomics was first formalized as a term in 1996 and combines “proteins” and “genomics.” Definitions of proteomics and the other “-Omics” seem to be evolving, however.

**Proteomics** is the genome-wide analysis of proteins and includes three aspects: 1) characterization of proteins and their posttranslational modifications, 2) “differential display” to compare protein levels and types, and 3) studies of protein–protein interactions. Proteomics uses mass spectrometry and two-dimensional protein gel electrophoresis ([Geisow 1998](#), [Dutt and Lee 2000](#), [Pandey and Mann 2000](#)).

Two-dimensional gel electrophoresis allows the identification of proteins whose expression changes in an interesting manner from that of a reference point. Two-dimensional gel electrophoresis separates proteins *by charge* using isoelectric focusing and *by size* using sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). Up to 11,000 proteins from a single mixture can be resolved ([Dutt and Lee 2000](#)). SDS-PAGE can purify proteins for amino-acid analysis, mass spectrometry, and amino-acid sequencing.

Mass spectrometry of proteins separated on two-dimensional gel electrophoresis generates different types of structural information about a protein. For

example, it provides information on the mass of a protein and also generates amino-acid sequence information (Dutt and Lee 2000). Furthermore, mass spectrometry can provide data on glycosylation patterns, phosphorylation, and other posttranslational modifications of proteins.

Proteomics (data on protein expression profiles) can be linked to data on nucleic-acid sequence. Several software packages are available to compare multiple protein-expression profiles to identify quantitative changes (Dutt and Lee 2000). For example, proteomics can identify proteins that are associated with growth control, or responses to high or low temperatures, or to different chemicals.

Proteomics data are available in databases, including the Protein Information Resource (PIR) (Barker et al. 2000) and SWISS-PROT (Geisow 1998, Bairoch and Apweiler 2000). These sites may contain search engines to compare sequence similarity and protein function. The SWISS-PROT site provides information on the function of a protein, its domains structure, its secondary structure ( $\alpha$  helix,  $\beta$  sheet), quaternary structure (homodimer, heterotrimer), similarity to other proteins, diseases associated with deficiencies in the protein, posttranslational modifications, and variants. SWISS-PROT has cross-references to additional databases (Bairoch and Apweiler 2000). The PIR site provides databases and search tools (BLAST, FASTA, pattern/peptide, pairwise alignments, multiple alignments, domain search, global or domain search, and GeneFIND) as well as technical bulletins and documentation (Barker et al. 2000).

A protein-complex network for *D. melanogaster* was completed in 2011, providing the largest metazoan protein-complex map to date (Guruharsha et al. 2011). It assigns functional links to 586 protein-coding genes that had no previous functional annotation and linked 2297 proteins in 10,969 interactions. The goal is to understand how all the functional units in the cell work together to control development and physiological status.

## 7.18 Functional Genomics

Functional genomics has been transformed from a concept that was considered futuristic in the 1980s to an accepted science. **Functional genomics**, the assignment of function to genes, includes understanding the organization and control of genetic pathways that come together to make up the physiology of an organism (Eisenberg et al. 2000).

Using DNA microarrays or gene chips or transcriptomics, scientists can analyze complex mixtures of RNA and DNA in a parallel and quantitative fashion.

DNA microarrays can be used to measure levels of gene expression (mRNA abundance) for tens of thousands of genes simultaneously ([Brazma and Vilo 2000](#), [Celis et al. 2000](#), [Lockhart and Winzeler 2000](#)). As a result of the microarray revolution, scientists are faced with an avalanche of data on mRNA expression, or, as expressed by [Eisenberg et al. \(2000\)](#) “piles of information but only flakes of knowledge.” Microarrays were described in Chapter 6. Another method used in functional genomics is TILLING.

**Tilling** (targeting induced local lesions in genomes) is a method of analyzing gene function that does not require the insertion of TEs to disrupt function ([Kurowska et al. 2011](#)). It involves mutagenesis with a chemical mutagen, such as ethyl methanesulfonate (EMS), and a sensitive method for identifying single-base mutations in a target gene. It is a type of **reverse genetics** (analysis from genotype to phenotype). Although about 15,000 *D. melanogaster* genes are annotated, only  $\approx$ 6000 genes have a TE insertion. TE insertions disrupt gene function, causing mutations, or can cause deletions in the gene (and subsequent mutations) when they move by excision. To study genes lacking TE insertions, [Winkler et al. \(2005\)](#) screened genomic DNA from 2086 mutagenized fly lines. They then screened the library for mutations in three genes. They concluded that TILLING is useful to obtain mutations in genes of interest in *Drosophila* so that function can be analyzed.

## 7.19 Structural Genomics—Another New Horizon?

The *Drosophila* and Human Genome Projects united a large group of geneticists and others in a coordinated effort to obtain massive amounts of genomic data in a relatively short time. Such large-scale biology projects were unprecedented in biology. Another new initiative began in 2000 called The Structural Genomics Project ([Smith 2000](#)). Large sums of money were allocated to the project by the United States and Japan. Once again, the project elicited concern and apprehension among biologists because it is difficult and expensive. The Structural Genomics Project was estimated to cost more and to be more complex than the Human Genome Project.

**Structural genomics** involves large-scale analysis of protein structures and functions based on gene sequences. Structural genomics developed after the genome-sequencing projects began, and advances in structural determination of proteins were obtained. The Structural Genomics Project aims to link sequence, structural, and functional information and enable the prediction of unknown structures by homology modeling. The Structural Genomics Project began in January 1998 and hopes to determine protein structures of 10,000 proteins, one

or more from each “fold family” within 10 years (Norvell and Machalek 2000, Terwilliger 2000). Once developed, this enormous body of structural data will be made available in public databases and promises to “accelerate scientific discovery in all areas of biological science” (Burley 2000).

## 7.20 Comparative Genomics

Now that multiple arthropod genomes have been completely sequenced, it is becoming possible to compare whole genomes to understand how genomes evolve (Kondrashov 1999, Rubin et al. 2000, Singh et al. 2009).

**Comparative genomics** attempts to learn the following: how many protein families are encoded in the genomes, the number of gene duplications, how similar genes are in different species, the degree of similarity of protein domains and families, and the similarity of developmental strategies and cellular processes (cell division, cell shape, cell–cell interactions) (Rubin et al. 2000). Approximately 30% of the predicted proteins in every organism bear no similarity to proteins in other organisms (Rubin et al. 2000). These genes are called orphan genes, and their origin remains poorly understood, although they are thought to be important for the specific aspects of biology unique to an organism (Tautz and Domazet-Loso 2011). Although many genes are thought to originate by duplication and rearrangements of the genome, there is speculation that orphan genes may derive by evolution from noncoding genomic areas.

Zdobnov and Bork (2007) compared the completely sequenced genomes of the honey bee, red flour beetle, silk moth, several *Drosophila* species (including *melanogaster*, *erecta*, *ananasseae*, *pseudoobscura*, *mojavensis*, *virilis*, and *grimshawi*) and two mosquitoes (*An. gambiae* and *Ae. aegypti*). They attempted to quantify the relationship between gene order and protein-sequence identity in 4632 single-copy orthologs (genes in different species that originated by descent from a common ancestor). Zdobnov and Bork (2007) reported, “Insect genomes are much more diverse than those of comparable vertebrate lineages.” The average protein-identity conservation ranged from 53 to 95% and seemed to vary based on evolutionary distance. Another comparison made was the difference in gene arrangements along the chromosomes (**synteny**), which are affected by inversions and translocations. The results “show that there is only limited selection for conservation of gene order and reveal a few hundred genes, proximity among which seems to be vital.”

Singh et al. (2009) reviewed the results obtained from an analysis of the 12 complete *Drosophila* species genomes (*Drosophila* 12 Genomes Consortium 2007), including genome sizes, genome structures, and genome contents. The

amount of coding DNA among the 12 species varies by approximately 1.7-fold, ranging from 111 Mbp in *D. simulans* to 187 Mbp in *D. willistoni*. The genomes are contained in six homologous chromosome arms, and the genes within the arms are largely conserved among the *Drosophila* species evaluated. Within each arm, however, the order and orientation of the genes can vary due to gene rearrangements caused by a large amount of gene movement by TEs. The comparisons of specific genes indicated that 11.9% of all new genes found were due to development of a totally new gene (not due to gene duplication). This indicates that novel genes provide novel aspects of *Drosophila* species’ biology. A great deal more was learned from these evaluations, including how to improve alignment and comparison methods.

## 7.21 Interactomes or Reactomes

Now that genomes have been sequenced and studied intensively, it is apparent that learning the DNA sequence is the first step in understanding the biology of an organism (*Nature Biotechnology* 2011). Researchers are now attempting to build networks of protein–protein interactions as a method for understanding signaling cascades and mechanisms of cell biology. Understanding how genes, and their proteins, work in pathways is now the goal (Beloqui et al. 2009, Bonetta 2010).

## 7.22 The Post-Genomic Era: Systems Genetics

During at least the past 50 years, biology was dominated by a reductionist approach, which narrowed the focus from the entire animal in its natural environment to increasingly smaller parts. The enormous complexity of a living organism overwhelmed the ability of biochemists, cell biologists, structural biologists, physiologists, and geneticists to study the whole animal (Vukmirovic and Tilghman 2000). Studies of organs, then cells and, finally, individual molecules became the focus of analyses.

The reductionist approach will continue to be productive, and necessary, to obtain detailed knowledge about gene function, gene regulation, and gene sequences of the genomes of organisms. One example of a sophisticated reductionist approach is the sequencing of the genomes of multiple individuals in a population to quantify molecular genetic variation. For example, Mackay et al. (2012) produced the *Drosophila melanogaster* Genetic Reference Panel, consisting of the genome sequences of 168 inbred lines derived from a natural population in North Carolina. The goal is to learn about quantitative traits, and the panel of fly lines is being made available to all geneticists interested in complex traits.

Biology is “suddenly awash in genome-based data ...” and “... is in the midst of an intellectual and experimental sea change,” which is revolutionizing the type of questions biologists can ask ([Vukmirovic and Tilghman 2000](#)). The future may become even more awash in genomic data soon: BGI, formerly known as Beijing Genomics Institute, has as its goal to “sequence every living thing on earth” ([Science 2011](#), [Callaway 2011](#)). Currently, BGI is able to produce the equivalent of 1500 human genomes each day using 160 sequencers (Illumina and SOLiD). It also provides analysis services and storage services ([Nature 2011](#)).

Leroy Hood, the leader of a team that invented the automated Sanger sequencer, noted recently “The future will be the study of the genes and proteins of organisms in the context of their informational pathways or networks” ([Smaglik 2000](#)). The next phase may be termed “systems genetics.” The goal of systems genetics is to understand “how genetic information is integrated, coordinated, and ultimately transmitted through molecular, cellular, and physiological networks to enable the higher order functions and emergent properties of biological systems ([Ala-Korpela et al. 2011](#), [Nadeau and Dudley 2011](#)). Systems genetics could produce predictions that result in experiments to confirm function.

In the near future, it should be possible to monitor simultaneously the expression of genes at the RNA or protein level, all possible protein–protein interactions, all alleles of all genes that affect a particular trait, and all protein-binding sites in a genome. It soon could be feasible to describe the properties of whole organisms in a precise and quantitative way.

Changing our focus in biology from a reductionist approach to an integrative one provides new challenges ([Palsson 2000](#)). Given the complexity of an organism, mathematical models and computer simulations will be required to study the integrated function of multiple gene products. Models will be required to analyze, interpret, and predict the relationship between genotype and phenotype. An early example of this integrative approach is a model that provides a comparative mathematical analysis of the genome and metabolic networks of 43 organisms representing the three domains of life ([Jeong et al. 2000](#)).

An integrative approach to biology will rely on improved bioinformatics methods and whole-systems analyses to understand the properties of cellular and tissue functions, and focus on the “**emergent properties**” properties that arise from the whole rather than the individual parts ([Palsson 2000](#)). Future geneticists will need to enhance their computer skills and have a higher level of mathematical and informatics training. The enormous increase in knowledge of genomics during the 21st century will have at least as large an effect on the

world as the changes brought about by the development of electronics and computation during the 20th century ([Brent 2000](#), [Nature Biotechnology 2011](#)).

## General References

- Gibson, G., and Muse, S.V., (2009). *A Primer of Genome Science*, 3rd ed. Sinauer, Sunderland, MA.
- Goldsmith, M.R., and Marec, F., Eds., (2010). *Molecular Biology and Genetics of the Lepidoptera*, CRC Press, Boca Raton, FL.
- Haddock, S., and Dunn, C., (2010). *Practical Computing for Biologists*. Sinauer Associates, Sunderland, MA.
- Harbers, M., and Kahl, G., Eds., (2012). *Tag-based Next Generation Sequencing*, Wiley-VCH Weinheim, Germany.
- Hartwell, L., Hood, L., Goldberg, M., Reynolds, A., and Silver, L., (2010). *Genetics: From Genes to Genomes*, 4th ed. McGraw Hill Science, New York.
- Lesk, A.M., (2012). *Introduction to Genomics*, 2nd ed. Oxford University Press, New York.
- Pevsner, J., (2009). *Bioinformatics and Functional Genomics*, 2nd ed. Wiley-Blackwell, Hoboken, NJ.
- Rapley, R., and Harbon, S., Eds., *Molecular Analysis and Genome Discovery*, 2nd ed. Wiley-Blackwell, Malden, MA.
- Sensen, C.W., (2008). *Essentials of Genomics and Bioinformatics*. John Wiley & Sons, New York.

## References Cited

- Adams, M.D., Celtniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D., and Amanatides, P.G., et al. (2000). The genome sequence of *Drosophila melanogaster*. *Science* **287**: 2185–2195.
- Adams, H.A., Southey, B.R., Robinson, G.E., and Rodriguez-Zas, S.L., (2008). Meta-analysis of genome-wide expression patterns associated with behavioral maturation in honey bees. *BMC Genomics* **9**: 503.
- Ala-Korpela, M., Kangas, A.J., and Inouye, M., (2011). Genome-wide association studies and systems biology: together at least. *Trends Genet.* **27**: 493–496.
- Alkan, C.S., Sajadian, and Eichler, E.E., (2011). Limitations of next-generation genome sequence assembly. *Nat. Methods* **8**: 61–65.
- Altincicek, B., Kovacs, J.L., and Gerardo, N.M., (2011). Horizontally transferred fungal carotenoid genes in the two-spotted spider mite. *Tetranychus urticae*. *Biol. Lett.* **8**: 253–257.
- Ambrose, B.J.B., and Pless, R.C., (1987). DNA sequencing: chemical methods. *Methods Enzymol.* **152**: 522–538.
- Archak, S., Meduri, E., Kumar, P.S., and Nagaraju, J., (2007). InSatDb: a microsatellite database of fully sequenced insect genomes. *Nucleic Acids Res.* **35**: D36–D39.
- Arenzberger, P., Megy, K., Waterhouse, R.M., Abrudan, J., Amedeo, P., and Antelo, B., et al. (2010). Sequencing of *Culex quinquefasciatus* establishes a platform for mosquito comparative genomics. *Science* **330**: 86–90.
- Ashburner, M., (2000). A biologist’s view of the *Drosophila* genome annotation assessment project. *Genome Res.* **10**: 392–393.
- Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., and Cherry, J.M., et al. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25**: 25–29.
- Attwood, T.K., (2000). The babel of bioinformatics. *Science* **290**: 471–472.
- Bairoch, A., and Apweiler, R., (2000). The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res.* **28**: 45–48.
- Balter, M., (1999). Gene sequencers target malaria mosquito. *Science* **285**: 508–509.
- Balter, M., (2001). Sequencing set for dreaded mosquito. *Science* **291**: 1873.

- Banerjee-Basu, S., Ryan, J.F., and Baxevanis, A.D., (2000). The homeodomain resource: a prototype database for a large protein family. *Nucleic Acids Res.* **28**: 329–330.
- Barker, W.C., Garavelli, J.S., Huang, H., McGarvey, P.B., Orcutt, B.C., and Srinivasarao, G.Y., et al. (2000). The protein information resource (PIR). *Nucleic Acids Res.* **28**: 41–44.
- Barnes, W.M., (1987). Sequencing DNA with dideoxyribonucleotides as chain terminators: hints and strategies for big projects. *Methods Enzymol.* **152**: 538–556.
- Baxevanis, A.D., (2001). The molecular biology database collection: an updated compilation of biological database resources. *Nucleic Acids Res.* **29**: 1–10.
- Baxevanis, A.D., and Ouellette, B.F., Eds., (2001). *Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins*, 2nd Ed. Wiley Publishing, New York.
- Beloqui, A., Guazzaroni, M.E., Pazos, F., Vieites, J.M., Godoy, M., and Golyshina, O.V., et al. (2009). Reactome array: forging a link between metabolome and genome. *Science* **326**: 252–257.
- Bennett, S.T., Barnes, C., Cox, A., Davies, L., and Brown, C., (2005). Toward the 1,000 dollars human genome. *Pharmacogenomics* **6**: 373–382.
- Bentley, D.R., (2006). Whole-genome re-sequencing. *Curr. Opin. Genet. Dev.* **16**: 545–552.
- Bevan, I.S., Rapley, R., and Walker, M.R., (1992). Sequencing of PCR-amplified DNA. *Genome Res.* **1**: 222–228.
- Bonasio, R., Zhang, G., Ye, C., Mutti, N.S., Fang, X., and Qin, W., et al. (2010). Genomic comparison of the ants *Camponotus floridanus* and *Harpegnathos saltator*. *Science* **329**: 1068–1071.
- Bonetta, L., (2010). Interactome under construction. *Nature* **468**: 851–854.
- Bork, P., (2000). Powers and pitfalls in sequence analysis: the 70% hurdle. *Genome Res.* **10**: 398–400.
- Bradley, R.D., and Hillis, D.M., (1997). Recombinant DNA sequences generated by PCR amplification. *Mol. Biol. Evol.* **14**: 592–593.
- Brazma, A., and Vilo, J., (2000). Gene expression data analysis. *FEBS Lett.* **480**: 17–24.
- Brent, R., (2000). Genomic biology. *Cell* **100**: 169–183.
- Brosius, J., (1999). Genomes were forged by massive bombardments with retroelements and retrosequences. *Genetica* **107**: 209–238.
- Burke, G.R., and Moran, N.A., (2011). Responses of the pea aphid transcriptome to infection by facultative symbionts. *Insect Mol. Biol.* **20**: 357–365.
- Burley, S.K., (2000). An overview of structural genomics. *Nat. Struct. Biol. (Suppl.)* **10**: 932–934.
- Caceres, M., Ranz, J.M., Barbadilla, A., Long, M., and Ruiz, A., (1999). Generation of a widespread *Drosophila* inversion by a transposable element. *Science* **285**: 415–418.
- Callaway, E., (2011). Genome giant offers data service. *Nature* **475**: 436–437.
- Capy, P., Gasperi, G., Hiemont, C., and Bazin, C., (2000). Stress and transposable elements: co-evolution or useful parasites? *Heredity* **85**: 101–106.
- Celis, J.E., Kruhoffer, M., Gromova, I., Frederiksen, C., Ostergaard, M., and Thykjaer, T., et al. (2000). Gene expression profiling: monitoring transcription and translation products using DNA microarrays and proteomics. *FEBS Lett.* **480**: 2–16.
- Chain, P.S.G., Graham, D.V., Fulton, R.S., FitzGerald, M.G., Hostetler, J., and Muzny, D., et al. (2009). Genome project standards in a new era of sequencing. *Science* **326**: 236–237.
- Chan, Q.W., and Foster, L.J., (2008). Changes in protein expression during honey bee larval development. *Genome Biol.* **9**: R156.
- Chen, X., and Tompa, M., (2010). Comparative assessment of methods for aligning multiple genome sequences. *Nat. Biotechnol.* **28**: 567–572.
- Cherf, G.M., Lieberman, K.R., Rashid, H., Lam, C.E., Karplus, K., and Akeson, M., (2012). Automated forward and reverse ratcheting of DNA in a nanopore at 5-Angstrom precision. *Nat. Biotechnol.* **30**: 344–348.
- Chintapalli, V.R., Wang, J., and Dow, J.A.T., (2007). Using FlyAtlas to identify better *Drosophila melanogaster* models of human disease. *Nat. Genet.* **39**: 715–720.

- Cinkosky, M.J., Fickett, J.W., Gilna, P., and Burks, C., (1991). Electronic data publishing and GenBank. *Science* **252**: 1273–1277.
- Cornman, R.S., Schatz, M.C., Johnston, J.S., Chen, Y.P., Pettis, J., and Hunt, G., et al. (2010). Genomic survey of the ectoparasitic mite *Varroa destructor*, a major pest of the honey bee *Apis mellifera*. *BMC Genomics* **11**: 602.
- Cox-Foster, D.L., Conlan, S., Holmes, E.C., Palacios, G., Evans, J.D., and Moran, N.A., et al. (2007). A metagenomic survey of microbes in honey bee colony collapse disorder. *Science* **318**: 283–287.
- Davies, R.W., (1982). DNA sequencing. In: Rickwood, D., and Hames, B.D., Eds., *Gel Electrophoresis of Nucleic Acids*, IRL Press., Oxford, UK, pp. 117–172.
- Depra, M., Valente, V.L., Margis, R., and Loreto, E.L.S., (2009). The hobo transposon and hobo-related elements are expressed as developmental genes in *Drosophila*. *Gene* **448**: 57–63.
- Derrington, I.M., Butler, T.Z., Collins, M.D., Manrao, E., Pavlenok, M., and Niederweis, M., et al. (2010). Nanopore DNA sequencing with MspA. *Proc. Natl. Acad. Sci. USA* **107**: 16060–16065.
- Doroszuk, A., Jonker, M.J., Pul, N., Breit, T.M., and Zwaan, B.J., (2012). Transcriptome analysis of a long-lived natural *Drosophila* variant: a prominent role of stress- and reproduction-genes in lifespan extension. *BMC Genomics* **13**: 167.
- Dressman, D., Yan, H., Traverso, G., Kinzler, K.W., and Vogelstein, B., (2003). Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations. *Proc. Natl. Acad. Sci. USA* **100**: 8817–8822.
- Drosophila 12 Genomes Consortium, 2007. Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* **450**: 203–218.
- Dutt, M.J., and Lee, K.H., (2000). Proteomic analysis. *Curr. Opin. Biotechnol.* **11**: 176–179.
- Eickbush, T.H., (1997). Telomerase and retrotransposons: which came first? *Science* **277**: 911–912.
- Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., and Otto, G., et al. (2009). Real-time DNA sequencing from single polymerase molecules. *Science* **323**: 133–138.
- Eisenberg, D., Maracotte, E.M., Xenarios, I., and Yeates, T.O., (2000). Protein function in the post-genomic era. *Nature* **405**: 823–826.
- Fan, W., and Li, R., (2012). Test driving genome assemblers. *Nat. Biotechnol.* **30**: 330–331.
- Fedurco, M., Romieu, A., Williams, S., Lawrence, I., and Turcatti, G., (2006). BTA, a novel reagent for DNA attachment on glass and efficient generation of solid-phase amplified DNA colonies. *Nucleic Acids Res.* **34**: e22.
- Field, D., Sansone, S.A., Collis, A., Booth, T., Dukes, P., and Gregurick, S.K., et al. (2009). ‘Omics data sharing. *Science* **326**: 234–235.
- FlyBase Consortium, 1997. FlyBase: a *Drosophila* database. *Nucleic Acids Res.* **25**: 63–66.
- Fortini, M.E., Skupski, M.P., Boguski, M.S., and Hariharan, I.K., (2000). A survey of human disease gene counterparts in the *Drosophila* genome. *J. Cell Biol.* **150**: F23–F29.
- Gadau, J., Helmkampf, M., Nygaard, S., Roux, J., Simola, D.F., and Smith, C.R., et al. (2012). The genomic impact of 100 million years of social evolution in seven ant species. *Trends Genet.* **28**: 14–21.
- Galperin, M.Y., and Fernandez-Suarez, X.M., (2011). The 2012 Nucleic Acids Research Database issue and the online Molecular Biology DataBase Collection. *Nucleic Acids Res.* **40**: D1–D8.
- Geisow, M.J., (1998). Proteomics: one small step for a digital computer, one giant leap for humankind. *Nat. Biotechnol.* **16**: 206.
- Gene Ontology Consortium, (2001). Creating the gene ontology resource: design and implementation. *Genome Res.* **11**: 1425–1433.
- George, J.A., Traverse, K.L., DeBaryshe, P.G., Kelley, K.J., and Pardue, M.L., (2010). Evolution of diverse mechanisms for protecting chromosome ends by *Drosophila* TART telomere retrotransposons. *Proc. Natl. Acad. Sci. USA* **107**: 21052–21057.
- Ghose, T., (2012). Sons of Next Gen. *Scientist* **6**: 56–58.
- Gingeras, T.R., (2009). Implications of chimaeric non-co-linear transcripts. *Nature* **461**: 206–211.

- Godfray, H.C.J., (2010). The pea aphid genome. *Insect Mol. Biol.* **19** (Suppl. 2): 1–4.
- Goldsmith, M.R., and Marec, F., Eds., (2010). *Molecular Biology and Genetics of the Lepidoptera*, CRC Press, Boca Raton, FL.
- Gourbiere, S., Dorn, P., Tripet, F., and Dumonteil, E., (2012). Genetics and evolution of triatomines: from phylogeny to vector control. *Heredity* **108**: 190–202.
- Grbic, M., Van Leeuwen, T., Clark, R.M., Rombauts, S., Rouze, P., and Grbic, V., et al. (2011). The genome of *Tetranychus urticae* reveals herbivorous pest adaptation. *Nature* **479**: 487–492.
- Guruharsha, K.G., Rual, J.F., Zhai, B., Mintseris, J., Valdya, P., and Valdya, N., et al. (2011). A protein complex network of *Drosophila melanogaster*. *Cell* **147**: 690–703.
- Harris, T.E., Buzby, P.R., Babcock, H., Beer, E., Bowers, J., and Braslavsky, I., et al. (2008). Single-molecule DNA sequencing of a viral genome. *Science* **320**: 106–109.
- Hass, B.J., and Zody, M.C., (2010). Advancing RNA-Seq analysis. *Nature Biotech.* **28**: 421–423.
- Hawley, R.S., and Walker, M.Y., (2000). *Drosophila melanogaster*: a genome sequenced. *Genome Res.* **10**: 593–594.
- Hayden, E.C., (2009). Genome sequencing: the third generation. *Nature* **457**: 768–769.
- Heliconius Genome Consortium, 2012. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature* **487**: 94–98.
- Hodgman, C., French, A., and Westhead, D., (2009). *Instant Notes in Bioinformatics*, 2nd ed. Bios Scientific Publishers, Oxford, UK.
- Holt, R., Subramanian, G.M., Halpern, A., Sutton, G.G., Charlab, R., and Nusskern, D.R., et al. (2002). The genome sequence of the malaria mosquito *Anopheles gambiae*. *Science* **298**: 129–149.
- Honeybee Genome Sequencing Consortium, 2006. Insights into social insects from the genome of the honeybee *Apis mellifera*. *Nature* **443**: 931–949.
- Howe, C.J., and Ward, E.S., Eds., (1989). *A Practical Approach*, IRL Press, New York.
- Hoy, M.A., (2009). The predatory mite *Metaseiulus occidentalis*: mitey small and mitey large genomes. *Bioessays* **31**: 581–590.
- Huerta-Cepas, J., Marcet-Houben, M., Pignatelli, M., Moya, A., and Gabaldon, T., (2010). The pea aphid phylome: a complete catalogue of evolutionary histories and arthropod orthology and paralogy relationships for *Acyrtosiphon pisum* genes. *Insect Mol. Biol.* **19** (Suppl. 2): 13–21.
- Hugenholtz, P., and Tyson, G.W., (2008). Metagenomics. *Nature* **455**: 481–483.
- International Aphid Genomics Consortium, (2010). Genome sequence of the pea aphid *Acyrtosiphon pisum*. *PLoS Biol.* **8** (2): e1000313.
- International Human Genome Sequencing Consortium, (2001). Initial sequencing and analysis of the human genome. *Nature* **409**: 860–921.
- Iyer, M.K., and Chinaiyan, A.M., (2011). RNA-Seq unleashed. *Nat. Biotechnol.* **29**: 599–600.
- Jabbari, K., and Bernardi, G., (2000). The distribution of genes in the *Drosophila* genome. *Gene* **247**: 287–292.
- Janitz, M., Ed., 2008. *Next-Generation Genome Sequencing: Towards Personalized Medicine* Wiley Science. New York.
- Jeong, H., Tombor, B., Albert, R., Oltvai, Z.N., and Barabasi, A.L., (2000). The large-scale organization of metabolic networks. *Nature* **407**: 651–654.
- Jeyaprakash, A., and Hoy, M.A., (1995). Complete sequence of a *mariner* transposable element from the predatory mite *Metaseiulus occidentalis* isolated by an inverse PCR approach. *Insect Mol. Biol.* **4**: 31–39.
- Johnson, J.M., Edwards, S., Shoemaker, D., and Schadt, E.E., (2005). Dark matter in the genome: evidence of widespread transcription detected by microarray tiling experiments. *Trends Genet.* **21**: 93–102.
- Kafatos, F.C., Louis, C., Savakis, C., Glover, D.M., Ashburner, M., and Link, A.J., et al. (1991). Integrated maps of the *Drosophila* genome: progress and prospects. *Trends Genet.* **7**: 155–161.
- Kahn, S.D., (2011). On the future of genomic data. *Science* **331**: 728–729.

- Kaminker, J.S., Bergman, C.M., Kronmiller, B., Carlson, J., Svirskas, R., and Patel, S., et al. (2002). The transposable elements of the *Drosophila melanogaster* euchromatin: a genomics perspective. *Genome Biol.* **3**: 20. research0084.1-0084.
- Kapitonov, V.V., and Jurka, J., (2008). A universal classification of eukaryotic transposable elements implemented in Repbase. *Nat. Rev. Genet.* **9**: 411–412.
- Kidwell, M.G., and Lisch, D.R., (2001). Transposable elements, parasitic DNA and genome evolution. *Evolution* **55**: 1–24.
- Kirkness, E.F., Haas, B.J., Sun, W., Braig, H.R., Perotti, M.A., and Clark, J.M., et al. (2010). Genome sequences of the human body louse and its primary endosymbiont provide insights into the permanent parasitic lifestyle. *Proc. Natl. Acad. Sci. USA* **102**: 12168–12173.
- Kondrashov, A.S., (1999). Comparative genomics and evolutionary biology. *Curr. Opin. Genet. Dev.* **9**: 624–629.
- Kurowska, M., Daszkowska-Golec, A., Gruszka, D., Marzec, M., Szurman, M., and Szarejko, I., et al. (2011). TILLING—a shortcut in functional genomics. *J. Appl. Genet.* **52**: 371–390.
- Lanave, C., Liuni, S., Licciulli, F., and Attimonelli, M., (2000). Update of the AMmtDB: a database of multi-aligned Metazoa mitochondrial DNA sequences. *Nucleic Acids Res.* **28**: 153–154.
- Lawson, D., Arensburger, P., Atkinson, P., Besansky, N.J., Bruggner, R.V., and Butler, R., et al. (2009). VectorBase: a data resource for invertebrate vector genomics. *Nucleic Acids Res.* **37**: D583–587.
- Legeal, F., Shigenobu, S., Gauthier, J.P., Colbourne, J., Rispe, C., and Collin, O., et al. (2010). AphidBase: a centralized bioinformatics resource for annotation of the pea aphid genome. *Insect Mol. Biol.* **19** (Suppl. 2): 5–12.
- Levin, H.L., and Moran, J.V., (2011). Dynamic interactions between transposable elements and their hosts. *Nat. Rev. Genet.* **12**: 615–627.
- Liu, X.S., (2007). Getting started in tiling microarray analysis. *PLoS Compt. Biol.* **3** (10): e183.
- Lockhart, D.J., and Winzeler, E.A., (2000). Genomics, gene expression and DNA arrays. *Nature* **405**: 827–836.
- Mackay, T.F.C., Richards, S., Stone, E.A., Barbadilla, A., Ayroles, J.F., and Zhu, D., et al. (2012). The *Drosophila melanogaster* genetic reference panel. *Nature* **482**: 173–178.
- Manrao, E.A., Derrington, I.M., Laszlo, A.H., Langford, K.W., Hopper, M.K., and Gillgren, N., et al. (2012). Reading DNA at single-nucleotide resolution with a mutant MspA nanopore and phi29 DNA polymerase. *Nat. Biotechnol.* **20**: 349–354.
- Marco, D., Ed., (2011). *Metagenomics: Current Innovations and Future Trends*, Caister Academic Press, Norfolk, UK.
- Mardis, E.R., (2008a). Next-generation DNA sequencing methods. *Annu. Rev. Genom. Human Genet* **9**: 387–402.
- Mardis, E.R., (2008b). The impact of next-generation sequencing technology on genetics. *Trends Genet* **24**: 133–141.
- Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., and Bemben, L.A., et al. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**: 376–380.
- Martin, J.A., and Wang, Z., (2011). Next-generation transcriptome assembly. *Nat. Rev. Genet.* **12**: 671–682.
- Maxam, A.M., and Gilbert, W., (1977). A new method for sequencing DNA. *Proc. Natl. Acad. Sci. USA* **74**: 560–564.
- Maxam, A.M., and Gilbert, W., (1980). Sequencing end-labeled DNA with base-specific chemical cleavages. *Methods Enzymol.* **65**: 499–560.
- Meldrum, D., (2000a). Automation for genomics, part one: Preparation for sequencing. *Genome Res.* **10**: 1081–1092.
- Meldrum, D., (2000b). Automation for genomics, part two: Sequencers, microarrays and future trends. *Genome Res.* **10**: 1288–1303.

- Men, A.E., Wilson, P., Siemering, K., and Forrest, S., (2008). Sanger DNA sequencing. In: Janitz, M., Ed., *Next-Generation Genome Sequencing: Towards Personalized Medicine* Wiley-VCH. Weinheim, Germany, pp. 2–11.
- Merriam, J., Ashburner, M., Hartl, D.L., and Kafatos, F.C., (1991). Toward cloning and mapping the genome of *Drosophila*. *Science* **254**: 221–225.
- Metzker, M.L., (2010). Sequencing technologies—the next generation. *Nat. Rev. Genet.* **11**: 31–46.
- Mierendorf, R.C., and Pfeffer, D., (1987). Direct sequencing of denatured plasmid DNA. *Methods Enzymol.* **152**: 556–562.
- Miller, W.J., McDonald, J.F., and Pinsker, W., (1997). Molecular domestication of mobile elements. *Genetica* **100**: 261–270.
- Miller, J.R., Koren, S., and Sutton, G., (2010). Assembly algorithms for next-generation sequencing data. *Genomics* **95**: 315–327.
- Misra, S., Crosby, M.A., and Drysdale, R.A., (2000). Using *Drosophila* genome databases. In: Sullivan, W., Ashburner, M., and Hawley, R.S., Eds., *Drosophila Protocols*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 509–523.
- Mita, K., Morimyo, M., Okano, K., Koike, Y., Nohata, J., and Kawasaki, H., et al. (2003). The construction of an EST database for *Bombyx mori* and its application. *Proc. Natl. Acad. Sci. USA* **100**: 14121–14126.
- Mita, K., Kasahara, M., Sasaki, S., Nagayasu, Y., Yamada, T., and Kanamori, H., et al. (2004). The genome sequence of silkworm. *Bombyx mori*. *DNA Res.* **11**: 27–35.
- Mockler, T.C., and Ecker, J.R., (2005). Applications of DNA tiling arrays for whole-genome analysis. *Genomics* **85**: 1–15.
- modENCODE Consortium, 2010. Identification of functional elements and regulatory circuits by *Drosophila* modENCODE. *Science* **330**: 1787–1797.
- Moran, N.A., and Jarvik, T., (2010). Lateral transfer of genes from fungi underlies carotenoid production in aphids. *Science* **328**: 624–627.
- Morozova, O., and Marra, M.A., (2008). Applications of next-generation sequencing technologies in functional genomics. *Genomics* **92**: 255–264.
- Mount, D.W., (2004). *Bioinformatics. Sequence and Genome Analysis*, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Munoz-Torres, M.C., Reese, J.T., Childers, C.P., Bennett, A.K., Sundaram, J.P., and Childs, K.L., et al. (2011). Hymenoptera Genome Database: integrated community resources for insect species of the order Hymenoptera. *Nucleic Acids Res.* **39 (Suppl. 1)**: D658–D662.
- Murray, V., (1989). Improved double-stranded DNA sequencing using the linear polymerase chain reaction. *Nucleic Acids Res.* **17**: 8889.
- Nadeau, J.H., and Dudley, A.M., (2011). Systems genetics. *Science* **331**: 1015–1016.
- Nadeau, N.J., Whibley, A., Jones, R.T., Davey, J.W., Dasmahapatra, K.K., and Baxter, S.W., et al. (2012). Genomic islands of divergence in hybridizing *Heliconius* butterflies identified by large-scale targeted sequencing. *Philos. Trans. R. Soc. London B* **367**: 343–353.
- Nagaraj, I.H., Gasser, R.B., and Ranganathan, S., (2006). A hitchhiker's guide to expressed sequence tag (EST) analysis. *Brief. Bioinform.* **8**: 6–21.
- Nasonia Genome Working Group, (2010). Functional and evolutionary insights from the genomes of three parasitoid *Nasonia* species. *Science* **327**: 343–348.
- Nature, 2011. Genome giant offers data service. *Nature* **475**: 436–437.
- Nature Biotechnology, (2010). Gathering clouds and a sequencing storm. *Nat. Biotechnol.* **28**: 1.
- Nature Biotechnology, (2011). The sequence is dead: long live the genome. *Nat. Biotechnol.* **29**: 463.
- Nene, V., Wortman, J.R., Lawson, D., Haas, B., Kodira, C., and Tu, Z., et al. (2007). Genome sequence of *Aedes aegypti*, a major arbovirus vector. *Science* **316**: 1718–1723.
- Norvell, J.C., and Machalek, A.Z., (2000). Structural genomics programs at the US National Institute of General Medical Sciences. *Nat. Struct. Biol. (Suppl.)* **10**: 931.

- Nunes, F.M.F., Valente, V., Sousa, J.F., Cunha, M.A.V., Pinheiro, D.G., and Maia, R.M., et al. (2004). The use of open reading frame ESTs (ORESTES) for analysis of the honey bee transcriptome. *BMC Genomics* **5**: 84.
- Nygaard, S., Zhang, G., Schiott, M., Li, C., Wurm, Y., and Hu, H., et al. (2011). The genome of the leaf-cutting ant *Acromyrmex echinatior* suggests key adaptations to advanced social life and fungus farming. *Genome Res.* **21** (8): 1339–1348.
- Oliver, B., (2006). Tiling DNA microarrays for fly genome cartography. *Nat. Genet.* **38**: 1101.
- Ozsolak, F., and Milos, P.M., (2011). RNA sequencing: advances, challenges and opportunities. *Nat. Rev. Genet.* **12**: 87–98.
- Palsson, B., (2000). The challenges of in silico biology. *Nat. Biotechnol.* **18**: 1147–1150.
- Pandey, A., and Mann, M., (2000). Proteomics to study genes and genomes. *Nature* **405**: 837–846.
- Papanicolaou, A., Gebauer-Jung, S., Blaxter, M.L., McMilland, W.O., and Jiggins, C.D., (2008). ButterflyBase: a platform for lepidopteran genomics. *Nucleic Acids Res.* **36**: D582–587.
- Pareek, C.S., Smoczyński, R., and Tretyn, A., (2011). Sequencing technologies and genome sequencing. *J. Appl. Genet.* **52**: 413–435.
- Pennisi, E., (2000a). Fruit fly genome yields data and a validation. *Science* **287**: 1374.
- Pennisi, E., (2000b). Ideas fly at gene-finding jamboree. *Science* **287**: 2182–2184.
- Perier, R.C., Praz, V., Junier, T., Bonnard, C., and Bucher, P., (2000). The eukaryotic promoter database (EPD). *Nucleic Acids Res.* **28**: 302–303.
- Pettersson, E., Lundeberg, J., and Ahmadian, A., (2009). Generations of sequencing technologies. *Genomics* **93**: 105–111.
- Pevsner, J., (2009). *Bioinformatics and Functional Genomics.*, 2nd Ed. Wiley-Blackwell, Hoboken, NJ.
- Podolak, E., (2010). Sequencing’s new race. *BioTechniques* **48**: 105–111.
- Poelchau, M.F., Reynolds, J.A., Denlinger, D.L., Elsik, C.G., and Armbruster, P.A., (2011). A de novo transcriptome of the Asian tiger mosquito, *Aedes albopictus*, to identify candidate transcripts for diapause preparation. *BMC Genomics* **12**: 619.
- Pop, M., and Salzberg, S.L., (2008). Bioinformatics challenges of new sequencing technology. *Trends Genet.* **24**: 142–149.
- Potter, C.J., Turencchalk, G.S., and Xu, T., (2000). *Drosophila* in cancer research. *Trends Genet.* **16**: 33–39.
- Reese, M.G., Hartzell, G., Harris, N.L., Ohler, U., Abril, J.F., and Lewis, S.E., (2000). Genome annotation assessment in *Drosophila melanogaster*. *Genome Res.* **10**: 483–501.
- Roberts, R.J., and Macelis, D., (2000). REBASE – restriction enzymes and methylases. *Nucleic Acids Res.* **28**: 306–307.
- Robinson, G.E., Kackett, K.J., Purcell-Miramontes, M., Brown, S.J., Evans, J.D., and Goldsmith, M.R., et al. (2011). Creating a buzz about insect genomes. *Science* **331**: 1386.
- Robison, K., (2011). Semiconductors charge into sequencing. *Nat. Biotechnol.* **29**: 805–806.
- Rubin, G.M., and Lewis, E.B., (2000). A brief history of *Drosophila*’s contributions to genome research. *Science* **287**: 2216–2218.
- Rubin, G.M., Yandell, M.D., Wortman, J.R., Kiklos, G.L.G., Nelson, C.R., and Hariharan, I.K., et al. (2000). Comparative genomics of the eukaryotes. *Science* **287**: 2204–2215.
- Sambrook, J., and Russell, D.W., (2001). DNA sequencing, *In Molecular Cloning: A Laboratory Manual*, 3d ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, Chapter 12.
- Sambrook, J., Fritsch, E.F., and Maniatis, T., (1989). DNA sequencing, *In Molecular Cloning: A Laboratory Manual*, 2nd Ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, Chapter 13.
- Sanger, F., Nicklen, S., and Coulson, A.R., (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**: 5463–5467.
- Schaack, S., Gilbert, C., and Feschotte, C., (2010). Promiscuous DNA: horizontal transfer of transposable elements and why it matters for eukaryotic evolution. *Trends Ecol. Evol.* **25**: 537–546.

- Schadt, E.E., Turner, S., and Kasarskis, A., (2010). A window into third-generation sequencing. *Hum. Mol. Genet.* **19**: R227–R240.
- Schisler, N.J., and Palmer, J.D., (2000). The IDB and IEDB: intron sequence and evolution databases. *Nucleic Acids Res.* **28**: 181–184.
- Science, 2011. Sequence everything: BGI thinks big. *Science* **334**: 1038.
- Sealey, P.G., and Southern, E.M., (1982). Electrophoresis of DNA. In: Rickwood, D., and Hames, B.D., Eds., *Gel Electrophoresis of Nucleic Acids*, IRL Press, Oxford, UK, pp. 39–76.
- Severson, D.W., and Behura, S.K., (2012). Mosquito genomics: progress and challenges. *Annu. Rev. Entomol.* **57**: 143–166.
- Shapiro, J.A., (1999). Transposable elements as the key to a 21st century view of evolution. *Genetica* **107**: 171–179.
- Shendure, J., and Ji, H., (2008). Next-generation DNA sequencing. *Nat. Biotechnol.* **26**: 1135–1145.
- Shendure, J., Porreca, G.J., Reppas, N.B., Lin, X., McCutcheon, J.P., and Rosenbaum, A.M., et al. (2005). Accurate multiplex polony sequencing of an evolved bacterial genome. *Science* **309**: 1728–1732.
- Singh, N.D., Larracuente, A.M., Sackton, T.B., and Clark, A.G., (2009). Comparative genomics on the *Drosophila* phylogenetic tree. *Annu. Rev. Ecol. Evol. Syst.* **40**: 459–480.
- Smaglik, P., (2000). For my next trick. *Nature* **407**: 828–829.
- Smith, T., (2000). A new era. *Nat. Struct. Biol. (Suppl.)* **10**: 927.
- Smith, C.D., Zimin, A., Holt, C., Abouheif, E., Benton, R., and Cash, E., et al. (2011a). Draft genome of the globally widespread and invasive Argentine ant (*Linepithema humile*). *Proc. Natl. Acad. Sci. USA* **108**: 5673–5678.
- Smith, C.R., Smith, C.D., Robertson, H.M., Helmkampf, M., Zimin, A., and Yandell, M., et al. (2011b). Draft genome of the red harvester ant *Pogonomyrmex barbatus*. *Proc. Natl. Acad. Sci. USA* **108**: 5667–5672.
- Stark, A., Lin, M.F., Kheradpour, P., Pedersen, J.S., Parts, L., and Carlson, J.W., et al. (2007). Discovery of functional elements in 12 *Drosophila* genomes using evolutionary signatures. *Nature* **450**: 232–291.
- Stein, L., (2001). Genome annotation. *Nat. Rev. Genet.* **2**: 493–503.
- Stormo, G.D., (2000). Gene-finding approaches for eukaryotes. *Genome Res.* **10**: 394–397.
- Suen, G., Teiling, C., Li, L., Holt, C., Abouheif, E., and Bornberg-Bauer, E., et al. (2011). The genome sequence of the leaf-cutter ant *Atta cephalotes* reveals insights into its obligate symbiotic lifestyle. *PLoS Genet.* **7** (2): 31002007.
- Tabor, S., and Richardson, C.C., (1987). DNA sequence analyses with a modified bacteriophage T7 DNA polymerase. *Proc. Natl. Acad. Sci. USA* **84**: 4767–4771.
- Tautz, D., and Domazet-Loso, T., (2011). The evolutionary origin of orphan genes. *Nat. Rev. Genet.* **12**: 692–702.
- Terwilliger, T.C., (2000). Structural genomics in North America. *Nat. Struct. Biol. (Suppl.)* **10**: 935–939.
- Tribolium Genome Sequencing Consortium, 2008. The genome of the model beetle and pest *Tribolium castaneum*. *Nature* **452**: 949–955.
- Van Zee, J.P., Geraci, N.S., Guerrero, F.D., Wikle, S.K., Stuart, J.J., and Nene, V.M., et al. (2007). Tick genomics: The *Ixodes* genome project and beyond. *Int. J. Parasitol.* **37**: 1297–1305.
- Venter, J.C., Smith, H.O., and Hood, L., (1996). A new strategy for genome sequencing. *Nature* **381**: 364–366.
- Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., and Sutton, G.G., et al. (2001). The sequence of the human genome. *Science* **291**: 1304–1351.
- Vukmirovic, O.G., and Tilghman, S.M., (2000). Exploring genome space. *Nature* **405**: 820–822.
- Wang, J., Xia, Q., He, X., Dai, M., Ruan, J., and Chen, J., et al. (2005). SilkDB: a knowledgebase for silkworm biology and genomics. *Nucleic Acids Res.* **33**: D399–D402.

- Waterhouse, R.M., Wyder, S., and Zdobnov, E.M., (2008). The *Aedes aegypti* genome: a comparative perspective. *Insect Mol. Biol.* **17**: 1–8.
- Wheeler, D.L., Chappay, C., Lash, A.E., Leipe, D.D., Madden, T.L., and Schuler, G.D., et al. (2000). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **28**: 10–14.
- Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J.L., Capy, P., and Chalhoub, B., et al. (2007). A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* **8**: 973–982.
- Winkler, S., Schwabedissen, A., Backasch, D., Bokel, C., Seidel, C., and Bonisch, S., et al. (2005). Target-selected mutant screen by TILLING. *Drosophila Genome Res.* **15**: 718–723.
- Wurm, Y., Wang, J., Riba-Grognuz, O., Corona, M., Nygaard, S., and Hunt, B.G., et al. (2011). The genome of the fire ant *Solenopsis invicta*. *Proc. Natl. Acad. Sci. USA* **108**: 5679–5684.
- Xia, Q., Zhou, Z., Lu, C., Cheng, D., Dai, F., and Li, B., et al. (2004). A draft sequence for the genome of the domesticated silkworm (*Bombyx mori*). *Science* **306**: 1937–1940.
- Yandell, M., and Ence, D., (2012). A beginner’s guide to eukaryotic genome annotation. *Nat. Rev. Genet.* **13**: 329–341.
- Zdobnov, E.M., and Bork, P., (2007). Quantification of insect genome divergence. *Trends Genet.* **23**: 16–20.
- Zerbino, D.R., Paten, B., and Haussler, D., (2012). Integrating genomes. *Science* **336**: 179–182.
- Zhan, S., Merlin, C., Boore, J.L., and Reppert, S.M., (2011). The monarch butterfly genome yields insights into long-distance migration. *Cell* **147**: 1171–1185.
- Zongyuan, M., Yu, J., and Kang, L., (2006). LocustDB: a relational database for the transcriptome and biology of the migratory locust (*Locusta migratoria*). *BMC Genomics* **7**: 11.

# DNA Amplification by the Polymerase Chain Reaction: Molecular Biology Made Accessible

## Chapter Outline

- 8.1 Overview 308**
- 8.2 Introduction 309**
- 8.3 The Basic Polymerase Chain Reaction (PCR) 312**
  - 8.3.1 The First Few Cycles are Critical 312
  - 8.3.2 PCR Power 313
  - 8.3.3 Standard PCR Protocols 313
  - 8.3.4 DNA Polymerases 313
  - 8.3.5 Other Thermostable DNA Polymerases 315
  - 8.3.6 Primers are Primary 315
  - 8.3.7 Storing Insects for the PCR 318
  - 8.3.8 Preparing DNA Samples 319
  - 8.3.9 PCR Automation 324
  - 8.3.10 Specificity of the PCR 324
  - 8.3.11 Detecting Primer Artifacts 326
  - 8.3.12 How Many Cycles Does a PCR Need? 326
  - 8.3.13 Reducing the Evils of Contamination 327
- 8.4 Some Modifications of the PCR 329**
  - 8.4.1 AFLP for DNA Fingerprinting 329
  - 8.4.2 Anchored PCR 333
  - 8.4.3 Arbitrary Primers 333
  - 8.4.4 Asymmetric PCR 334
  - 8.4.5 Degenerate Primers 334
  - 8.4.6 Hot-Start PCR 335
  - 8.4.7 Inverse PCR 335
  - 8.4.8 Long PCR or High-Fidelity PCR 336
  - 8.4.9 Multiplex PCR 339
  - 8.4.10 Nested PCR 341
  - 8.4.11 PCR-RFLP 341
  - 8.4.12 Quantitative PCR 342

- 8.4.13 Random Primers 342
- 8.4.14 Real-Time PCR 345
- 8.4.15 Reverse-Transcription PCR 346
- 8.4.16 TaqMan PCR 346
- 8.4.17 Digital PCR 346

**8.5 Some Research Applications 347**

- 8.5.1 Amplifying Ancient DNA 347
- 8.5.2 Amplifying Old DNA 350
- 8.5.3 Amplifying RNA 351
- 8.5.4 Analysis of mRNA Polyadenylation 352
- 8.5.5 Cloning a Gene 352
- 8.5.6 Detecting Gene Amplification 354
- 8.5.7 Detecting Methylation of DNA 354
- 8.5.8 Detecting Pathogens in Vector Arthropods 354
- 8.5.9 Detecting Pesticide Resistance 355
- 8.5.10 Developmental Biology 356
- 8.5.11 Engineering DNA 356
- 8.5.12 Evaluate Efficacy of Disease Control 357
- 8.5.13 Evolutionary Analyses 357
- 8.5.14 Sequencing DNA 358

**8.6 Multiple Displacement Amplification: Another Method to Amplify DNA 359****8.7 Concluding Remarks 359****References Cited 361**

## 8.1 Overview

Occasionally, a technique is developed that changes the kinds of questions that can be answered in biology. Within a few short years, the polymerase chain reaction (PCR) became just such a powerful tool for solving a myriad of basic and applied problems. Modifications of the PCR continue to be developed and these permit additional applications.

The PCR is a method for amplifying (copying) small amounts of DNA or RNA. It can be used to isolate specific DNA fragments, end label DNA, clone cDNA and genomic DNA, sequence DNA, mutate specific DNA sequences, alter promoters, quantitate the amount of RNA or DNA, and identify molecular markers for taxonomic or ecological studies. The PCR requires a DNA polymerase, dNTPs, template DNA, and primers. Information about sequences at each end of the DNA to be amplified is needed in order to synthesize appropriate primers for “standard” (allele-specific) PCR. When two specific primers are used, amplification of DNA theoretically is geometric, producing large quantities of DNA suitable for sequencing, cloning, or probing. PCR methods that use single primers, such as random amplified polymorphic DNA (RAPD)-PCR, also can result in DNA

amplification. RAPD-PCR uses short, randomly chosen primers to amplify multiple DNA segments in the genome. The resulting banding patterns (similar to bar codes) provide information about genetic variation within the entire genome of insects.

The power of the PCR to amplify DNA is dramatic; theoretically even a single molecule can be amplified, although efficiency usually is lower. This power creates formidable problems with contamination and requires careful organization of PCR experiments and the use of adequate controls. However, because the PCR is relatively easy, novices in molecular biology can use it to study molecular systematics, evolution, ecology, behavior, and development.

## 8.2 Introduction

The **polymerase chain reaction (PCR)** is an *in vitro* or cell-free method for synthesizing DNA sequences in nearly any amount required starting with a small initial quantity. The PCR is one of the most accessible and versatile techniques available to entomologists interested in both basic and applied problems. The PCR is powerful because it can be used to isolate specific DNA fragments, end label DNA, mutagenize specific DNA fragments, clone cDNA and genomic DNA, sequence DNA, quantitate RNA and DNA, and alter a variety of sequences to study gene expression.

DNA polymerase, as used in the PCR, was designated the “Molecule of the Year” by *Science* in 1989 ([Guyer and Koshland 1989](#)). The PCR became a standard laboratory method in an extraordinarily short time after it was invented in 1985 ([Mullis 1987](#), [Rabinow 1996](#)). In 1993, Kary Mullis received the Nobel Prize for chemistry for his work on the PCR, although some believe other scientists should have shared credit for the invention ([Rabinow 1996](#)).

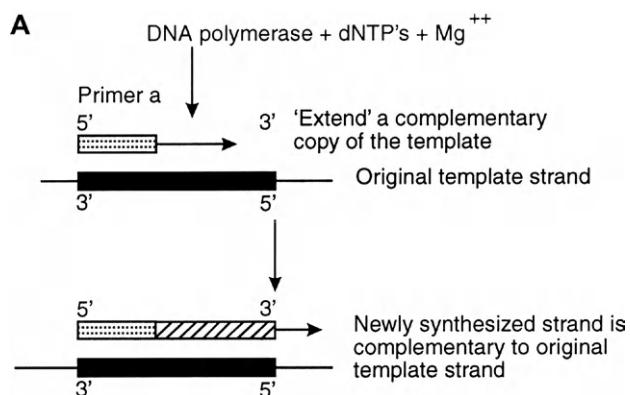
Since 1985, improvements in, and optimization of, the PCR have led to its use by numerous scientists ([Dieffenbach and Dveksler 2003](#), [Park 2010](#), [Kennedy and Oswald 2011](#)). The PCR has become a common procedure in forensics and diagnostics, and it is revolutionizing studies of basic biology, ecology, systematics, and evolution. This relatively simple technique has provided scientists with limited experience in molecular biology the opportunity to apply molecular techniques to diverse problems ([White et al. 1989](#), [Arnheim et al. 1990](#), [Erlich and Arnheim 1992](#), [King 2010](#)).

Although the PCR is conceptually simple, the process is, in fact, not completely understood. The PCR involves complex kinetic interactions between the template (or target) DNA, product DNA, oligonucleotide primers (polymers

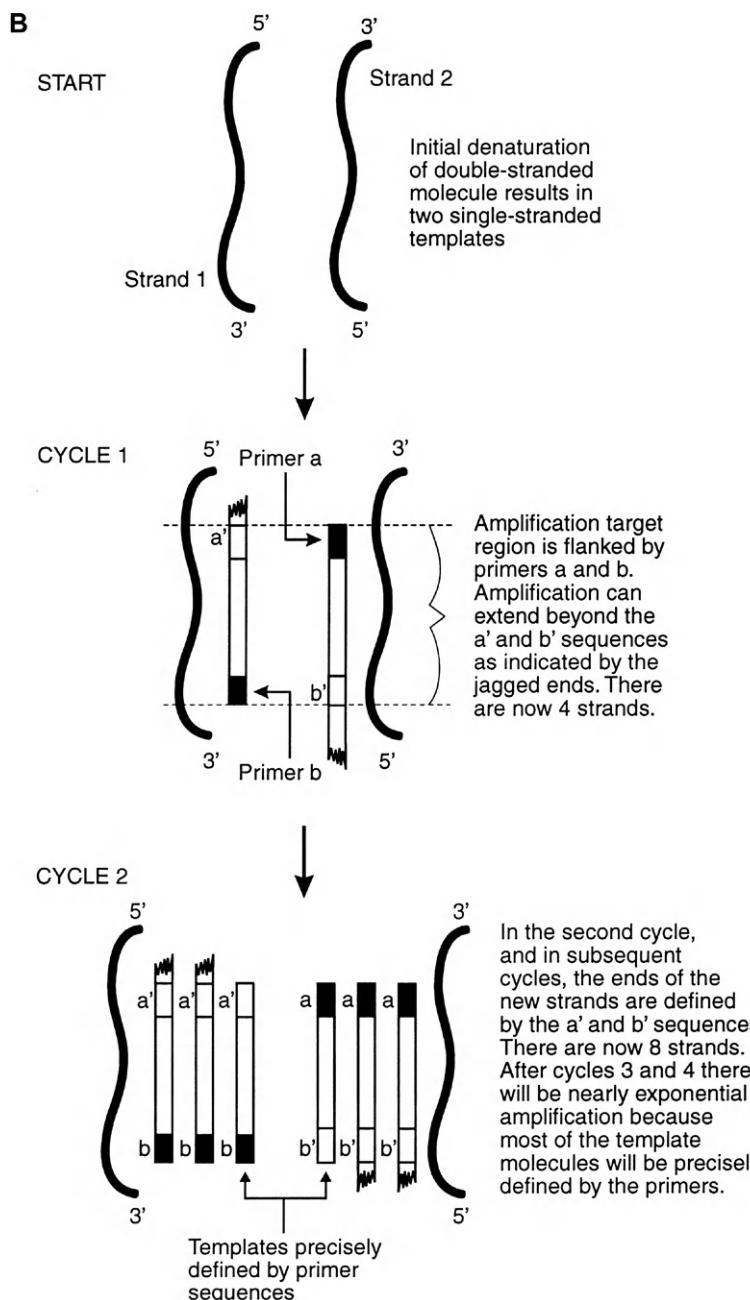
of 10 to  $\approx$ 30 nucleotides), deoxynucleotide triphosphates (dNTPs), buffer, and enzyme (one or more DNA polymerases). These relationships change during the course of the reaction (Figure 8.1).

The PCR works well with most DNA targets, but adjustments may be needed in the reaction parameters in order to improve specificity and yield. Parameters to be modified to optimize the PCR include: the reaction buffer (particularly the  $MgCl_2$  concentration); relative concentrations of template DNA, primers, dNTPs and DNA polymerase; annealing time and temperature; and extension time and temperature (Carbonari et al. 1993). No single protocol is appropriate for all situations and each new experiment requires optimization. For example, amplifying a 100-bp fragment is not equivalent to amplifying a 10-kb DNA fragment. Modifying each of the components of the PCR to develop an optimized reaction can be time consuming and tedious, but kits now are available that allow one to optimize the PCR with fewer steps.

This chapter describes what we know about the basic PCR, discusses some of the modifications of the basic method, identifies applications of the PCR, and provides references to additional information. PCR technology changes and new applications and methods of significance to entomologists will continue to become available.



**Figure 8.1** The standard, allele-specific PCR protocol. A) Template DNA is isolated and mixed with primers, dNTPs and *Taq* DNA polymerase in a buffer with  $Mg^{2+}$ . The double-stranded template DNA is heated to denature it so that the primer can anneal to single-stranded target DNA (only one template stand is shown). *Taq* synthesizes a new single strand of complementary DNA using the primer to initiate synthesis. The dNTPs are added in a sequence determined by the template DNA strand. This initial extension continues on beyond the desired end, as shown by the arrow on the newly synthesized strand in the lower diagram. The process of denaturation, annealing of the primer to the template, and DNA synthesis (or extension) is called a cycle.



**Figure 8.1 (Continued) B)** The process of amplification has low efficiency in the early cycles and some of the products produced lack defined ends. At the start, a single molecule (two strands) is denatured by heating and primer a and b, respectively, anneal to homologous sequences on strand 2 and 1. As shown above, amplification in cycle 1 extends some distance along the original DNA template strands. During cycle 2, the DNA is once again heated to denature it, and old and new strands can serve as templates for DNA synthesis. Note the DNA is synthesized from the 5' to the 3' direction. The primers are indicated by the small rectangles; the sequences homologous to the primers are designated at a' and b'. Once again in cycle 2, extension of some new strands occurs beyond the a' or b' sequences. However, for the first time templates are produced with precisely defined ends. After the next couple of cycles (cycles 3 and 4), most of the templates will be precisely defined and it is only then that amplification becomes nearly geometric.

## 8.3 The Basic Polymerase Chain Reaction (PCR)

The PCR involves combining a DNA sample (the template) with oligonucleotide primers, deoxynucleotide triphosphates (dNTPs), and a DNA polymerase in a buffer. The specificity of the basic PCR depends on complementary base pairing by the two primers to the template (target) DNA (Figure 8.1A).

The **primers**, single-stranded (ss) sequences that flank the DNA to be amplified, anneal to the template DNA that has been denatured by heating it so that it is single-stranded. Repeated PCR cycles involve heat denaturation to separate the template DNA strands, cooling to allow annealing of primers to the complementary DNA sequences of the ss template DNA, and “extension” (or replication) of new (product) DNA strands by DNA polymerase. The base sequence of the new strand is determined by the sequence of the ss template DNA. DNA synthesis proceeds across the region *between* the annealed primers (Figure 8.1B). This mixture is repetitively heated and cooled until the desired amount of template DNA has been amplified, usually after 25–30 cycles.

### 8.3.1 The First Few Cycles are Critical

All cycles begin by denaturing the template DNA (and any previously synthesized product) by heating it so that the template DNA and newly synthesized DNA become single-stranded (Figure 8.1A). As the temperature is lowered, the primers anneal to the complementary sequences of the single strands of DNA. The annealing step in the early cycles requires the primers to “scan” the template DNA for the correct target sequences to which they can anneal (Ruano et al. 1991, Dieffenbach 1995, Harris and Jones 1997). Because much of the template DNA will not have the correct complementary sequence, annealing during the early cycles may not be as efficient as it is during the middle cycles. Improper interactions of primers with template in the first few cycles can lead to nonspecific products and reduced yields. The PCR product will be specific *only* if the two primers bind to sites on the complementary strand of the DNA and, for the standard PCR, these sites are not more than  $\approx$ 1–2 kb apart. Thus, the first few cycles are very important if accurate, and high, yields are to be produced.

During the middle cycles, the DNA product previously synthesized is the *preferred* template for the primers, so this target template is perfectly demarcated (Figure 8.1B). Finally, in the late cycles, denatured amplified products that are present in high concentration can hybridize to themselves, thus blocking the primers from their complementary sites. In theory, DNA sequences up to  $\approx$ 10 kb in length can be synthesized by the standard protocol, but sequences of 2 kb, or less, are more readily obtained.

### 8.3.2 PCR Power

The power of the PCR is based on the fact that the products of one replication cycle serve as a template for the next. Each successive cycle, in theory, doubles the number of DNA molecules synthesized in the previous cycle, resulting in the exponential accumulation of the target DNA at approximately  $2^n$ , where  $n$  is the number of cycles. In practice, the PCR is never 100% efficient and less product will be produced. The early cycles are less efficient than the middle cycles because precisely defined template strands occur only after the first few cycles (Figure 8.1B). Late in the PCR, the reduced availability of reaction components may limit the yield. Reaction components that are reduced include primer concentration, dNTPs, or DNA polymerase (Czerny 1996). Primers and DNA polymerase can be degraded by the multiple heating cycles and dNTPs can be used up or degraded.

### 8.3.3 Standard PCR Protocols

Table 8.1 gives a procedure suitable for amplifying genomic DNA from *Drosophila* and demonstrates the relative simplicity of the technique. Table 8.2 discusses some of the issues that must be considered in setting up new PCRs.

The PCR is performed using commercially available temperature cyclers that allow the programming of the three fundamental reaction temperatures during denaturing, annealing, and extension (Figure 8.2). A typical amplification cycle involves denaturing the template DNA at 94 °C for 20 seconds, annealing the primers to the template at 55 °C for 20 seconds, and extending (or synthesizing) the DNA at 72 °C for 30 seconds. Because the instruments require time to heat and cool to a specific temperature, each actual cycle time may require 10 minutes or more, depending upon the specific machine used. If 25 cycles are performed, the total time will be ≈4 hours and the target DNA will have been amplified ≈1 million-fold, assuming a doubling has occurred in each cycle.

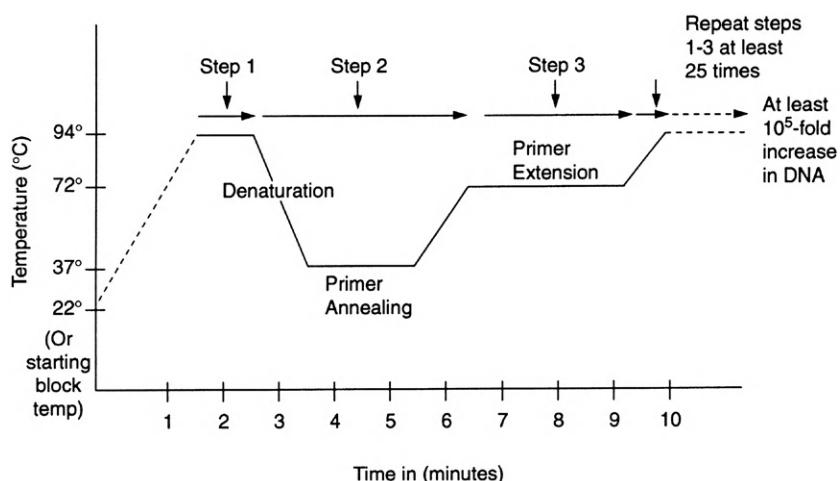
### 8.3.4 DNA Polymerases

The PCR, as first described in 1985, used the Klenow fragment of *Escherichia coli* DNA polymerase I to produce copies of target DNA (Saiki et al. 1985, Mullis and Faloona 1987). Because the Klenow fragment is heat sensitive, fresh enzyme had to be added to each cycle, making the PCR an exceedingly tedious procedure!

The efficiency and fidelity of the PCR was dramatically improved by using a heat-resistant polymerase (*Taq* DNA polymerase) so that the procedure could be carried out at high temperatures without having to add new enzyme before

**Table 8.1: Example of a Standard Allele-Specific PCR Reaction Protocol for Amplifying *Drosophila* DNA.**

1. Set up a 100- $\mu$ l reaction in a 0.5-ml microfuge tube, mix, and overlay with 75 $\mu$ l of mineral oil:
Template DNA ( $10^5$ – $10^6$ target molecules)
20 pmol each primer (each primer 18–30 nucleotides long)
100 mM Tris-HCl (pH 8.3 at 20°C)
10 mM MgCl <sub>2</sub>
0.05% Tween 20
50 $\mu$ M each dNTP
2 units of <i>Taq</i> DNA polymerase
2. Perform 25–35 cycles of PCR using the following temperature profile:
Denaturation    96 °C, 15 seconds
Primer annealing                                        55 °C, 30 seconds
Primer extension                                        72 °C, 1.5 minutes
3. Cycling should conclude with a final extension at 72 °C for 5 minutes. Stop reactions by chilling to 4 °C and/or adding ethylene dinitriilotetra-acetic acid (EDTA), a chelating agent to 10 mM.



**Figure 8.2** Example of a typical PCR protocol. Step 1 involves denaturating the double-stranded DNA template at 94 °C. Step 2 involves annealing the primer to the single-stranded target DNA by base pairing. Step 3 involves synthesis of new DNA from the 3' end of the primer (=primer extension) by DNA polymerase using dNTPs in a sequence determined by the template DNA. Steps 1–3 are a cycle, and approximately 25 cycles will yield an increase in DNA content by a factor of  $\approx 1$  million ( $2^{25}$ ).

each cycle (Saiki et al. 1988, Eckert and Kunkel 1991, Taylor 1991, Goodman 1995). *Taq* was isolated from the bacterium *Thermus aquaticus* that was collected from a hot spring in Yellowstone National Park. Because *Taq* can withstand repeated exposures to temperatures up to 94 °C, its use greatly increased the ease with which the PCR could be performed.

*Taq* DNA polymerase is a 94-kDa protein with a temperature optimum of  $\approx$ 75 to 80 °C. It can extend (add on)  $>$ 60 nucleotides per second at 70 °C with a GC-rich 30-mer primer. In a PCR mixture, *Taq* retains 50% of its activity after  $\approx$ 40 minutes at 94 °C. The use of *Taq* thus increased the specificity and yield of the PCR over that possible with the Klenow fragment because the primers could be annealed and extended at higher temperatures, which eliminated much of the *nonspecific* amplification. Longer PCR products were produced because the secondary structure of the template DNA was eliminated at these higher temperatures, as well. Fragments of  $\approx$ 500 bp can be synthesized with the Klenow fragment, but fragments up to 10 kb sometimes can be produced with *Taq*.

### 8.3.5 Other Thermostable DNA Polymerases

Genetically engineered variants of *Taq* have been developed and DNA polymerases from other sources now are available commercially in native and cloned form (Engelke et al. 1990, Erlich et al. 1991, Perler et al. 1996). For example, a recombinant *T. aquaticus* polymerase called the “Stoffel fragment” persists at 97.5 °C and exhibits optimal activity over a broad range of Mg<sup>2+</sup> concentrations.

Other thermostable DNA polymerases have been isolated from Eubacteria (*Thermus* and *Bacillus*) and Archaea (*Thermococcus*, *Pyrococcus*, and *Sulfolobus*). Many are commercially available and have specific attributes such as 3' to 5' exonuclease activity, “proofreading” ability, different molecular weights, different stabilities and temperature optima (Perler et al. 1996). For example, a thermostable enzyme isolated from *T. thermophilus* can reverse-transcribe RNA efficiently at high temperatures. The thermostability of this enzyme appears to minimize the importance of secondary structure in the RNA template and allow efficient cDNA synthesis at high temperatures.

A DNA polymerase isolated from the archaebacterium *S. acidocaldarius* carries out polymerization at 100 °C, which facilitates amplification of DNA regions with secondary structure (Arnheim and Erlich 1992). Polymerases from *Thermoplasma acidophilum*, *Thermococcus litoralis*, and *Methanobacterium thermoautotrophicum* have 3' to 5' exonuclease activities, which means that they can proofread, reducing the rate of misincorporation or errors.

### 8.3.6 Primers are Primary

Although all the components of a PCR are important, primers are truly crucial (Table 8.2). Well-designed primers can result in 100- to 1000-fold increases in sensitivity (He et al. 1994a,b). As a rule, longer primers are better than shorter for increasing specificity, but length is not the only consideration (Yuryev 2007).

**Table 8.2: Optimizing Standard PCR Protocols Involves Optimizing Reaction Components.**

PCR component	Issues to consider
Primers	<ol style="list-style-type: none"> <li>Select primers with a random base distribution and GC content similar to template DNA being amplified.</li> <li>Avoid primers with stretches of polypurines and polypyrimidines or other unusual sequences.</li> <li>Check primers for complementarity and avoid primers with 3' overlaps to reduce primer-dimer artifacts.</li> <li>Construct primers 20–30 nucleotides long.</li> <li>Optimize the amount of primers used.</li> <li>Design so the base at the 3' end of the primer is a G or C to enhance specificity (G-C clamp).</li> </ol>
Template DNA	<ol style="list-style-type: none"> <li>Template DNA should be free of proteases that could degrade the DNA polymerase.</li> <li>Template DNA with high levels of proteins or salts should be diluted or cleaned up to reduce inhibition of DNA polymerase activity.</li> <li>Highly concentrated template DNA may yield nonspecific product or inhibit the reaction.</li> <li>It is rare that template DNA concentration is too low.</li> </ol>
PCR buffer	<ol style="list-style-type: none"> <li>MgCl<sub>2</sub> concentration is very important.</li> <li>Excess Mg<sup>2+</sup> promotes production of nonspecific product and primer-dimer artifacts.</li> <li>Insufficient Mg<sup>2+</sup> reduces yields.</li> <li>Presence of EDTA or other chelators can reduce the availability of Mg<sup>2+</sup>.</li> </ol>
Taq polymerase	<ol style="list-style-type: none"> <li>Excessive Taq concentrations can yield nonspecific products and reduce yield. Recommended concentrations are between 0.5 and 2.5 units per 100-ml reaction. Add the Taq at 94 °C and mix thoroughly.</li> <li>Stringency can be increased by increasing the annealing temperature, adjusting dNTP concentrations, and minimizing incubation time.</li> </ol>
dNTPs	<ol style="list-style-type: none"> <li>dNTP concentrations should be equivalent to minimize misincorporation errors.</li> <li>Low dNTP concentrations minimize mispriming, but if too low, can reduce the amount of product.</li> </ol>
Cycle parameters	
Incubation	<ol style="list-style-type: none"> <li>Time varies with length of target being amplified; 1 minute/kb is average.</li> <li>Ramp time (time to change from one temperature to another) should be minimized to improve specificity.</li> <li>Insufficient step is a common problem; 94 °C results in complete separation, but excess time can cause denaturation of Taq polymerase.</li> </ol>

(Continued)

**Table 8.2: (Continued)**

PCR component	Issues to consider
Annealing	<ol style="list-style-type: none"> <li>Annealing temperature depends on length and GC content of primers; 55 °C is good for primers 20 nt long (50% GC).</li> <li>Higher annealing temperatures may be needed to increase primer specificity.</li> </ol>
Cycle number	<ol style="list-style-type: none"> <li>Optimum number varies with starting concentration of template DNA, and all of the above-mentioned parameters.</li> <li>Too many cycles increases amount of nonspecific product, whereas too few results in a low yield that can't be detected by gel electrophoresis.</li> <li>If additional product is required, it is better to reamplify, using an aliquot of the first reaction as the template, than to increase the number of cycles.</li> </ol>

Many of the modifications of the PCR have involved modifying the number, size, and specificity of the primers used, as described in Section 8.4.

What is a primer and where do you get them? A primer is a short (10 to  $\approx$ 30 nucleotides) single-stranded polymer of oligonucleotides. The standard (allele-specific) PCR requires that the specific sequence of the DNA targeted for amplification be known in order to synthesize a pair of primers. Thus, information is required about the gene/DNA to be amplified. Primers anneal to the target DNA by complementary base pairing, with A annealing to T, and C to G. The primers determine the length, specificity, and nature of the DNA fragment amplified.

**Allele-specific (standard) PCR** requires a pair of primers to flank the target DNA to be amplified; extension (copying of the single strand of template DNA) occurs from each 3'-OH end of the primer, so that the 5' ends of the primers define the ends of the amplified DNA. As a result, the length of the DNA generated is equal to the lengths of the two primers plus the length of the template DNA ([Figure 8.1](#)).

Most primers are synthesized to order on a DNA synthesizer. Many commercial suppliers provide this service, with the price determined by the number of bases in the primers. Primers may be called 10-, 20- or 30-mers, based on their length. Primers can be constructed that contain extensions so that restriction enzyme sites, regulatory codons, or labels can be added to the target DNA. These sequences will be incorporated into the 5' end of the target sequence, making the products more easily cloned or sequenced.

Selecting primers for allele-specific PCR remains somewhat empirical, although computer programs have been developed to aid in their design (Table 8.2) (Lexa et al. 2001). It is desirable, where possible, to select primer pairs with a G + C content of  $\approx$ 50% and a random base distribution (except at the 3' end). It is important to avoid complementary 3' ends of the primer pairs to avoid primer-dimer artifacts that will reduce the yield of the desired DNA.

Runs of three or more Cs or Gs at the 3' ends of primers may promote mispriming at G + C-rich regions. Primers with T, C, or G as the 3' (last) nucleotide results in a more specific PCR product than if the primer ends in an A (Ayyadevara et al. 2000). Amplification efficiency is reduced when T and A occupy the penultimate 3' position of the primer (Ayyadevara et al. 2000). Palindromic sequences within primers should be avoided. Sequences that will yield a significant secondary structure should be avoided. In some cases, primers with two Gs and/or Cs at the 3' end ("G/C clamp") will ensure the primer anneals strongly to the template to promote specific priming (Roux 1995).

Sometimes, suboptimal primers, perhaps containing high amounts of A and T, must be used due to the nature of the target sequence. Low concentrations of tetramethylammonium chloride (TMAC) could reduce mispriming and thus reduce nonspecific amplification (Chevet et al. 1995).

### 8.3.7 Storing Insects for the PCR

Ideal killing and storage techniques include placing the insects into an ultralow freezer ( $-80^{\circ}\text{C}$ ) or into liquid nitrogen, or dry ice. Rapid killing reduces damage to DNA by endogenous DNases. Storage of insects under inappropriate conditions can have detrimental effects on the quality and quantity of DNA available for the PCR (Dick et al. 1993). However, it is not always possible to kill and store insects under optimal conditions in remote field sites.

Alternative killing and storage methods include the use of ethanol (EtOH) at 95 or 100% (Quicke et al. 1999). The use of EtOH at <95% is undesirable because the water in insects dilutes the EtOH, which can result in degradation of DNA. If your insects are large, it may be desirable to kill them in 100% EtOH, pour it off, and replace it with fresh to reduce the dilution with endogenous water. Storage in methanol and chloroform, as well as low concentrations of EtOH, can result in poor preservation of DNA (Fukatsu 1999).

Other killing and storage methods may provide useful DNA, although loss in quality and quantity may occur. Desiccation with silica gel has preserved DNA of tiger beetles for several months (Vogler and Pearson 1996).

Other preservation methods include acetone, 2-propanol, diethyl ether, and ethyl acetate, which allow insects to be stored for  $\approx$ 6 months ([Fukatsu 1999](#)). [Fukatsu \(1999\)](#) recommended the use of acetone, which preserved the DNA of insects (as well as the DNA of microbial organisms within the insects) for >2 years at room temperature. Detecting microbial symbionts within stored insects can offer a challenge because symbionts may be low in titer; PCR-inhibiting substances may be present in the insect gut. [Zaspel and Hoy \(2008\)](#) compared the titer of *Wolbachia* in specimens of *Ephestia kuehniella* stored in 95% EtOH or 100% acetone and held at room temperature or at  $-80^{\circ}\text{C}$  for up to 101 weeks using high-fidelity and standard PCR protocols and real-time quantitative PCR methods. The least-effective storage method was the 95% EtOH and room-temperature protocol, but all protocols resulted in the detection of *Wolbachia* over the 2-year period.

Critical-point drying or other drying techniques are used to preserve many small insects in museums ([Austin and Dillon 1997](#)). Such dried insects sometimes can be used for molecular studies, especially if the target DNA is short and abundant (such as mitochondrial and ribosomal DNA). Specimens that were killed in 100% EtOH, stored at  $5^{\circ}\text{C}$ , and then dried yielded good quality DNA upon extraction ([Austin and Dillon 1997](#)). Amplification of long segments of single-copy genes from insects that have been poorly preserved is likely to be difficult.

### 8.3.8 Preparing DNA Samples

Template DNA used in the PCR generally should be free of proteases that could degrade the DNA polymerase. It should be free of nucleases that could degrade DNA, and free of DNA-binding proteins or high levels of heat-precipitable proteins that would inhibit amplification. Ideally,  $10^5$  to  $10^6$  template DNA molecules are available, although successful PCRs have been achieved with only a few DNA molecules.

Relatively crude DNA preparations *sometimes* can be used for the PCR. For example, when large numbers of individual insects must be processed, it is possible to do PCR on undissected larval or adult insects without prior isolation of the DNA ([Grevelding et al. 1996](#)). Apparently, the repeated denaturation steps at high temperature are sufficient to lyse cells so that sufficient template DNA is available, especially when the target DNA is present in high copy number in each cell (such as mitochondrial genes or ribosomal RNA genes). The advantage of using a crude lysate is that it reduces the time and costs to prepare the sample, which is important when hundreds or thousands of specimens must be evaluated. Such crude preparations do not allow the DNA to be stored.

Another crude preparation method involves boiling the insect with subsequent dilution. Lysing cells in boiling water is a quick and effective method of preparing DNA for the PCR, although only a small volume of the extract can be used because cellular debris may inhibit the PCR.

Cells in complex biological fluids or cells resistant to lysis require additional processing. Rapid and inexpensive DNA extractions can be achieved using Chelex®, a polyvalent chelating agent in resin form, which reduces degradation of DNA heated in low-ionic-strength buffers, probably by chelating heavy metal ions that may serve as catalysts in the breakdown of DNA (Singer-Sam et al. 1989). Adding Chelex during boiling appears to increase the amount of DNA produced from samples containing small amounts of template. Chelex is nontoxic, provides rapid results, and can be used to isolate DNA from hundreds of individual insects or mites suitable for either the standard PCR or RAPD-PCR (Edwards and Hoy 1994) (Table 8.3). A disadvantage to Chelex is that the extracted DNA is unstable and must be used within a few days.

A variety of DNA extraction methods have been tested for the PCR (Goldenberger et al. 1995, Steiner et al. 1995, Hammond et al. 1996, Shahjahan et al. 1995, Aljanabi and Martinez 1997), and many commercial kits are available. However, these may be expensive if used to extract DNA from large

**Table 8.3: A Rapid Method for Extracting DNA from a Single Insect or Mite Using Chelex® 100 Chelating Resin.**

1. Add a single insect or mite to a microcentrifuge tube. Insects can be alive or frozen at –80 °C.
2. Tap tube sharply to move the insect to the bottom of the tube. If the insect or mite is difficult to detect visually, add a small amount of buffer and spin in a microfuge tube to ensure the specimen is at the bottom.
3. Immerse the bottom of each microfuge in liquid nitrogen. Freeze a plastic pestle in liquid nitrogen. (Pestles are prepared in advance by melting the ends of 200-µl pipettor tips and molding them to the bottoms of microcentrifuge tubes.)
4. Macerate the frozen specimen well within the tube with the frozen pestle.
5. Add 200 µl of a 5% (w/v) Chelex® solution (Bio-Rad Laboratories).
6. Vortex the solution vigorously to thaw it.
7. Remove the pestle and place tube into a temperature cycler and heat to 56 °C for 15 minutes.
8. Centrifuge the sample (>100 g) for 15 seconds in a nanofuge to allow removal of the DNA solution from the top of the tube. Avoid removing any Chelex resin from the bottom of the tube. The DNA can be used for both traditional and RAPD-PCR. The DNA is not suitable for cutting with restriction enzymes, ligation reactions, or DNA sequencing. The DNA can be stored for a few days only at –20 °C.

(Adapted from Edwards and Hoy 1994.)

numbers of samples. The “best” DNA-extraction method will depend on the goals of your experiment. Preliminary experiments should be conducted to determine which DNA-extraction method is appropriate. For example, if the goal is to process large numbers of insects for ecological studies, then rapid, nontoxic and inexpensive extraction methods, such as Chelex, may be useful and sufficient, especially if amplifying multiple-copy genes in mitochondria or ribosomes. However, Chelex may yield some false-negative results and the extracted DNA cannot be stored for long periods of time. If it is important to maintain live insects for further study, it is possible to extract insect hemolymph from large insects, extract the DNA from the hemolymph with Chelex, and yield DNA suitable for microsatellite analysis (Gerken et al. 1998).

Sometimes, it is important to preserve the intact arthropod specimen so that morphological analyses can be made (Paquin and Vink 2009, Jeyaprakash and Hoy 2010). Several protocols have been developed that would allow nuclear and mitochondrial DNA to be extracted from whole insects or mites by soaking the specimen in an extraction buffer, yet allow the specimen to be preserved (Table 8.4). Methods developed to extract DNA and RNA from fossil bones and teeth have been used to extract DNA from arthropods (Cox 1968, Hoss and Paabo 1993, Rohland et al. 2004, Gilbert et al. 2007, Rowley et al. 2007, Jeyaprakash and Hoy 2010) without destroying the specimen. DNA and RNA can be extracted from fossil bones (Hoss and Paabo 1993) using a salt (guanidinium thiocyanate [GuSCN]) or guanidinium hydrochloride (GuHCl) (Table 8.4). It was

**Table 8.4: Extraction of DNA From a Mite by Soaking in a GuSCN Buffer Followed by Isolation Using a Silica Matrix Leaves the Body Intact for Morphological Analysis.**

1. Prepare GuSCN buffer with 5 M GuSCN, 50 mM Tris, pH 8.0, 25 mM NaCl, 20 mM ethylene dinitrioltetra-acetic acid (EDTA) and 1.3% Triton X-100.
2. Prepare the nucleic-acid-binding silica matrix by mixing 15 g of silicone dioxide with 125 ml of distilled water and allow it to settle for a week. After removing and discarding the supernatant, add fresh distilled water (125 ml), mix well, and allow the silica matrix to settle for another week. Resuspend the settled silica matrix in 25 ml of fresh distilled water plus 25 ml of HCl, then autoclave and store at room temperature in a dark bottle (Boom et al. 1990).
3. Place single specimens (fresh or stored in 95% EtOH) in 200 ml of GuSCN buffer at 60 °C for 16 hours. After soaking, 190 ml of the supernatant containing the DNA can be removed, leaving the specimen in 10 ml. It can be stored in alcohol or slide mounted.
4. To isolate the genomic DNA, the supernatant is mixed with 10 ml of silica matrix, incubated at 25 °C for 10 minutes and then centrifuged at 12,000 rpm at room temperature. The pellet produced is washed with 70% EtOH, resuspended in 10 ml of sterile water, incubated at 60 °C for 5 minutes, centrifuged again to remove the supernatant containing DNA from the silica matrix, then stored at –20 °C.

(Modified from Jeyaprakash and Hoy 2010.)

possible to extract sufficient DNA from mite eggs that PCR could be conducted using the GuSCN buffer followed by silica-matrix purification, which resulted in DNA suitable for amplification of mitochondrial and single-copy nuclear genes (Jeyaprakash and Hoy 2010).

Some experiments require specialized extraction methods (Mauel et al. 1999). For example, different numbers of false-negative results were obtained when DNA from ticks infected with a pathogen, granulocytic ehrlichiosis, was extracted by three different methods. Blood-fed ticks have inhibitors of the PCR that cannot be extracted easily with standard extraction methods. Inhibition of the PCR also was observed when amplifying *Borrelia burgdorferi* DNA from blood-fed ticks (Schwartz et al. 1997).

Some PCR protocols require higher-quality DNA than others. For example, Amplified Fragment Length Polymorphism PCR (AFLP-PCR) allows insects to be "finger-printed" (see Section 8.4.1), but requires very pure and high quality DNA that can be cut completely by restriction endonucleases. When different DNA extraction methods were compared, two of the complex methods failed to produce adequate amounts of DNA, one simple method produced only poor quality DNA, but three treatments (two complex DNA methods involving phenol treatments plus a CTAB-based protocol) produced an adequate quality and quantity of DNA (Reineke et al. 1998).

The PCR is inhibited by a variety of impurities including: complex polysaccharides, heme in blood, humic substances in soil, proteases, urea in urine, phenol, and detergents (Schwartz et al. 1997, Al-Soud and Radstrom 1998). One approach to reduce the amount of impurities is to dilute them (Table 8.2). Upon dilution, however, the template DNA must remain sufficiently abundant. Other methods for eliminating inhibitors include the use of dialysis or centrifugation in cesium chloride gradients, but these methods can result in the loss of large amounts of the template DNA.

Experimental procedures sometimes can introduce inhibitors of the PCR. For example, Lee and Cooper (1995) found that PCR carried out on DNA cloned into *E. coli* failed when the bacterial colonies containing the clones were picked from plates with wooden toothpicks. The nature of the water-soluble inhibitor in the wooden toothpicks is unknown, but the toothpicks negatively affected both *Taq* and Vent DNA polymerases.

If inhibition is a serious problem, it might be reduced by embedding whole cells in low-melting point agarose blocks, then immersing the block in a lysis buffer, which results in intact genomic DNA with minimal shearing damage. The agarose is then washed and cellular debris and other contaminants diffuse out during the lysis and washing steps, resulting in highly purified genomic

DNA free of contaminants (Moreira 1998). The agarose-embedded DNA can be used directly because the PCR is unaffected by the presence of high-quality low-melting-point agarose in concentrations up to 0.3%.

The use of degraded and fragmented DNA as a template for the PCR generally should be avoided because it reduces the efficiency of the PCR and limits the size of the products that can be amplified (Golenberg et al. 1996).

Concentrations of 0.05–1.0 µg of genomic DNA typically are used to amplify single-copy genes with the standard PCR. Less DNA (0.5–2 ng) can be used to amplify multiple-copy genes such as nuclear ribosomal RNA genes because these genes are repeated ≈200–500 times in the eukaryotic genome. It even may be possible to conduct PCR on ancient or degraded DNA if the template is “reconstructed” (Golenberg et al. 1996).

Table 8.5 describes a standard protocol for preparing genomic DNA suitable for the PCR from a single *D. melanogaster* (Jowett 1986). Other techniques are possible.

**Table 8.5: Extracting Genomic DNA From a Single *Drosophila melanogaster*.**

1. In a 1.5-ml microfuge tube, freeze a fly in liquid nitrogen. (Store at –70 °C until needed.)
2. Thaw and add 100 µl of 10 mM Tris·HCl (pH 7.5) 60 mM NaCl, 50 mM EDTA, 0.15 mM spermine, and 0.15 mM spermidine.
3. Grind fly with a yellow pipet tip.
4. Add 100 µl of 1.25% SDS, 0.3 M Tris·HCl, 0.1 M EDTA, 5% sucrose, and 0.75% freshly added diethylpyrocarbonate (DEP).
5. Mix and incubate 30–40 min at 60 °C to ensure lysis of the nuclei.
6. Cool and add 30 µl of 8 M potassium acetate.
7. Cool for 45 minutes on ice.
8. Spin for 1 minutes in a microfuge.
9. Remove supernatant, avoiding the lipid on the surface and add 2 volumes of ethanol.
10. Leave at room temperature.
11. Spin for 5 minutes and pour off the supernatant.
12. Wash the pellet with 70% EtOH.
13. Dry under vacuum.
14. Take up pellet in 25 µl of TE buffer (10 mM Tris and 1 mM EDTA).

The method lyses the nuclei once the tissue is broken up. The SDS and protein form complexes. DEP is a protein denaturant and nuclease inhibitor. The protein/SDS complexes are precipitated by adding potassium, leaving the DNA in solution. The final DNA is contaminated with RNA, which can be removed by adding RNase to a concentration of 100 µg/ml.

(Adapted from Jowett 1986.)

### 8.3.9 PCR Automation

The PCR involves repeated cycles with at least two, and generally three, temperatures (Figure 8.2). A high temperature is needed to denature (separate the two strands) the DNA template and subsequent product molecules. The lower temperature should allow annealing of the primer to the denatured ss DNA template. A third intermediate temperature close to the optima for DNA polymerase function is used for the extension (synthesis of the complementary sequence) phase. The annealing temperatures should be neither too low nor too high. An algorithm can be used to determine the optimal annealing temperature for a given pair of primers and template DNA based on the GC content of the primer-template sequences (Rychlik et al. 1990). This algorithm is available in computer programs (Osborne 1992) and at many websites of companies that provide primer synthesis services.

A variety of commercial thermal cyclers controlled by microprocessors automate the rapid and precise heating and cooling required for maximum efficiency of the PCR. Three basic categories of commercial temperature cyclers are available in which the reaction is 1) heated and cooled by fluids, 2) heated by electric resistances and cooled by fluids, or 3) heated by electric resistances and cooled by semiconductors. Accuracy and reproducibility in temperature control should be a concern when choosing a temperature cycler. Temperature cyclers are designed for use with 0.5- and 0.2-ml microfuge tubes or with 96-well plates.

Maintaining a close fit between the walls of the block and the microfuge tube or well plate is critical in maintaining accurate temperatures. Filling the wells with glycerol or mineral oil will encourage the transfer of heat. Thermocyclers should be checked periodically to determine their accuracy and calibrated if needed; differences of even 1–2 °C can be significant.

Commercially available temperature cyclers cost ≈US\$3000–\$8000. They differ in the design of the cooling system, control of ramping time between temperature steps, memory capacity for program storage, sequential linking of programs, and capacity of the heating block to hold different numbers of samples for amplification. Machines that do not provide a uniform temperature across a heating block can lead to variation in outcomes from the reactions taking place in different samples. Different models or brands of temperature cyclers, while ostensibly programmed to produce the same temperature profiles, may not be equivalent which can alter the outcomes of the PCR.

### 8.3.10 Specificity of the PCR

The specificity (or fidelity) of the PCR based on DNA synthesis by the Klenow DNA polymerase is low. The use of *Taq* and other DNA polymerases not only

simplifies the PCR but also increases the specificity and overall yield. The higher temperature optimum for *Taq* ( $\approx 75^\circ\text{C}$ ) allows the use of higher temperatures for primer annealing and extension, which increases the stringency of the reaction and minimizes the extension of primers that are mismatched with the template DNA.

The increase in specificity with the use of *Taq* also results in an increased yield of the target fragment because competition by nontarget products for DNA polymerase and primers is reduced. In the later cycles, the amount of polymerase may no longer be sufficient to extend all the annealed primer-template complexes in a single cycle interval, which results in reduced efficiency and a "plateau." This plateau is reached after  $\approx 30$  cycles when *Taq* is used rather than after 20 when the Klenow fragment is used.

Modifications of the standard PCR can enhance the outcome. Adjusting the annealing temperature can control the stringency of the annealing step; high-temperature annealing and extension ( $>55^\circ\text{C}$ ) and a balanced ratio of  $\text{Mg}^{2+}$  and dNTP concentrations give the greatest fidelity in the final product (Table 8.2). Various additives such as DMSO (2–5%), PEG 6000 (5–15% polyethylene glycol), glycerol (5–20%), nonionic detergents and formamide (5%) can be incorporated into the reaction to increase specificity (Roux 1995).

Optimizing the annealing temperature and minimizing the incubation time during the annealing and extension steps limits the amount of mispriming. Reducing primer and *Taq* concentrations also reduces mispriming. Changing the  $\text{MgCl}_2$  concentration can increase specificity by allowing a higher annealing temperature, which increases the stringency of the reaction.

Although *Taq* has no 3' to 5' exonuclease (proofreading) activity, its error rate is low compared with that of the Klenow fragment (Mullis et al. 1986, Keohavong and Thilly 1989), because *Taq* has a 5' to 3' exonuclease activity during polymerization (Erlich et al. 1991). Current estimates of misincorporation rates are  $10^{-5}$  nucleotides per cycle under optimized conditions. *Taq* appears to extend a mismatched primer-template significantly less efficiently than a correct primer-template. Misincorporated bases cannot be removed and this can promote termination of the extending DNA chain, which prevents propagation of the errors in subsequent PCR cycles but lowers the yield of the PCR. Because the accumulation of mutations in the PCR product is proportional to the number of DNA replications, the fewer cycles that are required to provide an adequate yield of DNA the better. Starting with adequate amounts of template (but not too much) reduces the number of PCR cycles required to produce a specific amount of product, and hence the amount of misincorporations.

Sometimes variability in PCR assays is due to bad batches of *Taq* DNA polymerase, but a functional assay can be carried out to test its performance (Wada et al. 1994).

**False-negative results** can occur for no apparent cause. These may be due to “interferences between our target DNA and the rest of the genome” (Baldrich et al. 1999). A solution can be to first digest the genomic DNA with a restriction enzyme that cuts outside the target region, followed by electrophoresis of the digested DNA, followed by recovering the restriction fragments of approximately the desired size by elution from the agarose gel. These fragments are then used as the template.

### 8.3.11 Detecting Primer Artifacts

Artifactual products consisting of low molecular-weight DNA products may be produced and are most obvious if the PCR is carried out with high primer concentrations, too much *Taq* in early cycles, small amounts of template DNA, or too many cycles. The artifacts may be “primer-dimers” or other artifacts derived from the primers. Methods have been developed to eliminate primer-dimer accumulation (Brownie et al. 1997).

**Primer-dimers** occur when the enzyme makes a product by reading from the 3' end of one primer across to the 5' end of the other. Because each primer serves as both primer and template, a sequence complementary to each primer is produced, which upon denaturation is a perfect template for further primer binding and extension. As the number of cycles is increased over 30, the probability of mispriming increases, as does the amount of artifact formed. The accumulation of a large amount of primer-dimers depletes primers and dNTPs from the reaction mixture and competes for enzyme with the desired target DNA.

If a PCR produces inadequate amounts of product, conducting a second amplification is a better solution to the problem than increasing the number of cycles of a single PCR. The second reaction is best done using 1 µl of the first reaction as template and a fresh reaction mixture.

### 8.3.12 How Many Cycles Does a PCR Need?

The answer is not too many and not too few. The optimum varies with the starting concentration of the template, the quality of the template, the amount of inhibitory substances in the reaction, as well as all the other parameters (Table 8.2).

Too many cycles can increase the amount of nonspecific background products. Too few cycles will give a low yield that can't be detected upon electrophoresis and staining with ethidium bromide. Too much template DNA actually can

inhibit the PCR. Too little template DNA can result in false negatives (Rameckers et al. 1997). The number of template molecules and cycles needed to give a good yield ( $\approx 10$  ng of DNA) was estimated, making the assumption that the efficiency of the PCR actually is approximately 70% (not 100%) and the product is 200 bp in length (Rameckers et al. 1997):

No. of template DNA molecules	Theoretical no. of cycles required
1	44
10	40
100	35
1,000	31
10,000	27
100,000	22

The efficiency of DNA amplification declines in the later cycles. This is called the **amplification plateau** because the product stops being produced exponentially and enters a linear or stationary phase (Kainz 2000). The plateau appears to be due to the binding of DNA polymerase to its amplification products. In general, it is better to set up multiple reactions if large amounts of DNA are needed.

### 8.3.13 Reducing the Evils of Contamination

It is crucial that laboratory techniques be meticulous to prevent contamination of the laboratory, supplies, and equipment with target DNA. Contamination can be an enormous problem because allele-specific PCR can generate copies of DNA from very small amounts of template (theoretically from a single molecule). Carryover of tiny quantities of PCR product can lead to **false positives** in subsequent reactions.

There is no simple and guaranteed method to prevent contamination. You must be thoughtful and careful *at all times*, using a variety of approaches to reduce the possibilities of contamination. Most importantly, you must have adequate controls to detect contamination.

Work surfaces can be decontaminated with 0.07 M sodium hypochlorite (10% bleach), which degrades DNA. Commercial products containing RNase solution eliminates RNA. UV irradiation of the workstation can be helpful, although dried DNA is less susceptible to UV irradiation than hydrated DNA (Roux 1995). Don't forget that UV irradiation of DNA polymerases, DNA template, and primers can damage them, reducing the efficiency of the PCR. UV light was reported to inhibit PCR amplification efficiency, even when only the water was irradiated, so routine decontamination with UV light should be used cautiously (Pao et al. 1993).

Autoclaving may not eliminate DNA contamination (Dwyer and Saksena 1992). In fact, PCR protocols published by the Cold Spring Harbor Laboratory recommend using microfuge tubes and tips without autoclaving them first to reduce the likelihood that undegraded DNA left over from previous autoclave cycles will contaminate them (Sambrook and Russell 2001).

It is crucial to separate physically the PCR amplification site from the location where the PCR products are evaluated by electrophoresis. Ideally, three separate sites, or rooms, will be available: one for DNA extraction, one for PCR amplification, and one for analysis of PCR products. Each separate room or containment unit should have a separate set of supplies and pipettors. Amplified DNA should never be brought into the area where DNA is being prepared for amplification or where it is being extracted. Reagents and supplies should never be taken from an area where PCR analyses are performed.

PCR reagents should be aliquotted to minimize the possibility of contamination. All reagents should be prepared, aliquotted, and stored in an area free of PCR products. Similarly, primers should be synthesized and purified in an environment free of PCR products.

To reduce contamination from barrels of pipettors, use positive displacement pipettors with disposable tips and plungers that are completely self contained. Don't "shoot" the tips off after use; that helps to make an aerosol of the DNA. Gently pull tips off the pipettor, especially after handling PCR products. Tips that are plugged with a filter should reduce contamination from DNA aerosolization. Contamination also can come from electrophoresis equipment, dot-blot apparatus, razor blades, microcentrifuges, water baths, and other equipment.

Contamination risks can be reduced by changing gloves frequently (especially between DNA extraction, PCR amplification, and analysis), wearing different laboratory coats for DNA preparation, PCR amplification and analysis, uncapping tubes carefully to reduce aerosol formation, and minimizing handling of DNA samples (Kitchin et al. 1990). Components of the PCR (mineral oil, dNTPs, primers, buffer, and enzyme) can be added to the tubes before adding the template DNA. Contamination will be reduced if each tube is capped before adding DNA to the next.

The use of positive controls can create a contamination problem. Ideally, if a **positive control** is necessary to demonstrate that your PCR is working appropriately, it should consist of template DNA that amplifies *weakly*, but consistently. Using DNA that produces strong positive responses will generate large amounts of amplified DNA, which is likely to cause contamination problems. It may be

undesirable, and unnecessary, to use a positive control after the PCR has been optimized.

By contrast, multiple **negative controls** *always* should be included in the PCR experiment because they will allow you to detect contamination (**false positives**) if used consistently and in adequate numbers. Negative controls consist of all reagents, but lack template DNA. A small number of contaminating template DNA molecules in the negative controls could lead to *sporadic* false-positive results. Thus, it is important to carry out multiple negative controls *each time* so that rare contaminants can be detected. How many negative controls should be used? There currently is no standard number, but a statistician might say that having *more* negative controls than experimental units is desirable (although expensive).

In conclusion, although UV irradiation has been recommended as effective in inactivating contaminating DNA and has been widely used, it should not be counted on as the only method to prevent contamination (Dwyer and Saksena 1992, Frothingham et al. 1992). Furthermore, autoclaving may not eliminate previously amplified PCR products (Dwyer and Saksena 1992). Thus, meticulous attention to the entire set of procedures, including the physical separation of DNA isolation, PCR amplification, and PCR analysis, is critical in minimizing contamination or carryover problems.

## 8.4 Some Modifications of the PCR

Up to now, the discussion has described allele-specific PCR for which primers can be designed because sequence information is available. What can you do if you want to amplify DNA from an arthropod for which little genetic information is available? What if you want single-stranded DNA rather than double-stranded DNA as a product? What if you want to find the sequence of DNA upstream or downstream from a specific gene? Some solutions to these, and other, problems have been achieved by modifying the types and numbers of primers used in the PCR (Table 8.6). Table 8.7 outlines some of the questions that can be answered by modifications of the PCR.

### 8.4.1 AFLP for DNA Fingerprinting

Amplified Fragment Length Polymorphism (AFLP) provides a method for developing DNA fingerprints that eliminates some of the problems inherent in RAPD-PCR, AP-PCR, and DAF (described below) (Savelkoul et al. 1999). AFLP has its own limitations, namely that it generates dominant rather than codominant markers (Mueller and Wolfenbarger 1999), and it requires absolutely clean template DNA in consistent quantities.

**Table 8.6: Some Modifications of the PCR Use Different Types of Primers.**

PCR type, Primer number, nt length	Potential uses
<b>Standard allele-specific</b> Paired, 15–30 nt each	Amplify DNA for which sequence information is available for the target DNA.
<b>AFLP</b> Generic primers based on restriction site sequences and “adapter” sequences	DNA is digested by restriction enzymes, and oligo “adapters” are ligated to digested DNA and amplified using generic primers that use the restriction-site sequences and adaptor sequences as target sites for primer annealing. AFLP-PCR yields multiple fragments.
<b>Anchored</b> One known primer, second is made	Amplify DNA when only one primer sequence is known. Synthesis of cDNA with the known primer is carried out using mRNA; a poly(G) tail is added to the cDNA. The second primer is made by synthesizing a primer with a poly(C) sequence, which allows amplification of a second DNA strand that is complementary to the cDNA.
<b>Arbitrary</b> Single primer, 18–30 nt arbitrary sequence	Amplify regions of DNA internal to regions to which arbitrary primers (such as M13 sequencing primer, M13 reverse sequencing primer, or T3 sequencing primer) anneal on opposite strands. One or more DNA fragments will be produced, and these fragments can be used to generate genome maps or discriminate between individuals, populations, or species.
<b>Asymmetric</b> Paired primers, 10–30 nt in a 1:50 to 1:100 ratio	Amplify ss DNA for sequencing.
<b>Degenerate</b> Multiple types, 15–30 nt	Amplify DNA that is related to genes for which the sequence or part of the sequence is known in a related species, or for members of a gene family. The degeneracy of the DNA code, and codon bias, for amino acids determines how many primer types are needed in the reaction.
<b>Inverse</b> Paired primers, 15–30 nt inverse orientation	Amplify regions of DNA of unknown sequence that flank known sequences; used for identifying upstream/downstream sequences. Primers are oriented so DNA synthesis occurs away from the known “core” DNA.

(Continued)

**Table 8.6: (Continued)**

PCR type, Primer number, nt length	Potential uses
<b>Multiplex</b> Multiple primers	More than one pair of primers amplify several DNA targets simultaneously. Careful optimization of PCR conditions is required to produce consistent results.
<b>PCR-RFLP</b> Paired primers, 18–30 nt	Nuclear DNA is amplified by the standard PCR then the product is cut with restriction enzymes. Banding patterns are visualized on a gel after staining with ethidium bromide.
<b>Quantitation of mRNA</b> Paired primers, 15–30 nt	Several methods: 1. Two different cDNAs are amplified and the absolute level of one is calculated if the other is known. 2. The sample is spiked with a known amount of control DNA, and target and control DNA are amplified and compared to estimate the amount of target DNA.
<b>RAPD-PCR</b> Single primer, 10-nt random sequence	Random amplified polymorphic DNA PCR. Amplify regions of DNA that are flanked by the random primer sequences. Multiple DNA fragments may be produced and used as markers for genome mapping or identifying individuals, populations, or species.
<b>RNA amplification</b> 18–22 nt	mRNA is reverse transcribed and the cDNA is amplified by PCR.

As is described below, RAPD-PCR, AP-PCR, and DAF fingerprinting methods are based on amplifying random genomic DNA fragments using arbitrarily selected PCR primers, which means that “DNA fingerprints” can be generated from any DNA without prior knowledge of the DNA sequence. These PCRs are performed at low annealing temperatures to allow the primers to anneal to the template at multiple loci, which makes them very sensitive to reaction conditions, including DNA quality and quantity, and PCR temperature profiles.

AFLP eliminates most of these problems because it is based on detecting restriction fragments by PCR amplification; AFLP can be used on DNAs of any origin and complexity, without requiring prior knowledge of sequence, and using a limited set of generic primers (Vos et al. 1995, Savelkoul et al. 1999).

**Table 8.7: Some Entomological Problems and Potential PCR Protocols.**

Problem	PCR technique(s)
Amplify ancient DNA	Standard allele-specific PCR
Amplify mRNA	Reverse transcriptase PCR
Chromosome walking	Inverse PCR; Long PCR
Cloning a gene	Blunt-end cloning; sticky-ended cloning; anchored PCR; PCR with degenerate primers; Long PCR
Constructing a genetic map	AP-PCR; RAPD; inverse PCR
Constructing a phylogeny	Standard PCR with primers having polylinkers for cloning/sequencing; asymmetric PCR and sequencing; PCR-RFLP; multiplex PCR
Detecting gene expression	RNA PCR; TaqMan PCR
Detecting mutations	Standard PCR; RAPD; AP-PCR; PCR-RFLP
Detecting pathogens in arthropod vectors	Standard PCR; Long PCR
Detecting transgenic arthropods	Standard PCR
Engineering DNA	
Introduce restriction sites into DNA fragments	Attach sequences to 5' end of primers and conduct standard PCR probes, or isolating DNA strands on a column
Label DNA with $^{32}\text{P}$ or biotin for sequencing	
Assemble overlapping DNA segments to make synthetic DNA	Alter primer sequence when synthesizing, then standard PCR
Introduce substitutions, deletions, or insertions in product DNA	
Evolutionary analyses	Standard PCR; RAPD; AP-PCR; DNA sequencing; PCR-RFLP
Identify species	Standard PCR; RAPD; AP-PCR; PCR-RFLP
Identify strains, races, or biotypes	Standard PCR; RAPD; AP-PCR; PCR-RFLP
Identifying upstream/downstream sequences	Inverse PCR; single-specific primer PCR (SSP-PCR)
Monitoring dispersal of individuals	Standard PCR; RAPD; AP-PCR
Sequencing a gene	Asymmetric PCR to produce ss DNA; dideoxynucleotide chain-termination sequencing method with <i>Taq</i> polymerase; cycle sequencing; direct sequencing

There are three steps in AFLP-PCR: 1) the DNA is digested by restriction enzymes and oligo "adapters" are ligated to the digested DNA; 2) sets of restriction fragments are selectively amplified using the adapter and restriction site sequences as target sites for primer annealing; and 3) the amplified fragments are analyzed by electrophoresis.

The number of fragments produced in a single AFLP-PCR can be determined by selecting specific primer sets. Annealing conditions can be stringent in AFLP. AFLP-PCR allows analysis of closely related populations and species. For example, populations of an introduced pest in Europe, the grape phylloxera, were compared by

AFLP-PCR with North American populations, from which they were suspected to have derived ([Forneck et al. 2000](#)). Two distinct populations were found in Europe, and AFLP-PCR patterns suggest that two different introductions occurred, one from the northeastern United States and the other from the south central United States.

A simplified version of AFLP was developed to discriminate between European and African honey bees ([Suazo and Hall 1999](#)). The protocol involved digesting DNA and ligating the adapters in one reaction rather than two; one restriction enzyme was used rather than two; and amplification was accomplished in one reaction rather than two. Finally, the PCR products were electrophoresed in agarose-Synergel instead of polyacrylamide and visualized by ethidium bromide staining rather than autoradiography of labeled primers. These modifications in AFLP-PCR may reduce the amount of polymorphism detected.

#### **8.4.2 Anchored PCR**

If only one sequence is known that is suitable for a primer (rather than two), **anchored PCR** can be used. The procedure involves synthesis of cDNA with the known primer from mRNA ([Collasius et al. 1991](#)). A poly(G) tail is added to the cDNA. The second primer is developed by synthesizing a primer with a poly(C) sequence, which allows amplification of a second DNA strand that is complementary to the cDNA. Subsequent cycles yield amplified DNA from both strands.

#### **8.4.3 Arbitrary Primers**

Ecologists, evolutionary biologists, and geneticists often wish to develop genetic markers for insects for which little genetic information is available. **Arbitrarily primed PCR** (AP-PCR) can produce a characteristic “fingerprint” pattern for any genome, which could be useful for developing markers for breeding programs, genetic mapping, population genetics, or epidemiology ([Welsh and McClelland 1990](#), [Welsh et al. 1992](#), [McClelland and Welsh 1994](#)).

AP-PCR involves two cycles of low-stringency amplification, followed by cycles conducted at higher stringency, using a single primer of arbitrary sequence. The term **stringency** refers to PCR conditions such as the annealing temperature. If a high annealing temperature is used, then the primers will only anneal to the template DNA if a high proportion of the sequences match. Lower annealing temperatures allow some mismatches. Full-length primers (20–34 nt long) that have been used include the Universal M13 sequencing primer, the M13 reverse sequencing primer, and the T3 sequencing primer.

How does AP-PCR work? At lower temperatures, an arbitrary primer can anneal to many sequences with some mismatches. By chance, some primers will

be able to anneal to the target DNA within a few hundred bases of each other and on opposite strands. Sequences between these positions then will be amplified. The extent to which sequences amplify depends on the efficiency of priming and the efficiency of extension. During early cycles, those sequences that prime most efficiently will predominate. During later cycles, those that amplify most efficiently will predominate.

Between three and 20 DNA products typically are produced in AP-PCR, which allows differentiation between closely related strains of some species (Welsh et al. 1990). AP-PCR also has been used to amplify RNA in order to detect and clone mRNAs that are differentially expressed in different cells (McClelland et al. 1995). Clones of the aphid *Ceratovacuna nekoashi* from a single gall were shown by AP-PCR to be genetically identical, whereas aphids from different galls on the same twig were successfully differentiated, indicating that members of a gall constitute a clonal population, a gall is founded by a single female, and intergall migration is absent or rare (Fukatsu and Ishikawa 1994).

A modification of AP-PCR was developed and called DALP, or Direct ALength Polymorphisms (Desmarais et al. 1998). DALP uses the M13 sequencing forward primer as a core sequence for the forward primer and the M13 reverse primer. These primers produce specific multibanded patterns that show inter-individual length variations. Each band then can be sequenced with the universal sequencing primers.

#### **8.4.4 Asymmetric PCR**

Single-stranded DNA can be produced by asymmetric PCR. By providing an excess of primer for one of the two strands, typically in ratios of 50:1 to 100:1, amplification results in product that is primarily single-stranded. Early in the reaction, both strands are produced, but as the low-concentration primer is depleted, the strand primed by the abundant primer accumulates arithmetically. Such ss DNA is particularly useful for sequencing (see Chapter 7).

#### **8.4.5 Degenerate Primers**

If only a limited portion of a protein sequence is known for a target gene, **degenerate primers** may allow detection of new or uncharacterized sequences in a related family of genes, or may amplify members of a gene family. Degenerate primers are a mixture of oligonucleotides varying in base sequence, but with the same number of bases.

Designing degenerate primers for the PCR requires several considerations. You will recall that the genetic code is degenerate (with more than one codon

for most amino acids). Methionine and tryptophan are encoded by a single codon, but the other amino acids are encoded by two to six different codons. When designing degenerate primers, it is useful to choose a segment of the protein in which the amino acids have minimal degeneracy. The lower the degeneracy in the primers, the higher the specificity of the PCR. The degeneracy of the primer may be restricted further by considering which codons are most often used in a particular species (codon bias), if it is known. Furthermore, degeneracy may be reduced if primers containing fewer (15–20) nucleotides are used. Because a single mismatch, especially at the 3' end of the primer, may prevent *Taq* from extending, degeneracy at the 3' end should be avoided. Empirical testing of primers may be necessary and modifications made to ensure that the desired product is synthesized.

#### 8.4.6 Hot-Start PCR

A hot-start PCR protocol can optimize the yield of the desired product while limiting the likelihood of nonspecific amplification. Hot-start PCR is achieved by leaving an essential component out of the reaction mixture until the mixture has been heated to a temperature that inhibits nonspecific priming and extension. Typically, all PCR components are added and held at high temperature before the DNA polymerase is added.

A modification of this method involves using wax to provide a physical barrier between the components of the reaction. The primers, Mg<sup>2+</sup>, dNTPs and buffer can be mixed at room temperature in the bottom of the reaction tube and then covered with melted wax that melts at low temperature (53–55 °C). The remaining components are then added on top of the wax barrier. During the first cycle of the PCR the wax barrier melts during the denaturation step, allowing the components to combine. The melted wax floats to the top of the mixture where it acts as a barrier to evaporation. Hot-start *Taq* DNA polymerase is now available in which the enzyme is activated only after the reaction reaches 94 °C or higher, allowing all components to be mixed at room temperature.

Hot-start PCR especially is useful when nonspecific amplification is a problem because there is too little template DNA, the template DNA is complex, or several pairs of primers are used (multiplex PCR) (Sambrook and Russell 2001).

#### 8.4.7 Inverse PCR

An unknown sequence that flanks a “core” region with a known sequence can be amplified by **inverse PCR** (Ochman et al. 1990, Sambrook and Russell 2001). Inverse PCR involves digesting the template DNA with a restriction endonuclease

that cuts *outside* the region of known sequence to produce a fragment of  $\approx$ 3–4 kb. Southern-blot analyses may be necessary to identify restriction enzymes that produce fragments of suitable size for circularization and amplification. If an enzyme is used that cleaves within the core region, either the upstream or downstream segment of DNA will be amplified.

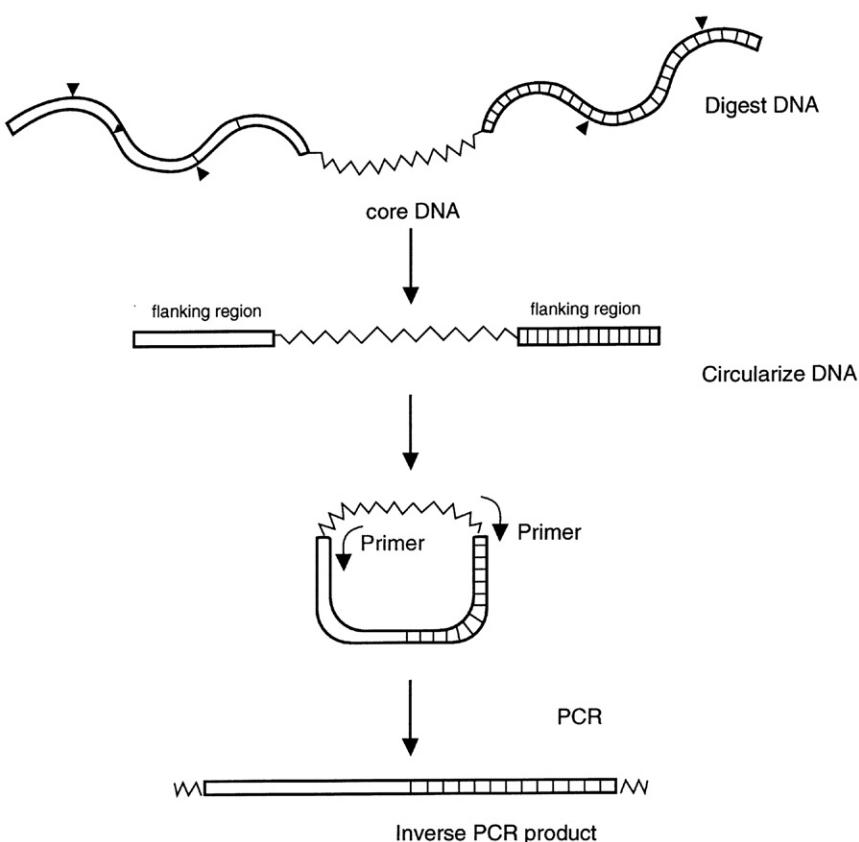
Once the DNA has been digested, the ends of the fragment are ligated to form a circular molecule. Ligation is performed with T4 DNA ligase in a dilute DNA concentration to favor formation of monomeric circles. Amplifying the flanking DNA outside the “core” region is carried out using primers oriented in the *opposite direction* of the usual (Silver 1991). Primers for inverse PCR thus are homologous to the ends of the core region so that DNA synthesis proceeds across the *uncharacterized* region of the circle rather than across the characterized core region (Figure 8.3).

#### 8.4.8 Long PCR or High-Fidelity PCR

In theory, standard PCR using a *Taq* DNA polymerase can amplify up to  $\approx$ 10 kb of DNA sequence. In practice, amplification of such long sequences is difficult; most amplifications are limited to 1–2 kb. However, large (up to 40 kb) DNA fragments can be amplified with high fidelity and yield if two DNA polymerases are used (Barnes 1994; Cheng et al. 1994a,b; Lahr and Katz 2009).

The effectiveness of Long PCR is due to the two DNA polymerases working together. A polymerase such as *Pfu* has a 3'-exonuclease activity (which *Taq* does not have), which means that *Pfu* can remove any accidental mismatches introduced into the growing DNA molecule. The theory is that one of the deterrents to truly long PCR products may be due to the incorporation by *Taq* of mismatched or damaged nucleotides every few kilobases, thus eliminating those DNA molecules from further amplification. The incorporation of errors occurs because *Taq* lacks a proofreading function. Adding a small amount of a proofreading polymerase such as *Pfu* to an excess of *Taq* provides *Taq* with a proofreading “helper.” Primer mismatchs are corrected and *Taq* is able to extend primers for longer distances. Several combinations of DNA polymerases have been used in long PCR; Cheng et al. (1994a) found that rTth (from *Thermus thermophilus*) and Vent (from *Thermococcus litoralis*) polymerases were the most reliable combination under their test conditions.

Several other modifications in the Long PCR protocol enhance the likelihood of obtaining longer products. Long PCR typically uses long primers (at least 30 nt), works best at a higher pH (8.8–9.0), uses a high rate of change in temperatures (rapid cycling), and a longer synthesis interval.



**Figure 8.3** Inverse PCR allows amplification of DNA flanking the “core” DNA, for which sequence information is available. Step 1 involves digesting the template DNA with an appropriate restriction enzyme to produce fragments  $\approx 2\text{--}4\text{ kb}$  long, with the “core DNA” in the middle. Step 2 involves circularizing the DNA by ligation. Primers, dNTPs, and DNA polymerase are added and the PCR is carried out. Primers are oriented so that synthesis of DNA occurs away from the core DNA into the flanking regions. The PCR product consists of the two flanking regions.

Why not simply use a single DNA polymerase with 3'-exonuclease activity to edit out the mismatches in extension? Barnes (1994) suggested that the enzymes with 3'-exonuclease activity could degrade PCR primers, especially during the long synthesis times. Thus, only small amounts of polymerase with 3'-exonuclease activity should be used. In addition, it is especially important that the template strands be completely denatured at high temperatures to prevent renaturation before primers can anneal and be extended (Cheng 1995).

Primer design for Long PCR, as usual, should avoid the potential for secondary structure and dimer formation (Cheng et al. 1994a). Primers of 21–34 nt that

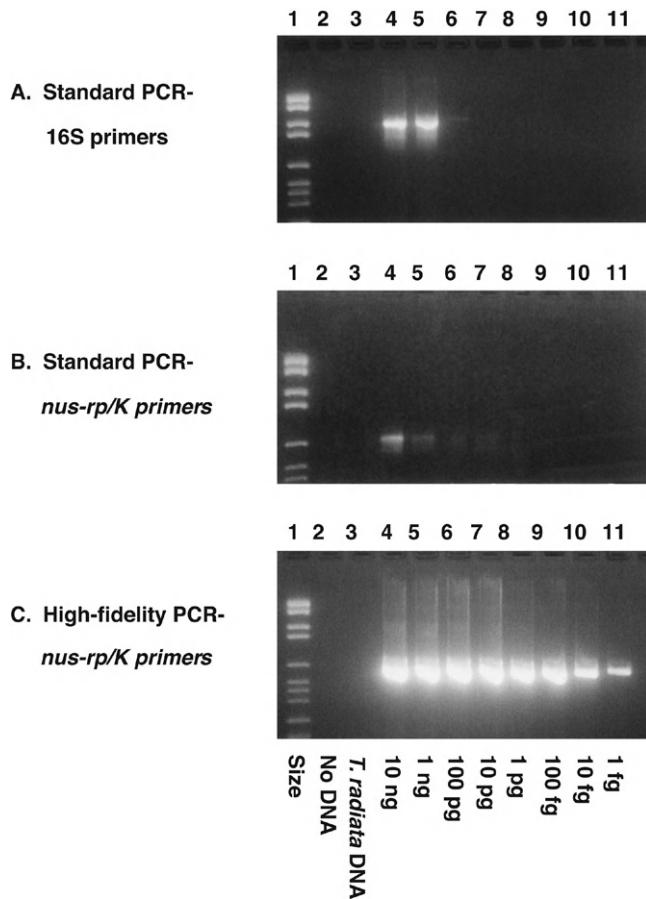
have melting temperatures near 65–70 °C permit the use of higher annealing temperatures to enhance reaction specificity. Thermal cycling profiles in Long PCR typically use a hot start at 78–80 °C, initial denaturation at 94°C for 1 minute, 25–40 cycles of denaturation at 94°C for 15 seconds, and annealing and extension steps at 60–68°C for 30–60 kb of target DNA. Typically, extension times are increased for each subsequent cycle to facilitate production of long product molecules.

Other factors that influence the success of the Long PCR include the integrity of the target DNA, which means that DNA extraction methods must be carefully considered. Longer targets can be amplified best from DNA with little shearing damage ([Cheng and Kolmodin 1998](#)). Several DNA extraction methods produce large DNA fragments from insects ([Ebert 1996](#)), although shearing of DNA may be difficult to avoid when extracting DNA from adult insects because their exoskeleton can damage the DNA during grinding. One solution is to extract DNA from embryos ([Rabinow et al. 1993](#)).

The Long PCR protocol has been used for another application—amplifying microbial DNA when mixed with arthropod DNA ([Jeyaprakash and Hoy 2000](#), [Hoy et al. 2001](#)). When insect and microbial DNA are mixed, efficient amplification of the microbial DNA seems to be inhibited, for unknown reasons. For example, the ability to detect the endosymbiont *Wolbachia* within the bodies of various arthropods was greatly enhanced when the Long PCR, rather than a standard PCR, protocol was used; *Wolbachia* were found in 76% of the 63 arthropods examined in 13 orders. The Long PCR thus can be used to increase sensitivity or fidelity of the PCR even when shorter DNA targets are amplified.

The Long PCR protocol is ≈5–7 orders of magnitude more sensitive in amplifying *Wolbachia* DNA than the standard PCR ([Jeyaprakash and Hoy 2000](#)). When standard and Long PCR protocols were compared using known amounts of *Wolbachia* template DNA mixed with known amounts of insect DNA, the Long PCR could amplify as few as 100 copies of *Wolbachia* DNA consistently. By contrast, standard PCR was only able to reliably detect *Wolbachia* DNA when at least 100 million copies of plasmid DNA were present. Similar results were obtained in the amplification of the bacterium causing citrus greening disease ([Hoy et al. 2001](#)) ([Figure 8.4](#)).

Long PCR should have many applications whenever long DNA fragments are useful; for example, Long PCR has been used to develop rapid restriction maps of DNA fragments of 8–18 kb ([Her and Weinshilboum 1995](#)). Long PCR can be used to clone large genes or be a labor-saving alternative for studying larger genome segments such as entire mitochondria that are 16–20 kb ([Nelson et al. 1996](#)).



**Figure 8.4** Long (or high-fidelity) PCR is more sensitive by  $\approx 6$  orders of magnitude than standard, allele-specific PCR when microbial DNA is mixed with insect DNA (Hoy et al. 2001). The same template DNA (a plasmid containing both the *nusG-rp/K* and 16S sequences) from the greening bacterium was serially diluted from 10 nanograms (ng) to 1 femtogram (fg) and added to 10 ng of parasitoid, *Tamarixia radiata*, DNA. A) 16S primers were used in a standard, allele-specific PCR protocol. Detectable products are in lanes 4 and 5 only. B) *nusG-rp/K* primers were used in a standard, allele-specific PCR protocol. Weak products are seen in lanes 4 and 5. C) *nusG-rp/K* primers were used with a Long PCR protocol. Strong products were produced in lanes 4–11. Lane 1, DNA size marker; lane 2, no template-DNA control; lane 3, 10 ng of *T. radiata* DNA only control; lanes 3–11 contain *T. radiata* DNA + 10 ng, 1 ng, 100 pg, 10 pg, 1 pg, 100 fg, 10 fg, and 1 fg of the plasmid DNA, respectively.

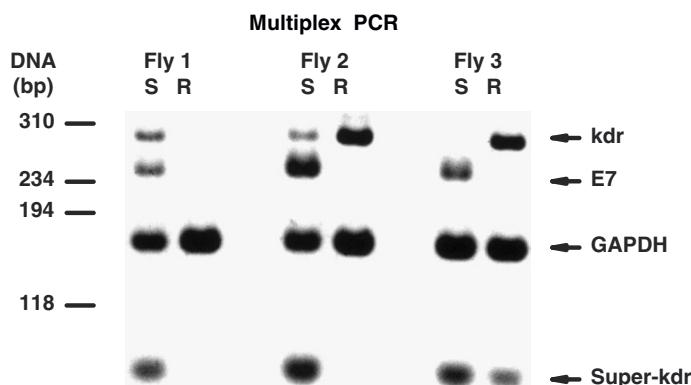
#### 8.4.9 Multiplex PCR

More than one pair of primers can be used to amplify multiple PCR products (Sambrook and Russell 2001, Hamiduzzaman et al. 2010, Meeus et al. 2010, Ravikumar et al. 2011). In multiplex PCR, the goal is to amplify several segments

of target DNA simultaneously, which should reduce time, minimize costs and increase efficiency (Figure 8.5). In reality, the yield of each product usually is reduced in proportion to the number of primer pairs included in the reaction. Up to eight primer pairs have been used simultaneously before the yield of each product is too low to be visualized by staining with ethidium bromide on an agarose gel.

Despite the potential benefits, multiplex PCR can be difficult to develop because all the primers must have approximately the same melting temperature, the primers should not interact with one another to produce primer-dimers, and the amplified products should be of approximately the same size, but still be distinguished from each other by gel electrophoresis.

Steps to develop multiplex PCR include the following: 1) Determine that all target DNA can be amplified efficiently using the same PCR temperature profile. 2) Titrate the amount of each primer pair to achieve maximum amplification in separate reactions using the same program and reaction conditions. 3) Balance the amount of each primer pair to achieve acceptable levels of amplification of all targets in the multiplex reaction. One solution to problems with step 3 usually involves increasing progressively the concentration of



**Figure 8.5** Multiplex PCR allows more than one gene to be sampled in a single reaction. In this example, three pesticide resistance genes (*kdr*, *E7*, and *Super-kdr*) were monitored in three hornflies, *Haematobia irritans*. Fly 1 has only susceptible alleles for *kdr*, *Super-kdr*, and *E7*. Fly 2 has one copy of the susceptible and one of the resistant alleles of *kdr*, but it has susceptible alleles only for the *E7* and *Super-kdr* genes. Fly 3 has only *kdr* alleles (is homozygous resistant), susceptible alleles of *E7*, and is heterozygous for *Super-kdr*. Amplification of the *GABDH* gene provides a control to demonstrate that the PCR is working correctly. (Photo kindly provided by Felix Guerrero.)

the nonworking primer pairs while reducing the concentration of the effective primer pairs. Other recommendations for optimizing multiplex PCR can be found in [Sambrook and Russell \(2001\)](#). Examples of the use of multiplex PCR include efforts to detect honey bee viruses (acute bee paralysis virus, Kashmir bee virus and Deformed wing virus) in bumble bees using reverse-transcriptase PCR ([Meeus et al. 2010](#)) and *Nosema apis* and *Nosema ceranae* infections in honey bees ([Hamiduzzaman et al. 2010](#)). [Rugman-Jones et al. \(2011\)](#) developed a method to identify any life stage of 10 parasitoids of soft scales in California citrus, using ribosomal RNA genes and three multiplex PCR protocols.

#### 8.4.10 Nested PCR

Nested PCR involves a two-step procedure in which one pair of primers is used to amplify a fragment. Subsequently, a second pair of primers is used to amplify a smaller fragment from an aliquot of the product of the first PCR. Nested PCR is designed to be both sensitive and specific.

Nested PCR of a 16S rRNA gene from the causative agent of granulocytic ehrlichiae (*Ehrlichia chaffeensis*), a disease of humans, was found to be so sensitive that as few as two copies of the 16S gene could be detected when a spiking experiment was conducted ([Massung et al. 1998](#)). Spiking experiments were conducted using known quantities of a plasmid containing the 16S rRNA gene spiked into background human genomic DNA. The use of serial dilutions to determine how repeatable and reliable a PCR assay is should be done whenever it is important to resolve how often false negatives are likely to occur in an experiment.

#### 8.4.11 PCR-RFLP

PCR-RFLP eliminates some of the disadvantages to traditional restriction fragment length polymorphism (RFLP) analysis for analyzing population variation using DNA isolated from individual insects ([Karl and Avise 1993](#)). If no primers are available from the literature, a genomic DNA library is constructed and clones are isolated. Clones with inserts of 500–2000 bp are chosen and sequences of the first 100–150 nt from both ends are obtained so that PCR primers can be designed. Nuclear DNA is amplified by the PCR using these primers and digested with appropriate restriction enzymes. The cut DNA is visualized after electrophoresis by staining with ethidium bromide. The advantage to PCR-RFLP is that DNA extracted from a single individual is sufficient, after amplification, to provide electrophoretic bands that can be visualized without having to be hybridized with radiolabeled probes.

#### 8.4.12 Quantitative PCR

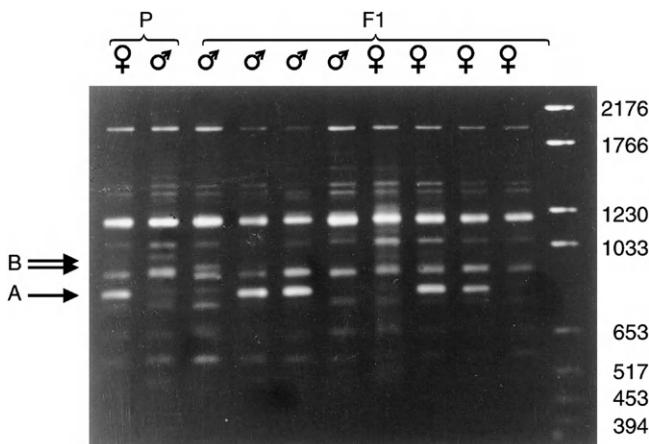
Methods have been developed to quantify the amount of DNA or RNA present in a sample (Arnheim and Erlich 1992, Siebert and Larrick 1992, Foley et al. 1993, Sambrook and Russell 2001, Bustin 2004, Baker 2011). Quantitative PCR requires some form of standard with which the target sequence concentration is compared. For example, estimation of the number of amplified esterase genes in insecticide-resistant mosquitoes used a nonamplified esterase gene as an internal control (Weill et al. 2000). Quantification of infection of fleas with the plague bacterium *Yersinia pestis* was based on standard, curve-based, competitive PCR (Hinnebusch et al. 1998). This quantitative PCR method was found to be equally accurate and precise as a colony-count method when evaluated using mock samples and laboratory-infected fleas.

Quantitative PCR is more difficult than other types of PCR due to the nature of the PCR. Because PCR is an exponential (or nearly so) process, small differences in efficiency at each cycle can lead to large differences in yield. Anything that affects exponential amplification can disrupt quantitation. Thus, different amounts of inhibitors in samples containing the same amount of template DNA could result in different amounts of product, as could small differences in efficiency between the primer pairs used to amplify the standard (control) and target sequences.

Quantification of amplified products can be achieved by gel electrophoresis or by fluorescently labeled primers quantified with an automated DNA sequencer, fluorometry, analysis of gel images stained with ethidium bromide or other intercalating dyes, or measurement of radioactivity incorporated during amplification (Sambrook and Russell 2001). Another method for quantitating PCR products is through real-time PCR (see Section 8.4.14) or TaqMan PCR (see Section 8.4.16).

#### 8.4.13 Random Primers

A method similar to AP-PCR was developed when Williams et al. (1990) demonstrated that genomic DNA from diverse organisms could be amplified using a single short (9- or 10-nt) primer composed of “random” oligonucleotides (Figure 8.6). The “random primers” can be designed without the experimenter having any genetic information for the organism being tested. The only constraints are that the primers should have 50–80% G + C content and no palindromic sequences. Different random primers used with the same genomic DNA produce different numbers and sizes of PCR products (Ellsworth et al. 1993, Kernodle et al. 1993, Meunier and Grimont 1993, MacPherson et al. 1993, Williams et al. 1993). The amplified DNA can



**Figure 8.6** Photograph of a gel showing mode of inheritance of RAPD-PCR DNA fragments from the parasitic wasp *Trioxyx pallidus*. A single 10-mer primer anneals to different regions of the genome and if two primers anneal in the opposite orientation, amplification of several DNA sequences occurs. The size marker on the right lane indicates the size of the bands. Arrow A indicates the band in the mother is inherited in approximately half of her haploid male and diploid female progeny, which is consistent with the hypothesis that the mother was heterozygous for this band. Arrow B identifies faint bands that are not inherited in a Mendelian manner and are not used in analyses (Edwards and Hoy 1994).

be detected as bands in ethidium bromide-stained agarose gels (Figure 8.6). This modified PCR method was called **RAPD-PCR** because it produced Random Amplified Polymorphic DNA bands.

RAPD-PCR has been used to develop genetic maps and to identify molecular markers in populations or species, as well as determine paternity in dragonflies (Hadrys et al. 1992, 1993; Tingey and del Tufo 1993; Schierwater 1995). RAPD-PCR makes it possible to identify hundreds of new markers in a short time, which allows genetic maps to be developed rapidly. RAPD-PCR is particularly valuable for genome mapping in those species for which other genetic markers are lacking or rare (Laurent et al. 1998).

Genomic DNA sequences differing by only a single base may not be amplified in the RAPD protocol, or they may result in a complete change in the number and size of the amplified DNA segments. Thus, RAPD-PCR may detect small differences in the genomes of individual arthropods, different populations, or species. RAPD fingerprinting can be carried out on very small arthropods, such as single aphid embryos, while preserving the mother for morphometric or karyotyping analyses (Chan et al. 1999) and is especially useful for discriminating between tiny parasitoid species or biotypes (Edwards and Hoy 1993, 1994;

Vanlerberge-Masutti 1994). Differences in RAPD-PCR patterns are correlated with the evolution of different taxa, allowing limited estimates of evolutionary divergence (Espinasa and Borowsky 1998).

RAPD-PCR products can be cloned (Comes et al. 1997) and sequenced so that “allele-specific” primers can be developed for future PCR analyses. Sequence-Characterized Amplified Region (SCAR) primers will produce allele-specific PCR products. Agusti et al. (2000) used a SCAR primer pair to amplify single bands of 310 bp to detect the whitefly *Trialeurodes vaporariorum* in the gut of the predator *Dicyphus tamaninii*.

Two or more primers have been used simultaneously to generate reproducible RAPD fragments that are different from those obtained with each single primer (Micheli et al. 1993, Sall et al. 2000).

RAPD-PCR has been criticized for its lack of reproducibility (Ayliffe et al. 1994, Lamboy 1994, Micheli et al. 1994, Hallden et al. 1996, Jones et al. 1997, Khandka et al. 1997, McEwan et al. 1998, Perez et al. 1998). Different RAPD banding patterns can be obtained if different DNA extraction methods are used, probably due to the presence of different kinds or amounts of contaminants or different amounts of template DNA (Micheli et al. 1994). Different DNA polymerases also may amplify different RAPD products (Schierwater and Ender 1993).

RAPD-PCR is sensitive to both DNA template concentration and quality, so bands may vary in intensity or even disappear if template concentration is not controlled or DNA is sheared (Khandka et al. 1997). Reproducibility also can be poor if different PCR machines or pipettors are used, resulting in different temperature cycling conditions or different concentrations of the PCR mixture (He et al. 1994a,b; Schweder et al. 1995). Occasionally, heteroduplex molecules that are formed between allelic sequences can cause artifactual RAPD bands (Ayliffe et al. 1994). Thus, it is critical that researchers use primers only if they produce bright, consistent banding patterns in the particular thermocycler used. Likewise, researchers should obtain a reference profile for their own work rather than comparing their results to those generated by another (He et al. 1994a).

Another criticism of RAPD-PCR is that all bands are inherited as dominant alleles. This means that heterozygotes cannot be identified unless progeny testing is conducted, although this is not an issue when RAPD-PCR is conducted on haploid males of arrhenotokous species (Edwards and Hoy 1993). Another problem is that comigration of similar sized bands with different sequences can occur, but may not be detected, unless the bands are cloned and sequenced.

Use of RAPD markers to calculate genetic similarity coefficients can result in false positives and false negatives if RAPD artifacts are present (Lamboy 1994).

As a result, Nei and Li's coefficient is recommended for computing genetic similarities with RAPD data, particularly if PCR artifacts are present.

Some of the negative aspects of RAPD-PCR can be eliminated by a method called Sequencing With Arbitrary Primer Pairs (SWAPP) (Burt et al. 1994). In this procedure, amplified random bands are purified from the gel, reamplified with the same two primers used in the initial amplification, and repurified. One of the primers then is added back and annealed to the product and sequenced. SWAPP allows polymorphisms in populations to be characterized at the nucleotide level, eliminates non-Mendelian inheritance, and allows bands to be produced reliably. The technique requires only small amounts of low-quality DNA and no prior genetic information for the organism. Lunt et al. (1999) developed a method for isolating microsatellite data from RAPD fragments that reduced the time and expense of traditional microsatellite isolation methods.

#### 8.4.14 Real-Time PCR

Real-time PCR can quantify gene expression and confirm differential expression of genes (Dorak 2006, Bustin et al. 2009, Baker 2011). Real-time PCR uses commercially available fluorescence-detecting thermocyclers to amplify specific nucleic-acid sequences and measure their concentration simultaneously (Sambrook and Russell 2001). Target sequences are amplified and quantified in the same PCR machine.

Internal standards are not required to quantify the amount of DNA or RNA present in real-time PCR. The ability to quantify the amplified DNA during the exponential phase of the PCR, when none of the reaction components are in limited amounts, results in improved precision in quantification of target sequences. Real-time PCR can measure the initial concentration of target DNA over a range of 5 or 6 orders of magnitude. At present, the limit of detection when fluorescent dyes are used is  $\approx$ 10–100 copies of template DNA in the starting reaction (Sambrook and Russell 2001). The TaqMan method of real-time PCR is described in Section 8.4.16.

Significant advantages of real-time PCR include its ability to measure DNA concentrations over a large range, its sensitivity, its ability to process multiple samples simultaneously, and its ability to provide immediate information. A disadvantage is that the machines are more expensive than traditional PCR machines. Real-time PCR is especially useful in large commercial laboratories that process a large number of samples of a similar type. Bustin et al. (2009) recommend guidelines that will improve the reliability of quantitative PCR data and allow others to reproduce the work.

#### 8.4.15 Reverse-Transcription PCR

Messenger RNA can be reverse transcribed and the resultant cDNA then can be amplified using *Taq* DNA polymerase. Reverse Transcriptase-PCR (RT-PCR) allows detection of gene expression in small numbers of specific cells or tissues. Reactions have been carried out with RNA isolated from as few as 10–1000 cells.

The process involves 1) isolation of mRNA, 2) reverse transcription of mRNA into cDNA, and 3) amplification of cDNA by DNA polymerase. Primers for the amplification should be 18–22 nt long, and should occur in separate exons to inhibit amplification of any contaminating genomic DNA in the RNA preparation. Multiple reverse transcriptases can be used in RT-PCR to increase sensitivity and product yield (Nevett and Louwrier 2000).

RT-PCR has been used to monitor for the presence of rabbit hemorrhagic disease virus in fly species in Australia (Asgari et al. 1998). This calcivirus causes a lethal disease in European rabbits, but little was known about how it spreads in the field. RT-PCR provided a sensitive and reliable method for detecting the virus in flies and fly spots, which allows it to be used to study the epizootiology and vector biology of the virus.

#### 8.4.16 TaqMan PCR

TaqMan PCR is a type of real-time PCR. TaqMan PCR uses a nucleic-acid probe complementary to an internal segment of the target DNA. The probe is labeled with two fluorescent moieties. The emission spectrum of one overlaps the excitation spectrum of the other, resulting in “quenching” of the first fluorophore by the second. The probe is present during the PCR and if product is made, the probe is degraded via the 5'-nuclease activity of *Taq* polymerase that is specific for DNA hybridized to template (=TaqMan activity). The degradation of the probe allows the two fluorophores to separate, which reduces quenching and increases intensity of the emitted light. Because this assay involves fluorescence measurements that can be performed without opening the PCR tube, the risk of contamination is greatly reduced. Furthermore, no electrophoresis is required, so labor costs are reduced (Kalinina et al. 1997, Sambrook and Russell 2001, Baker 2011).

#### 8.4.17 Digital PCR

Digital PCR (dPCR) involves diluting a sample of DNA and partitioning it into hundreds or even millions of separate reaction chambers so that each chamber contains one or no copies of the target sequence (Lo et al. 2007). After

conducting PCR, the number of positive and negative reactions is counted and this allows the scientist to estimate how many copies of DNA molecule were in the sample. This technique is more expensive to conduct than quantitative PCR, although the use of nanofabrication and microfluidics has reduced the amount of reagents required and, thus, the cost. Several dPCR machines are commercially available and the use of dPCR has been used to quantify the number of viruses in individual cells (Tadmor et al. 2011, White et al. 2012) or the differential expression of alleles (Pekin et al. 2011, Whale et al. 2012). dPCR provides a high resolution; you can distinguish between two and three copies using 200 chambers, but need 8000 chambers to distinguish between 10 and 11 copies. dPCR does not require the same level of calibration or controls as qPCR, but artifacts can occur if too many copies occur in the chambers (i.e., the DNA was not sufficiently diluted), inhibitors prevent reactions from occurring, and specificity can be poor if pseudogenes are present.

## 8.5 Some Research Applications

The PCR can be applied to a diverse array of both basic and applied problems (Table 8.6). Protocols for the different methods are available in books (Erlich 1989, Erlich et al. 1989, Innis et al. 1990, Ausubel et al. 1991, McPherson et al. 1991, Sambrook and Russell 2001, Gariepy et al. 2007) or individual journal papers.

The following examples provide evidence of the versatility of the PCR, but are only an abbreviated introduction to the diversity of applications to which this tool can be applied. Modifications of the PCR continue to be made to resolve diagnostic, ecological, evolutionary, genetic, and developmental biology questions.

### 8.5.1 Amplifying Ancient DNA

The film *Jurassic Park* implied it was possible to amplify dinosaur DNA from insects preserved in amber; this captured the imagination of the public and created a climate in which the PCR was perceived to be an unusually powerful key to analyzing the past. Subsequently, the PCR was used to amplify DNA fragments from a number of insects preserved in ancient amber. Unfortunately, these results have been controversial (Box 8.1), as have been the results of amplifying dinosaur DNA (Austin et al. 1997a,b; Rollo 1998; Hofreiter et al. 2001; Hebsgaard et al. 2005).

The most common ancient DNA amplified by the PCR is usually mitochondrial DNA because it is so abundant. However, this abundance makes it easy to contaminate the ancient sample with modern mitochondrial DNA. The amplification of ancient DNA remains highly controversial because technical

### Box 8.1 Amplifying Ancient DNA from Insects in Amber: Controversial Results?

Why the controversy? Is amber a special form of preservative that allows DNA to persist for unusually long periods (millions of years)? Amber entombs insect specimens completely, after which they completely dehydrate so the tissue is effectively mummified. The terpenoids, major constituents of amber, could inhibit microbial decay (Austin et al. 1997a). Certainly, preservation of amber-embedded insects seems to be exceptional and insect tissues in amber appear comparable in quality to the tissues of the frozen woolly mammoth (which is “only” 50,000 years old). But is the DNA in these tissues preserved and can it be amplified by the PCR?

Claims have been made that DNA can be extracted from insects in amber, including a fossil termite, *Mastotermes electrodominicus*, estimated to be 25–30 million years old (DeSalle et al. 1992); a 120- to 130-million-year-old conifer-feeding weevil (Coleoptera: Nemonychidae) (Cano et al. 1993a); and a 25- to 40-million-year-old bee (Cano et al. 1993b). These are extraordinary ages for DNA!

The DNA sequences obtained from all amber-preserved insects meet several, but not all, criteria of authenticity; the fossil DNA sequences “make phylogenetic sense” and DNA has been isolated from more than one specimen in several cases (although the weevil DNA was derived from a single specimen) (Austin 1997a).

Yet extraction and amplification of fossil DNA sequences from amber-preserved insects must be reproduced in independent laboratories, to meet quality criteria. This has cast doubt on the authenticity of the reports (Austin 1997a,b; Sykes 1997; Walden and Robertson 1997; Gutierrez and Marin 1998; Hofreiter et al. 2001).

One of the most controversial claims involved the isolation of a “living” bacterium from the abdomen of amber-entombed bee. Bacterial DNA from a 25-million-year-old bee was obtained and sequenced. Cano and Boruki (1995) reported reviving a bacterial spore, culturing it, and identifying it. The classification of the bacterium is controversial (Beckenbach 1995, Priest 1995) because the bacterium could have come from a currently undescribed species of the *Bacillus sphaericus* complex. The modern *B. sphaericus* complex is incompletely known, so the “ancient” sequence obtained could be that of a modern, but previously unidentified, bacterium. These bacteria often are isolated from the soil (Yousten and Rippere 1997).

Other claims of amplifying ancient DNA have been disproved. For example, the mitochondrial cytochrome *b* sequence of an 80-million-year-old dinosaur from the Upper Cretaceous in Utah was discovered to be, most probably, of human origin (Hedges and Schweitzer 1995). Likewise, a 20-million-year-old magnolia leaf produced sequences that were similar to those of modern magnolias. The authenticity of the magnolia sequences were cast into doubt because they were exposed to water and oxygen during preservation and DNA is especially vulnerable to degradation under such conditions. The jury is out on the authenticity of ancient DNA in insects embedded in amber. Fortunately, the scientific criteria for resolving the controversy are clear.

High-throughput sequencing machines (NextGen sequencing) work well with small DNA fragments in ancient bones, hair, and teeth. Next-generation sequencing has resulted in the entire genomes of an ancient cave bear, a mammoth, and the Neanderthal (Gibbons 2010).

difficulties are enormous (Rollo 1998, Cooper and Poinar 2000). Theoretical considerations indicate that maximal DNA survival of 50 thousand to 1 million years could be possible (Hebsgaard et al. 2005). In fact, most DNA decays relatively rapidly and amplifiable DNA rarely is found in fossils one hundred to a few thousand years old, and these are rarely longer than 100 bp unless the cadaver has been frozen.

DNA decays spontaneously, mainly through hydrolysis and oxidation. Hydrolysis causes deamination of the nucleotide bases and cleavage of base-sugar bonds, creating baseless sites. Deamination of cytosine to uracil and depurination (loss of purines adenine and guanine) are two types of hydrolytic damage. Baseless sites weaken the DNA, causing breaks that fragment the DNA into smaller and smaller pieces. Oxidation leads to chemical modification of bases and destruction of the ring structure of base and sugar residues (Austin 1997a). As a result, it is almost impossible to obtain long amplification products from ancient DNA (Handt et al. 1994). It is possible to use overlapping primer pairs if longer sequences are needed, but there usually is an inverse relationship between efficiency and length of the PCR products. When such an inverse relationship is *not* seen, the amplification product often turns out to be due to contamination (Handt et al. 1994, Hofreiter et al. 2001).

PCR products from ancient DNA often are “scrambled.” This is due to the phenomenon called “jumping PCR,” which occurs when the DNA polymerase reaches a template position which carries either a lesion or a strand break that stops the polymerase (Handt et al. 1994). The partially extended primer can anneal to another template fragment in the next cycle and be extended up to another damaged site. Thus, *in vitro* recombination (jumping) can take place until the whole stretch encompassed by the two primers is synthesized and the amplification enters the exponential part of the PCR (Handt et al. 1994). This phenomenon makes it essential that cloning and sequencing of multiple clones be carried out to eliminate this form of error.

Most archeological and paleontological specimens contain DNA from exogenous sources such as bacteria and fungi, as well as contaminating DNA from contemporary humans (Poinar and Stankiewicz 1999). Aspects of burial conditions seem to be important in DNA preservation, especially low temperature during burial (Poinar and Stankiewicz 1999). The oldest DNA sequences reported and confirmed by other laboratories are from the remains of a wooly mammoth found in the Siberian permafrost; these sequences are “only” 50,000 years old—not millions of years old (Poinar and Stankiewicz 1999).

More recently, ancient DNA has been sequenced using new Next-Generation sequencing methods because small fragments of (degraded) DNA are sequenced efficiently with these methods. Subsequent bioinformatic analysis allowed the scientists to recognize and sort out the short bits of ancient DNA from longer pieces of contaminant DNA. For example, the entire mammoth genome was sequenced from DNA extracted from mammoth hair (Miller et al. 2008) and the entire Neanderthal genome was sequenced in 2010 from DNA from bone fragments estimated to be ≈44,000 years old (Green et al. 2010). Thus, the PCR may not always be the most appropriate method for analysis of ancient DNA.

Theoretical calculations and empirical observations suggest DNA should only be able to survive, in a highly fragmented and chemically modified form, for 50,000–100,000 years (Austin et al. 1997a, Rollo 1998, Hofreiter et al. 2001). Because only tiny amounts of DNA usually can be extracted from an archeological specimen, stringent precautions and multiple controls are required to avoid accidental contamination with modern DNA.

A methodology to deal with ancient specimens has been proposed that includes careful selection of well-preserved specimens, choice of tissue samples that are likely to have the best DNA preservation, and surface sterilization to eliminate surface contamination (Hebsgaard et al. 2005). The operations should be carried out in a laboratory dedicated to work on ancient specimens and work on ancient DNA should be separated from that on modern DNA (Austin et al. 1997a, Cooper and Poinar 2000, Hofreiter et al. 2001). Most importantly, *multiple* negative controls should be performed during DNA extraction and PCR set up, although a lack of amplifications in the negative controls is not definitive proof of authentic ancient DNA.

Another consideration is the likelihood of amplifying nuclear copies of mitochondrial genes (den Tex et al. 2010). Universal primers are often used to amplify DNA from mitochondrial genes in ancient specimens. However, den Tex et al. (2010) found that nuclear copies were often amplified instead of mitochondrial DNA, especially when universal primers were used.

Another crucial step is the authentication of the results. Putatively ancient DNA sequences should be obtained from different extractions of the same sample and from different tissue samples from different specimens (Austin et al. 1997a, Cooper and Poinar 2000). The ultimate test of authenticity should be independent replication in two separate laboratories (Rollo 1998). So far, this type of replication has not been achieved for DNA from amber-preserved arthropod specimens (Austin et al. 1997b, Walden and Robertson 1997, Gutierrez and Marin 1998).

Thomsen et al. (2009) report a method for nondestructive sampling of ancient insect DNA, which allowed insects from a museum collected in 1820 to be amplified, but was less successful with samples from permafrost sediments that were 3280–1800 years old.

### 8.5.2 Amplifying Old DNA

Amplification of old DNA from museum specimens is less difficult and less controversial (Paabo 1990, 1991; Jackson et al. 1991; Cano et al. 1993a,b; Townson et al. 1999; Watts et al. 2007). DNA from pathogens contained within museum specimens of arthropods can be amplified by the PCR. For example, Lyme

disease spirochete (*Borrelia burgdorferi*) DNA extracted from the midgut of ticks (*Ixodes dammini*) stored for 50 years in 70% EtOH could be amplified by the PCR (Persing et al. 1990). Individual tick specimens were removed from the EtOH with flame-sterilized forceps and air-dried on filter paper disks for 5 minutes. Then, 200 µl of 0.5-mm glass beads were incubated in 1 ml of 1% bovine serum albumin in distilled water at 37 °C for 30 minutes and then washed twice in 1 ml of distilled water. Ticks were placed whole into 0.6-ml microcentrifuge tubes containing slurries (20 µl) of the treated glass beads. Specimens were crushed into the beads with a disposable plastic dowel for 30–45 seconds to liberate the midgut contents, and 25 µl of PCR buffer was added. Samples were boiled for 5 minutes, and then cooled on ice; 5-µl portions of the supernatant fluid were used for the PCR.

A simpler protocol was used by Azad et al. (1990) to determine whether individual ticks or fleas were infected with rickettsia. Individual ticks or fleas were placed in 100 µl of *brain heart* infusion broth and boiled for 10 minutes. The PCR was carried out with 10 µl of the suspension. Because the PCR can be applied to frozen or formalin-fixed tissues, dried museum specimens, and alcohol-preserved specimens, PCR can reduce potential dangers involved in maintaining and transporting live infectious disease vectors from the field. Detection of pathogens by the PCR is significantly more sensitive than enzyme-linked immunosorbent assay (ELISA) (Azad et al. 1990).

Dried, pinned specimens of the *Anopheles gambiae* mosquito complex, ranging in age from 15–93 years, were tested to determine whether ribosomal DNA could be amplified by the PCR (Townson et al. 1999). Most of the specimens yielded amplifiable DNA from entire abdomens, but extractions from single hind legs from these old, dried specimens were unsuccessful. By contrast, single legs from a fresh specimen produced sufficient DNA to yield a PCR product. Note, however, that ribosomal genes are present in high copy numbers.

The PCR has been used to amplify DNA from tissues preserved in formalin followed by paraffin embedding. Specimens up to 40 years old have yielded DNA up to 800 bp in length (Wright and Manos 1990). However, Watts et al. (2007) found that air-dried insect legs >50 years old had limited usefulness for studies of nuclear loci. The integrity of the DNA and the duration of fixation affect the length of the product that can be amplified.

### 8.5.3 Amplifying RNA

The PCR can be used to amplify messenger RNA sequences using reverse transcriptase (Kawasaki 1990). This process allows analysis of gene expression during

development, quantitation of mRNA from specific tissues, rearrangements of DNA during cell differentiation, and RNA processing.

#### ***8.5.4 Analysis of mRNA Polyadenylation***

Measurement of poly(A) tail length is important when studying mRNA stability. A simple PCR-based method has been developed ([Eguchi and Eguchi 2000](#)).

#### ***8.5.5 Cloning a Gene***

The PCR can generate microgram quantities of a specific DNA fragment and these can be cloned ([Scharf 1990](#)), although many products of the PCR are “recalcitrant to cloning” ([Sambrook and Russell 2001](#)). One reason is that several of the DNA polymerases used in the PCR have the ability to add a single, unpaired nucleotide at the 3' end of the PCR product (terminal transferase activity). The nucleotide added depends both on the adjacent base and on the particular polymerase used; for example, when the 3'-terminal base of the template DNA is cytidine, *Taq* will add an adenine to the end of the completed PCR product ([Sambrook and Russell 2001](#)).

One solution to this problem is to use the 3' to 5' exonuclease activity of bacteriophage T4 DNA polymerase or *Pfu* DNA polymerase to “polish” the ends of the PCR products that contain the added bases; the polished DNA fragments can then be phosphorylated by T4 kinase and cloned into a blunt-ended dephosphorylated vector ([Costa and Weiner 1994](#), [Costa et al. 1994](#)). Unfortunately, blunt-ended cloning is notoriously inefficient (10- to 100-fold less efficient than cloning with DNA fragments with cohesive termini). Furthermore, blunt-ended cloning allows no opportunity to direct the orientation of the fragment within the vector.

A second reason for potential difficulties in cloning a PCR product is that *Taq* (and perhaps other polymerases) can survive extraction with methods used to purify the PCR products ([Bennett and Molenaar 1994](#), [Sambrook and Russell 2001](#)). The residual polymerase and dNTPs may make it difficult to tailor the ends of the amplified DNA for cloning.

Currently, the most popular and efficient method for cloning PCR products involves a method that relies on ligation of cohesive ends of the PCR product and a vector ([Sambrook and Russell 2001](#)). *Taq* DNA polymerase typically adds an A at the 3' end of the product. If a plasmid vector is used with a protruding 3' T residue at each of its ends, cloning is more efficient because the T and A can base-pair. It is perhaps surprising that pairing of single bases is sufficient to increase efficiency of cloning, but it is estimated to be ≈50-fold more efficient than blunt-ended cloning.

Restriction sites can be inserted at the 5' end of each primer so that the amplified DNA can be cloned directly into a vector after digestion of the amplified DNA (Kaufman and Evans 1990). Because the restriction sites can be the same or different in the two primers, the researcher can tailor the ends of the PCR product to the specific vectors required for the project (Sambrook and Russell 2001). Various commercial kits allow direct cloning of PCR products.

The isolation of a gene requires some prior knowledge of the gene sequence (Clackson et al. 1991). If a probe (primer) is available from another species, genomic DNA can be screened by the PCR using standard or degenerate primers (McPherson et al. 1991, Clackson et al. 1991). This success of this approach was illustrated by the cloning of a sodium channel gene from *Drosophila* and the house fly, *Musca domestica*, using degenerate primers (Knipple et al. 1991). Several vertebrate sodium channel genes had been cloned and comparisons of the inferred amino-acid sequences of the alpha subunits of sodium channels from rat brain and rat skeletal muscle to that of the electric eel revealed a 70% homology when conservative substitutions were taken into account. Two sodium channel genes (*para* and DSC1) cloned from *Drosophila* were homologous to the vertebrate sodium channel genes. Using this information, it was possible to generate DNA primers to isolate a segment of the gene homologous to *para* from the house fly.

The PCR was performed on genomic house fly DNA using degenerate primers. The 5'-end primer consisted of a 256-fold degenerate sequence 20 nt long. The 3'-end primer consisted of a 64-fold degenerate sequence 21 nt long. (When using degenerate primers, the primer length may be shorter than the typical 25–30 nt.) Both had additional sequences appended to their 5' ends to provide a *Hind*III and *Xba*I restriction-enzyme recognition sequence, respectively, to facilitate cloning the amplification products. The PCR product was 104 bp, consisting of 87 bp of coding sequence plus the flanking sequences attached to the 5' ends of the primers. To confirm that the PCR-generated DNA was derived from the house fly, amplified DNA was labeled with  $^{32}\text{P}$  and used as a probe of genomic Southern blots containing digests of house fly, *Drosophila*, and mouse DNA. The only specific hybridization signal after high-stringency washing was to the fly DNA. The PCR products were cloned and sequenced, and the sequence isolated from the house fly differed from that of *Drosophila* at only 16 nucleotides (81.6% similarity). The substitutions, primarily in the third base of the codon, had no effect on the amino-acid sequence.

Doyle and Knipple (1991) subsequently used the same degenerate primers to amplify DNA from seven insects and an arachnid, including the tobacco budworm, *Heliothis virescens*; *Aedes aegypti*; the diamondback moth, *Plutella xylostella*; the gypsy moth, *Lymantria dispar*; the cabbage looper, *Trichoplusia ni*; the

Colorado potato beetle, *Leptinotarsa decemlineata*; the American cockroach, *Periplaneta americana*; and the twospotted spider mite, *Tetranychus urticae*. After amplification, the PCR products were sequenced, and only five of 60 clones were not derived from *para* homologs. This study, and others, suggests that degenerate primers derived from conserved segments of characterized *D. melanogaster* genes can be used to clone genes from diverse arthropods. Interestingly, intraspecific polymorphisms were found in the sequence from three moth species, which could reflect the presence of duplicated genes or allelic variants in the populations.

Similarly, primers for the conserved *Actin* gene(s) in insects were used to clone these genes from the predatory mite *Metaseiulus occidentalis*, despite the long evolutionary separation of insects and mites (Hoy et al. 2000). The rich source of sequence information in GenBank for the complete *Drosophila* genome and those of other species makes this approach increasingly feasible.

#### **8.5.6 Detecting Gene Amplification**

Sometimes it is important to determine whether a gene has been amplified (increased in copy number), leading to increased levels of gene product. Some insects are resistant to pesticides due to amplification of esterase genes. A method called comparative PCR can be used to detect gene amplification (Brass et al. 1998).

#### **8.5.7 Detecting Methylation of DNA**

Genomic imprinting is often due to DNA methylation at several sites in the genome. A methylation-specific PCR assay can be used to detect methylation of specific genes more quickly than the use of Southern-blot assays (Kubota et al. 1997).

#### **8.5.8 Detecting Pathogens in Vector Arthropods**

Arthropod vectors such as ticks, fleas, and mosquitoes are involved in maintaining and transmitting (vectoring) pathogens to humans and other vertebrates. Aphids and leafhoppers transmit (vector) viruses and mycoplasma to plants.

The detection of pathogenic microorganisms within vector arthropods is important in conducting epidemiological studies and developing control strategies. Several antigen-detection techniques have been developed, including direct or indirect immunofluorescence tests and enzyme-linked immunosorbent assays (ELISA) using polyclonal or monoclonal antibodies. Other techniques involve recovery of the microorganisms from vectors by culture in embryonated eggs or tissue culture cells or by experimental infections in laboratory animals. The recovery of pathogenic microorganisms by these methods requires

either live or properly frozen specimens. These techniques are expensive, time-consuming, and may involve dissection and preparation of specimens from live arthropods, which can be hazardous (Barker 1994).

The PCR offers another approach to detecting and identifying pathogenic microorganisms if sequence information is available to design appropriate primers (Wise and Weaver 1991, Higgins and Azad 1995, Crowder et al. 2010). The PCR can be carried out with material from dead specimens, is more sensitive than most immunological techniques, and is more rapid. For example, using primers that amplify a 434-bp fragment of a protein from *Rickettsia rickettsii*, infected fleas and ticks have been identified (Azad et al. 1990). Malarial DNA has been detected in both infected blood and individual mosquitoes (Schriefer et al. 1991). Using seminested PCR, as few as three *Leishmania* parasites could be detected in infected sand flies (Aransay et al. 2000). Trypanosome infections could be detected in wild tsetse flies in Cameroon (Morlais et al. 1998). The heartwater fever pathogen *Cowdria ruminantium* could be detected in vector ticks (*Amblyomma*) with high levels of specificity when  $10^7$  to  $10^4$  organisms were present. The reliability of the assay dropped when ticks had only  $10^3$  to  $10^2$  organisms, which highlights the need to conduct quantitative analyses for sensitivity before using PCR assays in epidemiological studies (Peter et al. 2000).

West Nile virus was detected in human clinical specimens, field-collected mosquitoes, and bird samples by a TaqMan reverse-transcriptase PCR assay (Lanciotti et al. 2000). This rapid, specific, and sensitive assay can be used in the diagnostic laboratory for testing humans and as a tool for conducting surveillance of West Nile virus in mosquitoes and birds in the field (Anderson et al. 1999). Sequencing of the West Nile virus causing encephalitis in the northeastern United States indicated the virus was most closely related to a virus isolated in Israel in 1998 (Lanciotti et al. 1999).

A quantitative PCR protocol was used to assay densities of the plague bacteria *Yersinia pestis* in fleas and mice. The assays indicated fleas needed  $\approx 10^6$  bacteria to be able to transmit the bacteria to mice (Engelthaler et al. 2000). Random primers (hexanucleotides) were used to develop primers for a multiplex reverse-transcriptase PCR to detect five potato viruses and a viroid in aphids, leaves, and potato tubers (Lie and Singh 2001). Hamiduzzaman et al. (2010) developed a multiplex PCR assay to diagnose and quantify *Nosema apis* and *N. ceranae* infections in honey bees.

### 8.5.9 Detecting Pesticide Resistance

The malaria vectors *Anopheles gambiae* and *An. arabiensis* were screened for permethrin resistance (nerve-insensitivity, *kdr*-type) (Brooke et al. 1999).

The results indicated that one of the populations was resistant to permethrin through a different biochemical method, indicating that both PCR and bioassay data should be obtained for monitoring resistance-allele frequencies and the mechanism(s) of insecticide resistance.

### 8.5.10 Developmental Biology

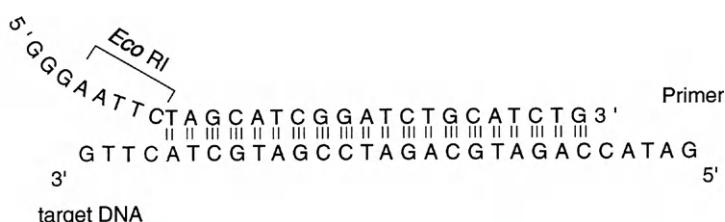
It is possible to detect the presence of specific mRNAs in tissues or cells by reverse transcription and DNA amplification by the PCR.

### 8.5.11 Engineering DNA

DNA can be engineered in several ways by the PCR. Sequences can be added to the 5' end of primers. Such sequence changes are readily accepted, even though these add-ons do not base-pair with the template DNA. The DNA being synthesized contains the add-on because the primers are incorporated in the synthesized DNA fragment. For example, it is possible to add a restriction site sequence to DNA being amplified by the PCR by attaching the restriction site sequence to the primers (Figure 8.7). Such restriction sites facilitate subsequent manipulations of the final PCR product.

The T7 promoter located at the 5' end of one primer can be added to PCR products. This promoter allows RNA copies to be generated from the DNA synthesized in the PCR reaction. Although the add-on sequences in the primers don't base-pair to the template DNA, in most cases they have little effect on the specificity or efficiency of the amplification. The 3' end of the primer apparently is most important in imparting specificity.

One PCR product strand or both can be tagged with a radioactive, biotin, or fluorescent label (Chehab and Kan 1989, Mertz and Rashtchian 1994). DNA



**Figure 8.7** A 5' add-on of a restriction site sequence (*Eco*RI) to a primer, which is annealed to a target DNA sequence. Although this add-on does not specifically match the template DNA, this does not significantly affect the PCR. The extra bases that are added 5' to the *Eco*RI site ensure that the efficiency of the restriction enzyme cleavage is maintained.

sequences also can be altered at any position by modifying primers so that substitutions, additions, or deletions are made in the amplified DNA.

#### ***8.5.12 Evaluate Efficacy of Disease Control***

In regions of the world where malaria is endemic, the use of bed nets impregnated with a pyrethroid insecticide has been used as a method to control the mosquito vector (Gokool et al. 1992). The PCR was used to determine whether pesticide-treated bed nets prevent mosquito transmission of malaria. PCR was used to amplify and fingerprint the human DNA contained within a mosquito blood meal. DNA fingerprints also were obtained from the blood of individuals sleeping under the bed nets and compared to the fingerprints from the mosquitoes (Gokool et al. 1992). The many hypervariable regions of the human genome produce individual-specific patterns of DNA fragments, and the banding patterns obtained indicated that few mosquitoes had fed on individuals protected by treated bed nets.

#### ***8.5.13 Evolutionary Analyses***

Analysis of evolution involves reconstruction of phylogenies and analysis of population genetics (for further details, see Chapters 12 and 13). The relative ease and simplicity of the PCR enhances molecular studies of evolution. The PCR makes it possible to directly sequence amplified gene fragments from individuals and populations, increasing the resolving power and phylogenetic range of comparative studies. For most applications of the PCR, it is necessary to know a sequence to synthesize primers. However, by choosing sequences that are highly conserved among widely divergent species it is possible to design “universal primers” to amplify a particular nuclear or organelle gene fragment from many members of a major taxonomic group. This allows comparisons of sequences from classes or phyla for taxonomic work, as well as enhancing population ecology studies that involve identifying individuals and biotypes.

Universal primers have been developed for a number of nuclear and mitochondrial genes. For example, primers that amplify a region of  $\approx 515$  bp of the 18S rDNA from many fungi, protozoa, algae, plants, and animals are available. The primers are based on conserved sequences among the 18S rDNA, but they do not amplify bacterial or mitochondrial rRNA genes.

Mitochondrial gene sequences are useful for many studies in evolutionary and population biology (Kapsa et al. 1997). Primers have been developed that allow a number of different gene fragments to be amplified from different insect orders (Simon et al. 1994, Kambhampati and Smith 1995). It even may be

possible to amplify complete insect mitochondrial genomes in “two easy pieces” (Roehrdanz 1995). Because mitochondria are inherited maternally, sequence analyses of mitochondrial DNA allow construction of maternal phylogenies. Mitochondrial DNA evolves at a higher rate than nuclear DNA because mitochondria lack a proofreading function to correct errors in DNA synthesis; this makes mitochondria especially useful for analyses of closely related populations or species.

#### 8.5.14 Sequencing DNA

Both the Maxim–Gilbert and Sanger sequencing methods have been modified to sequence DNA amplified by the PCR (Ausubel et al. 1991; Ellingboe and Gyllensten 1992; Kocher 1992; Olsen et al. 1993; Rao 1994a,b). Both methods permit the rapid determination of sequences without the need to construct a library or screen the library for the gene(s) of interest. Sequencing of PCR products can be either direct (Landweber and Kreitman 1995) or after amplification and cloning (Olsen et al. 1993, Sambrook and Russell 2001). Kits are available that aid in sequencing by various methods.

Direct sequencing of ds PCR products should, in theory, save time and effort. Unfortunately, it can be “...unreliable in practice” (Sambrook and Russell 2001). Direct sequencing of PCR products can only be successful if the PCR was optimized to reduce mispriming and the PCR product was cleaned of residual primers, DNA polymerase, unused dNTPs, and nonspecific copies of the original template (Sambrook and Russell 2001). Effective cleaning of the PCR product may be achieved through spun column chromatography or centrifugal ultrafiltration to remove residual primers and unused dNTPs. Residual thermostable DNA polymerase and residual dNTPs may “befoul chain-termination sequencing reactions” catalyzed by other thermostable enzymes, such as Sequenase and AmpliTaq, unless eliminated (Sambrook and Russell 2001). Running the DNA on a low-melting-point agarose gel can eliminate nonspecific PCR products by separating the band of the appropriate size from the gel.

Direct sequencing of DNA amplified by a PCR protocol that uses only a single primer (such as RAPD-PCR) requires a different procedure (Iizuka et al. 1996). If ss DNAs first are isolated by polyacrylamide gel electrophoresis, sequencing can be conducted on the single strands using the corresponding single primers.

Direct sequencing of PCR-amplified DNA is not appropriate if the starting sample contained one normal allele and one deleted allele because the deleted allele will be masked. Likewise, if the sample contains multiple alleles, direct sequencing would result in a composite sequence ladder, making it impossible

to decipher the sequence of individual alleles (Rao 1994a). Under these circumstances, the amplified DNA should be cloned and the sequences of many clones should be determined to resolve the sequences of individual alleles.

Direct sequencing of PCR-amplified DNA is inappropriate if nonspecific products were produced. Thus, direct sequencing requires good PCR conditions: quality DNA template, highly specific primers, high annealing temperature, initiating the PCR by the “hot-start” method, and performing a rapid cycling protocol for as few cycles as possible using an adequate amount of target DNA (Rao 1994a).

## 8.6 Multiple Displacement Amplification: Another Method to Amplify DNA

Multiple displacement amplification (MDA) is a non-PCR-based DNA amplification method that can amplify very small amounts of DNA, even from a single cell (Cheung et al. 1996, Dean et al. 2001, 2002, Gorrochotegui-Escalante and Black 2003, Hosono et al. 2003, Jeyaprakash and Hoy 2004, Panelli et al. 2006). After amplification by MDA, the PCR can be used to further amplify specific genes. MDA can amplify genomic DNA within a few hours without a thermal cycler at 30 °C using exonuclease-resistant thiophosphate-modified hexamers (random nucleotides) as primers and bacteriophage Phi29 DNA polymerase, which is a high-fidelity enzyme. It is thought that the hexamers bind at random over the genome, allowing Phi29 DNA polymerase to synthesize DNA strands up to 7–10 kb, enriching each amplified DNA strand by approximately 10,000-fold in ~3 hours (Dean et al. 2001, 2002, Hosono et al. 2003, Trannah et al. 2003).

MDA can be used to amplify DNA from single adult legs or larvae of a mosquito, producing sufficient DNA for subsequent amplification by the PCR or analysis by the Southern blot (Gorrochotegui-Escalante and Black 2003). Whole genome amplification (WGA) can be used to amplify microbial DNA from arthropods; Jeyaprakash and Hoy (2004) produced sufficient DNA from *Wolbachia* or *Cardinium* bacteria in a predatory mite egg that subsequent amplification by high-fidelity PCR was successful. A sensitivity analysis suggested that WGA followed by the high-fidelity PCR was able to detect as little as 0.01 fg of the target DNA, equivalent to about one copy. Kumar et al. (2008) found that MDA amplification of single human cells resulted in high-quality DNA with few artifacts.

## 8.7 Concluding Remarks

The speed, specificity, versatility, and sensitivity of the PCR have had a significant effect on genetics, immunology, forensic science, evolutionary biology, systematics, ecology, and population biology (Arnheim and Erlich 1992, Dieffenbach 1995).

The PCR has revolutionized the way in which much of our research is conducted. However, the speed and ease of using the PCR and MDA can also result in erroneous conclusions (Box 8.2).

End users and commercial developers continue to develop sophisticated techniques for integrating the PCR into more applications (Waters et al. 1998, Smyth et al. 2010). For example, it has long been a goal to reduce the time to obtain results in conventional PCR machines. Thermal cycler heating and cooling rates have constrained this goal and typical reactions take an hour or more. However, Wheeler et al. (2011) found that a new thermocycler could amplify DNA in less than three minutes using two commercially available enzymes; the ultrafast PCR machine has a superior heat transfer due to convective flow through a porous media substrate, allowing very short cycle times.

### Box 8.2 A Cautionary Tale of the Insidious Nature of Contamination and the Sensitivity of the PCR and Whole Genome Amplification

A brief history of analyses of *Wolbachia* in the predatory mite *Metaseiulus occidentalis* will illustrate some of the issues associated with advances in PCR methods, the high degree of sensitivity of the PCR and whole genome amplification (WGA), as well as the possibility of contamination leading to false conclusions.

This predatory mite was known to have populations that were partially incompatible with others and to contain two microbial symbionts, although the bacteria were not identified at that time. Once *Wolbachia* was discovered and 16S PCR primers became available, positive results were obtained using an allele-specific PCR protocol, leading us to conclude that *M. occidentalis* was infected with *Wolbachia*, although not all reactions were positive. Furthermore, heat treatment resulted in strains that appeared to lack *Wolbachia* based on negative PCR results and, when “*Wolbachia*-negative” females were mated with males positive for *Wolbachia*, they exhibited reproductive incompatibility. However, the two-spotted spider mite *Tetranychus urticae*, the prey fed to these predators, also were infected with *Wolbachia* and sequencing of 16S sequences from both predator and prey indicated they had identical sequences. This led to speculation that horizontal transfer could have occurred between predator and prey. Subsequent PCR analyses using *wsp* sequences also indicated both predator and prey had nearly identical sequences, supporting this hypothesis. Furthermore, although we starved the predators for 6 to 8 hours prior to extracting their DNA, they still retained the *Wolbachia* sequences. However, PCRs of groups of predator eggs gave inconsistent results; some were positive and some were negative, even using the highly sensitive high-fidelity PCR protocol. We expected all eggs to be positive if transovarial transmission was involved, as it is with other *Wolbachia* infections of arthropods. When whole genome amplification became available, we used this technique to amplify *Wolbachia* sequences prior to conducting high-fidelity PCR analyses on single eggs. The results were again variable; some eggs were positive and some were negative (Jeyaprakash and Hoy 2004). During this time, *Wolbachia* was the only symbiont known to cause reproductive incompatibility.

*Cardinium*, another bacterium associated with reproductive incompatibility, subsequently was discovered, so we tested *M. occidentalis* and *T. urticae* and discovered that *M. occidentalis* was infected with *Cardinium* but their *T. urticae* prey was not. During these studies, surface decontamination of the mites to eliminate contaminating DNA was not done. However, when bleach was used to surface-decontaminate eggs and adults of *M. occidentalis*, combined with starvation for more than 72 hours, the positive high-fidelity PCR signal for *Wolbachia* disappeared. This indicated that previous detection of *Wolbachia* in *M. occidentalis* was

likely caused by contamination by minute amounts of *Wolbachia* DNA on *M. occidentalis* surfaces or by undigested, prey-derived *Wolbachia* present in the guts due to insufficient starvation periods used in prior studies (Wu and Hoy 2012a). Furthermore, when a strain of *M. occidentalis* was discovered to lack *Cardinium* naturally and the appropriate crosses were conducted, reproductive incompatibility was observed, indicating that *Cardinium* was the likely agent of the incompatibility rather than *Wolbachia* (Wu and Hoy 2012b).

The data illustrate several points: predators can retain prey (and the prey symbiont DNA) in their guts for a long time, resulting in false positives in their predator. Thus, when studying symbionts that appear to occur both in predator and prey (and parasitoid and host?), a long series of starvation times should be tested using sensitive PCR protocols. Because WGA and high-fidelity PCR are so sensitive (able to detect as few as one or two DNA molecules), PCR can pick up very small amounts of contamination. The false-positive egg data probably are due to external contamination of the eggs by *T. urticae* *Wolbachia* DNA. Resolving the sensitivity of your PCRs with sensitivity analyses (PCR amplification of cloned gene fragments mixed with arthropod DNA that has been serially diluted) will clarify just how sensitive your PCR is (Jeyaprakash and Hoy 2000). Failure to know about the presence of other putative symbionts, however, also can result in erroneous conclusions. Surveys for symbionts in both predator and prey (or host and parasitoids) should be conducted so that function(s) of the symbionts can be attributed correctly. For example, Plantard et al. (2012) recently found that *Wolbachia* thought to occur in a tick was actually in a parasitoid of the tick. Finally, external decontamination of predators should be conducted to reduce false positives.

## References Cited

- Agusti, N., de Vicente, M.C., and Gabarra, R., (2000). Developing SCAR markers to study predation on *Trialeurodes vaporariorum*. *Insect Mol. Biol.* **9**: 263–268.
- Aljanabi, S.M., and Martinez, I., (1997). Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. *Nucleic Acids Res.* **25**: 4692–4693.
- Al-Soud, W.A., and Radstrom, P., (1998). Capacity of nine thermostable DNA polymerases to mediate DNA amplification in the presence of PCR-inhibiting samples. *Appl. Environ. Microbiol.* **64**: 3748–3753.
- Anderson, J.F., Andreadis, T.G., Vossbrinck, C.R., Tirrell, S., Wakem, E.M., and French, R.A., et al. (1999). Isolation of West Nile virus from mosquitoes, crows and a Cooper's hawk in Connecticut. *Science* **286**: 2331–2333.
- Aransay, A.M., Scoulica, E., and Tselenitis, Y., (2000). Detection and identification of *Leishmania* DNA within naturally infected sand flies by seminested PCR on minicircle kinetoplastid DNA. *Appl. Environ. Microbiol.* **66**: 1933–1938.
- Arnheim, N., and Erlich, H., (1992). Polymerase chain reaction strategy. *Annu. Rev. Biochem.* **61**: 131–156.
- Arnheim, N., White, T., and Rainey, W.E., (1990). Application of PCR: organismal and population biology. *BioScience* **40**: 174–182.
- Asgari, S., Hardy, J.R.E., Sinclair, R.G., and Cooke, B.D., (1998). Field evidence for mechanical transmission of rabbit haemorrhagic disease virus (RHDV) by flies (Diptera: Calliphoridae) among wild rabbits in Australia. *Virus Res.* **54**: 123–132.
- Austin, A.D., and Dillon, N., (1997). Extraction and PCR of DNA from parasitoid wasps that have been chemically dried. *Aust. J. Entomol.* **36**: 241–244.
- Austin, J.J., Smith, A.B., and Thomas, R.H., (1997a). Palaeontology in a molecular world: the search for authentic ancient DNA. *Trends Evol. Ecol.* **12**: 303–306.
- Austin, J.J., Ross, A.J., Smith, A.B., Fortey, R.A., and Thomas, R.H., (1997b). Problems of reproducibility—does geologically ancient DNA survive in amber-preserved insects? *Proc. R. Soc. Lond. B* **264**: 467–474.

- Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.M. Seldman, J.G. and Smith, J.A. Eds., (1991). *Current Protocols in Molecular Biology*, Greene Publishing, New York.
- Ayliffe, M.A., Lawrence, G.J., Ellis, J.G., and Pryor, A.J., (1994). Heteroduplex molecules formed between allelic sequences cause nonparental RAPD bands. *Nucleic Acids Res.* **22**: 1632–1636.
- Ayyadevara, S., Thaden, J.J., and Shmookler Reis, R.J., (2000). Discrimination of primer 3'-nucleotide mismatch by *Taq* DNA polymerase during polymerase chain reaction. *Analyt. Biochem.* **284**: 11–18.
- Azad, A.F., Webb, L., Carl, M., and Dasch, G.A., (1990). Detection of rickettsiae in arthropod vectors by DNA amplification using the polymerase chain reaction, pp. 557–563. In: *Rickettsiology: Current Issues and Perspectives, Volume 590*. Annals of the New York Academy of Sciences, The New York Academy of Sciences, New York.
- Baker, M., (2011). qPCR: quicker and easier but don't be sloppy. *Nat. Methods* **8**: 207–212.
- Baldrich, E., Xamena, N., and Cabre, O., (1999). Overcoming false negatives due to the genomic context in polymerase chain reaction amplification. *J. Biochem. Biophys. Methods* **40**: 45–48.
- Barker Jr., R.H., (1994). Use of PCR in the field. *Parasitol. Today* **10**: 117–119.
- Barnes, W.M., (1994). PCR amplification of up to 35-kb DNA with high fidelity and high yield from  $\lambda$  bacteriophage templates. *Proc. Natl. Acad. Sci. USA* **91**: 2216–2220.
- Beckenbach, A.T., (1995). Age of bacteria from amber. *Science* **270**: 2015–2016.
- Bennett, B.L., and Molenaar, A.J., (1994). Cloning of PCR products can be inhibited by *Taq* polymerase carryover. *BioTechniques* **16**: 32–33.
- Brass, N., Heckel, D., and Meese, E., (1998). Comparative PCR: An improved method to detect gene amplification. *BioTechniques* **24**: 22–26.
- Brooke, B.D., Hunt, R.H., Koekemoer, L.L., Dossou-Yovo, J., and Coetzee, M., (1999). Evaluation of a polymerase chain reaction assay for detection of pyrethroid insecticide resistance in the malaria vector species of the *Anopheles gambiae* complex. *J. Am. Mosq. Control Assoc.* **15**: 565–568.
- Brownie, J., Shawcross, S., Theaker, J., Whitcombe, D., Ferrie, R., and Newton, C., et al. (1997). The elimination of primer-dimer accumulation in PCR. *Nucleic Acids Res.* **25**: 3235–3241.
- Burt, A., Carter, D.A., White, T.J., and Taylor, J.W., (1994). DNA sequencing with arbitrary primer pairs. *Mol. Ecol.* **3**: 523–535.
- Bustin, S.A., (2004). *A-Z of Quantitative PCR (IUL Biotechnology, No. 5)*. International University Line, La Jolla, CA.
- Bustin, S.A., Benes, V., Garson, J.A., Hellmann, J., Huggett, J., and Kubista, M., et al. (2009). The MIQE Guidelines: Minimum Information for publication of Quantitative real-time PCR Experiments. *Clin. Chem.* **55**: 611–622.
- Cano, R.J., and Borucki, M.K., (1995). Revival and identification of bacterial spores in 25- to 40-million-year-old Dominican amber. *Science* **268**: 1060–1063.
- Cano, R.J., Poinar, H.N., Pieniazek, N.J., Acra, A., and Poinar, G.O., (1993a). Amplification and sequencing of DNA from a 120–135-million-year-old weevil. *Nature* **363**: 536–538.
- Cano, R.J., Poinar, H.N., Roubik, D.W., and Poinar, G.O., (1993b). Enzymatic amplification and nucleotide sequencing of portions of the 18s rRNA gene of the bee *Proplebeia dominicana* (Apidae: Hymenoptera) isolated from 25–40 million year old Dominican amber. *Med. Sci. Res.* **20**: 619–622.
- Carbonari, M., Sbarigia, D., Cibati, M., and Fiorilli, M., (1993). Optimization of PCR performance. *Trends Genet.* **9**: 42–43.
- Chan, C.K., Petersen, D.J., and Vrain, T.C., (1999). DNA fingerprinting of single aphid embryos by random amplified polymorphic DNA polymerase chain reaction (RAPD-PCR). *Can. Entomol.* **131**: 229–230.
- Chehab, F.F., and Kan, Y.W., (1989). Detection of specific DNA sequences by fluorescence amplification: a color complementation assay. *Proc. Natl. Acad. Sci. USA* **86**: 9178–9182.
- Cheng, S., (1995). Longer PCR amplifications. In: Innis, M.A., Gelfand, D.H., and Sninsky, J.J., Eds., *PCR Strategies*, Academic Press, San Diego, CA, pp. 313–324.

- Cheng, S., and L.A. Kolmodin (1998). XL PCR amplification of long targets from genomic DNA, pp. 17–29. In: *Methods in Molecular Biology. Volume 67: PCR Cloning Protocols: From Molecular Cloning to Genetic Engineering*. B. A. White, Ed., Humana Press, Totowa, NJ.
- Cheng, S., Fockler, C., Barnes, W.M., and Higuchi, R., (1994a). Effective amplification of long targets from cloned inserts and human genomic DNA. *Proc. Natl. Acad. Sci. USA* **91**: 5695–5699.
- Cheng, S., Chang, S.Y., Gravitt, P., and Respass, R., (1994b). Long PCR. *Nature* **369**: 684–685.
- Cheung, V.G., and Nelson, S.F., (1996). Whole genome amplification using a degenerate oligonucleotide primer allows hundreds of genotypes to be performed on less than one nanogram of genomic DNA. *Proc. Natl. Acad. Sci. USA* **93**: 14676–14679.
- Chevret, E., Lemaitre, G., and Katinka, M.D., (1995). Low concentrations of tetramethylammonium chloride increase yield and specificity of PCR. *Nucleic Acids Res.* **23**: 3343–3344.
- Clackson, T., Gussow, D., and Jones, P.T., (1991). General application of PCR to gene cloning and manipulation. In: McPherson, M.J., Quirke, P., and Taylor, G.R., Eds., *PCR. A Practical Approach*, IRL Press, Oxford, UK, pp. 187–214.
- Collasius, M., Puchta, H., Schlenker, S., and Valet, G., (1991). Analysis of unknown DNA sequences by polymerase chain reaction (PCR) using a single specific primer and a standardized adaptor. *J. Virol. Methods* **32**: 115–119.
- Comes, A.M., Humbert, J.F., and Laurent, F., (1997). Rapid cloning of PCR-derived RAPD probes. *BioTechniques* **23**: 211–212.
- Cooper, A., and Poinar, H.N., (2000). Ancient DNA: do it right or not at all. *Science* **289**: 1139.
- Costa, G.L., and Weiner, M.P., (1994). Polishing with T4 or *Pfu* polymerase increases the efficiency of cloning of PCR fragments. *Nucleic Acids Res.* **22**: 2423.
- Costa, G.L., Grafsky, A., and Weiner, M.P., (1994). Cloning and analysis of PCR-generated DNA fragments. *PCR Methods Appl.* **2**: 338–345.
- Cox, R.A., (1968). The use of guanidinium chloride in the isolation of nucleic acids. *Methods Enzymol.* **12B**: 120–129.
- Crowder, C.D., Bounds, M.A., Phillipson, C.A., Picuri, J.M., Matthews, H.E., and Halverson, J., et al. (2010). Extraction of total nucleic acids from ticks for the detection of bacterial and viral pathogens. *J. Med. Entomol.* **47**: 89–94.
- Czerny, T., (1996). High primer concentration improves PCR amplification from random pools. *Nucleic Acids Res.* **24**: 985–986.
- Dean, F.B., Nelson, J.R., Giesler, T.L., and Lasken, R.S., (2001). Rapid amplification of plasmid and phage DNA using Phi29 DNA polymerase and multiply-primed rolling circle amplification. *Genome Res.* **11**: 1095–1099.
- Dean, F.G., Ghosono, S., Fang, L., Wu, X., Faruqi, A.F., and Bray-Ward, P., et al. (2002). Comprehensive human genome amplification using multiple displacement amplification. *Proc. Natl. Acad. Sci. USA* **99**: 5261–6266.
- den Tex, R.J., Maldonado, J.E., Thorington, R., and Leonard, J.A., (2010). Nuclear copies of mitochondrial genes: another problem for ancient DNA. *Genetica* **138**: 979–984.
- DeSalle, R., Gatesy, J., Wheeler, W., and Grimaldi, D., (1992). DNA sequences from a fossil termite in Oligo-Miocene amber and their phylogenetic implications. *Science* **257**: 1933–1936.
- Desmarais, E., Lanneluc, I., and Lagnel, J., (1998). Direct amplification of length polymorphisms (DALP), or how to get and characterize new genetic markers in many species. *Nucleic Acids Res.* **26**: 1458–1465.
- Dick, M., Bridge, D.M., Wheeler, W.C., and DeSalle, R., (1993). Collection and storage of invertebrate samples. *Methods Enzymol.* **224**: 51–65.
- Dieffenbach, C.W., Ed., (1995). *PCR Primer: A Laboratory Manual* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Dieffenbach, C.W., and Dveksler, G.S., (2003). *PCR Primer: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Dorak, M.T., Ed., (2006). *Real-Time PCR (Advanced Methods)*. Taylor & Francis, Oxford, UK.

- Doyle, K.E., and Knipple, D.C., (1991). PCR-based phylogenetic walking: isolation of para-homologous sodium channel gene sequences from seven insect species and an arachnid. *Insect Biochem.* **21**: 689–696.
- Dwyer, D.E., and Saksena, N., (1992). Failure of ultra-violet irradiation and autoclaving to eliminate PCR contamination. *Mol. Cell. Probes* **6**: 87–88.
- Ebert, P.R., (1996). Isolation of megabase-sized DNA fragments from the honey bee (*Apis mellifera*) genome. *J. Apic. Res.* **35**: 124–125.
- Eckert, K.A., and Kunkel, T.A., (1991). The fidelity of DNA polymerases used in the polymerase chain reactions. In: McPherson, M.J., Quirke, P., and Taylor, G.R., Eds., *PCR. A Practical Approach* IRL, Oxford, UK, pp. 225–245.
- Edwards, O.W., and Hoy, M.A., (1993). Polymorphisms in two parasitoids detected using random amplified polymorphic DNA (RAPD) PCR. *Biol. Control* **3**: 243–257.
- Edwards, O.W., and Hoy, M.A., (1994). Monitoring laboratory and field biotypes of the walnut aphid parasite, *Trioxys pallidus*, in population cages using RAPD-PCR. *Biocontrol Sci. Technol* **5**: 313–327.
- Eguchi, Y., and Eguchi, T., (2000). PCR-based method for the rapid analysis of mRNA polyadenylation. *Biotechnol. Lett.* **22**: 1583–1586.
- Ellingsboe, J., and Gyllensten, U.B., Eds., (1992). *The PCR Technique: DNA Sequencing* Eaton Publishing, Natick, MA.
- Ellsworth, D.L., Rittenhouse, K.D., and Honeycutt, R.L., (1993). Artifactual variation in randomly amplified polymorphic DNA banding patterns. *BioTechniques* **14**: 214–217.
- Engelke, D.R., Krikos, A., Bruck, M.E., and Ginsburg, D., (1990). Purification of *Thermus aquaticus* DNA polymerase expressed in *Escherichia coli*. *Analyt. Biochem.* **191**: 396–500.
- Engelthaler, D.M., Hinnebusch, B.J., Rittner, C.M., and Gage, K.L., (2000). Quantitative competitive PCR as a technique for exploring flea-*Yersina pestis* dynamics. *Am. J. Trop. Med. Hyg.* **62**: 552–560.
- Erlich, H.A., Ed., (1989). *PCR Technology, Principles and Applications of DNA Amplification*. Stockton Press, New York.
- Erlich, H.A., and Arnheim, N., (1992). Genetic analysis using the polymerase chain reaction. *Annu. Rev. Genetics* **26**: 479–506.
- Erlich, H.A., Gibbs, R., and Kazazian Jr., H.H., Eds., (1989). *Polymerase Chain Reaction. Current Communications in Molecular Biology*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Erlich, H.A., Gelfand, D., and Sninsky, J.J., (1991). Recent advances in the polymerase chain reaction. *Science* **252**: 1643–1651.
- Espinasa, L., and Borowsky, R., (1998). Evolutionary divergence of AP-PCR (RAPD) patterns. *Mol. Biol. Evol.* **15**: 408–414.
- Foley, K.P., Leonard, M.W., and Engel, J.D., (1993). Quantitation of RNA using the polymerase chain reaction. *Trends Genet.* **9**: 380–385.
- Forneck, A., Walker, M.A., and Blaich, R., (2000). Genetic structure of an introduced pest, grape phylloxera (*Daktulosphaira vitifoliae* Fitch), in Europe. *Genome* **43**: 669–678.
- Frothingham, R., Blitchington, R.B., Lee, D.H., Greene, R.C., and Wilson, K.H., (1992). UV absorption complicates PCR decontamination. *BioTechniques* **13**: 208–210.
- Fukatsu, T., (1999). Acetone preservation: a practical technique for molecular analysis. *Mol. Ecol.* **8**: 1935–1945.
- Fukatsu, T., and Ishikawa, H., (1994). Differentiation of aphid clones by arbitrarily primed polymerase chain reaction (AP-PCR) DNA fingerprinting. *Mol. Ecol* **3**: 187–192.
- Gariepy, T.D., Kuhlmann, U., Gillott, C., and Erlandson, M., (2007). Parasitoids, predators and PCR: the use of diagnostic molecular markers in biological control of arthropods. *J. Appl. Entomol.* **131**: 225–240.

- Gerken, T., Kurtz, J., Sauer, K.P., and Lubjuhn, T., (1998). DNA preparation and efficient microsatellite analysis from insect hemolymph. *Electrophoresis* **19**: 3069–3070.
- Gibbons, A., (2010). Tiny time machines revisit ancient life. *Science* **330**: 1616.
- Gilbert, M.T.P., Moore, W., Melchior, L., and Worobey, M., (2007). DNA extraction from dry museum beetles without conferring external morphological damage. *PLoS One* **2** (3): e272.
- Gokool, S.D., Smith, F., and Curtis, C.F., (1992). The use of PCR to help quantify the protection provided by impregnated bednets. *Parasitol. Today* **8**: 347–350.
- Goldenberger, D., Perschil, I., Ritzler, M., and Altweig, M., (1995). A simple “universal” DNA extraction procedure using SDS and proteinase K is compatible with direct PCR amplification. *PCR Methods Appl.* **4**: 368–370.
- Golenberg, E.M., Bickel, A., and Weihs, P., (1996). Effect of highly fragmented DNA on PCR. *Nucleic Acids Res.* **24**: 5026–5033.
- Goodman, M. F. (1995). DNA polymerase fidelity: misinsertions and mismatched extensions, pp. 17–31. In: *PCR Strategies*. Academic Press, San Diego, CA.
- Gorrochotegui-Escalante, N., and Black, W.C., (2003). Amplifying whole insect genomes with multiple displacement amplification. *Insect Mol. Biol.* **12**: 195–200.
- Green, R.E., Krause, J., Briggs, A.W., Maricic, T., Stenzel, U., and Kircher, M., et al. (2010). A draft sequence of the Neandertal genome. *Science* **328**: 710–722.
- Grevelding, C.G., Kampkötter, A., Hollmann, M., Schafer, U., and Kunz, W., (1996). Direct PCR on fruitflies and blood flukes without prior DNA isolation. *Nucleic Acids Res.* **24**: 4100–4101.
- Gutierrez, G., and Marin, A., (1998). The most ancient DNA recovered from an amber-preserved specimen may not be as ancient as it seems. *Mol. Biol. Evol.* **15**: 926–929.
- Guyer, R.L., and Koshland Jr., D.E., (1989). The molecule of the year. *Science* **246**: 1543–1546.
- Hadrys, H., Balick, M., and Schierwater, B., (1992). Applications of random amplified polymorphic DNA (RAPD) in molecular ecology. *Mol. Ecol.* **1**: 55–63.
- Hadrys, H., Schierwater, B., Dellaporta, S.L., DeSalle, R., and Buss, L.W., (1993). Determination of paternity in dragonflies by random amplified polymorphic DNA fingerprinting. *Mol. Ecol.* **2**: 79–87.
- Hallden, C., Hansen, M., Nilsson, N.O., and Hjerdin, A., (1996). Competition as a source of errors in RAPD analysis. *Theor. Appl. Genet.* **93**: 1185–1192.
- Hamiduzzaman, M.M., Guzman-Novoa, E., and Goodwin, P.H., (2010). A multiplex PCR assay to diagnose and quantify *Nosema* infections in honey bees (*Apis mellifera*). *J. Invertebr. Path.* **105**: 151–155.
- Hammond, J.B.W., Spanswick, G., and Mawn, J.A., (1996). Extraction of DNA from preserved animal specimens for use in randomly amplified polymorphic DNA analysis. *Analyt. Biochem.* **240**: 298–300.
- Handt, O., Hoss, M., Krings, M., and Paabo, S., (1994). Ancient DNA: methodological challenges. *Experientia* **50**: 524–529.
- Harris, S., and Jones, D.B., (1997). Optimization of the polymerase chain reaction. *Br. J. Biomed. Sci.* **54**: 166–173.
- He, Q., Viljanen, M.K., and Mertsola, J., (1994a). Effects of thermocyclers and primers on the reproducibility of banding patterns in randomly amplified polymorphic DNA analysis. *Mol. Cell. Probes* **8**: 155–160.
- He, Q., Marjamaki, M., Soini, H., Mertsola, J., and Viljanen, M.K., (1994b). Primers are decisive for sensitivity of PCR. *BioTechniques* **17**: 82–86.
- Hebsgaard, M.B., Phillips, M.J., and Willerslev, E., (2005). Geologically ancient DNA: fact or artefact? *Trends Microbiol.* **13**: 212–220.
- Hedges, S.B., and Schweitzer, M.H., (1995). Detecting dinosaur DNA. *Science* **268**: 1191.
- Her, C., and Weinshilboum, R.M., (1995). Rapid restriction mapping by use of Long PCR. *BioTechniques* **19**: 530–532.

- Higgins, J.A., and Azad, A.F., (1995). Use of polymerase chain reaction to detect bacteria in arthropods: a review. *J. Med. Entomol.* **32**: 213–222.
- Hinnebusch, B.J., Gage, K.L., and Schwan, T.G., (1998). Estimation of vector infectivity rates for plague by means of a standard curve-based competitive polymerase chain reaction method to quantify *Yersinia pestis* in fleas. *Am. J. Trop. Med. Hyg.* **58**: 562–569.
- Hofreiter, M., Serre, D., Poinar, H.N., Kuch, M., and Paabo, S., (2001). Ancient DNA. *Nat. Rev. Genetics* **2**: 353–359.
- Hosono, S., Faruqi, A.F., Dean, F.B., Du, Y., Sun, Z., and Wu, X., et al. (2003). Unbiased whole-genome amplification directly from clinical samples. *Genome Res.* **13**: 954–964.
- Hoss, M., and Paabo, S., (1993). DNA extraction from Pleistocene bones by a silica-based purification method. *Nucleic Acids Res.* **21**: 3913–3914.
- Hoy, M.A., Jeyaprakash, A., Morakote, R., Lo, P.K.C., and Nguyen, R., (2000). Genomic analyses of two populations of *Ageniaspis citricola* (Hymenoptera: Encyrtidae) suggest that a cryptic species may exist. *Biol. Control* **17**: 1–10.
- Hoy, M.A., Jeyaprakash, A., and Nguyen, R., (2001). Long PCR is a sensitive method for detecting *Liberobacter asiaticum* in parasitoids undergoing risk assessment in quarantine. *Biol. Control* **22**: 278–287.
- Iizuka, M., Sugiyama, Y., Iida, S., and Seikiya, T., (1996). Direct sequencing of DNA fragments amplified with single primers by polymerase chain reaction without cloning. *Analyt. Biochem.* **241**: 136–139.
- Innis, M.A., Gelfand, D.H., Ninsky, and J.J., White, T.J., Eds., (1990). *PCR Protocols, A Guide to Methods and Applications*, Academic, San Diego, CA.
- Jackson, D.P., Hayden, J.D., and Quirke, P., (1991). Extraction of nucleic acid from fresh and archival material. In: McPherson, M.J., Quirke, P., and Taylor, G.R., Eds., *PCR. A Practical Approach*, IRL, Oxford, UK, pp. 29–50.
- Jeyaprakash, A., and Hoy, M.A., (2000). Long PCR improves *Wolbachia* DNA amplification: *wsp* sequences found in 76% of sixty-three arthropod species. *Insect Mol. Biol.* **9**: 393–405.
- Jeyaprakash, A., and Hoy, M.A., (2004). Multiple displacement amplification in combination with high-fidelity PCR improves detection of bacteria from single females or eggs of *Metaseiulus occidentalis* (Nesbitt) (Acar: Phytoseiidae). *J. Invertebr. Path.* **86**: 111–116.
- Jeyapraish, A., and Hoy, M.A., (2010). A DNA extraction procedure that allows mite specimens to be slide mounted: phytoseiid species evaluated as a model. *Exp. Appl. Acarol.* **52**: 131–140.
- Jones, C.J., Edwards, K.J., Castaglione, S., Winfield, M.O., Sala, F., and van de Wiel, C., et al. (1997). Reproducibility testing of RAPD, AFLP and SSR markers in plants by a network of European laboratories. *Mol. Breed.* **3**: 381–390.
- Jowett, T., (1986). Preparation of nucleic acids. In: Roberts, D.B., Ed., *Drosophila, A Practical Approach*. IRL, Oxford, UK, pp. 275–286.
- Kainz, P., (2000). The PCR plateau phase—toward an understanding of its limitations. *Biochim. Biophys.* **1494**: 23–27.
- Kalinina, O., Lebedeva, I., Brown, J., and Silver, J., (1997). Nanoliter scale PCR with TaqMan detection. *Nucleic Acids Res.* **25**: 1999–2004.
- Kambhampati, S., and Smith, P.T., (1995). PCR primers for the amplification of four insect mitochondrial gene fragments. *Insect Mol. Biol.* **4**: 233–236.
- Kapsa, R., Siregar, N., Quigley, A., Ojaimi, J., Katsabanis, S., and Sue, C., et al. (1997). The polymerase chain reaction in the study of mitochondrial genetics. *J. Biochem. Biophys. Methods* **36**: 31–50.
- Karl, S.A., and Avise., J.C., (1993). PCR-based assays of Mendelian polymorphisms from anonymous single-copy nuclear DNA: techniques and applications for population genetics. *Mol. Biol. Evol.* **10**: 342–361.
- Kaufman, D.L., and Evans, G.A., (1990). Restriction cleavage at the termini of PCR products. *BioTechniques* **8**: 304–306.

- Kawasaki, E.S. (1990). Amplification of RNA, pp. 21–28. In: *PCR Protocols: A Guide to Methods and Applications*, Academic, New York.
- Kennedy, S., and Oswald, N., Eds., (2011). *PCR Troubleshooting and Optimization: The Essential Guide*, Caister Academic Press, Norfolk, UK.
- Keohavong, P., and Thilly, W.G., (1989). Fidelity of DNA polymerases in DNA amplification. *Proc. Natl. Acad. Sci. USA* **86**: 9253–9257.
- Kernodle, S.P., Cannon, R.E., and Scandalios, J.G., (1993). Concentration of primer and template qualitatively affects products in random-amplified polymorphic DNA PCR. *BioTechniques* **14**: 362–364.
- Khandka, D.K., Tuna, M., Tal, M., Nejidat, A., and Golan-Goldhirsh, A., (1997). Variability in the pattern of random amplified polymorphic DNA. *Electrophoresis* **18**: 2852–2856.
- King, N., Ed., (2010). *RT-PCR Protocols*, 2nd ed. Humana Press, Totowa, NJ.
- Kitchin, P.A.Z., Szotyori, C., Fromholc, and Almond, N., (1990). Avoidance of false positives. *Nature* **344**: 201.
- Knipple, D.C., Payne, L.L., and Soderlund, D.M., (1991). PCR-generated conspecific sodium channel gene probe for the house fly. *Arch. Insect Biochem. Physiol.* **16**: 45–53.
- Kocher, T.D., (1992). PCR, direct sequencing, and the comparative approach. *PCR Methods Appl* **1**: 217–221.
- Kubota, T., Das, S., Christian, S.L., Baylin, S.B., Herman, J.G., and Ledbetter, D.H., (1997). Methylation-specific PCR simplifies imprinting analysis. *Nat. Genet.* **16**: 16–17.
- Kumar, G., Garnova, E., Reagin, M., and Vidali, A., (2008). Improved multiple displacement amplification with Phi29 DNA polymerase for genotyping of single human cells. *BioTechniques* **44**: 879–890.
- Lahr, D.J., and Katz, L.A., (2009). Reducing the impact of PCR-mediated recombination in molecular evolution and environmental studies using a new-generation high-fidelity DNA polymerase. *BioTechniques* **47**: 857–866.
- Lamboy, W.F., (1994). Computing genetic similarity coefficients from RAPD data: the effects of PCR artifacts. *PCR Methods Appl.* **4**: 31–37.
- Lanciotti, R.S., Roehrig, J.T., Deubel, V., Smith, J., Parker, M., and Steele, K., et al. (1999). Origin of the West Nile virus responsible for an outbreak of encephalitis in the northeastern United States. *Science* **286**: 2333–2337.
- Lanciotti, R.S., Kerst, A.J., Nasci, R.S., Dogsey, M.S., Mitchell, C.J., and Savage, H.M., et al. (2000). Rapid detection of West Nile virus from human clinical specimens, field-collected mosquitoes and avian samples by a TaqMan reverse transcriptase-PCR assay. *J. Clin. Microbiol.* **38**: 4066–4071.
- Landweber, L.F., and M. Kreitman (1995). Producing single-stranded DNA in polymerase chain reaction for direct genomic sequencing. pp. 579–588. In: *Recombinant DNA Methodology II*. Academic Press, San Diego, CA.
- Laurent, V., Wajnberg, E., Mangin, B., Schiex, T., Gaspin, C., and Vanlerberghe-Masutti, F., (1998). A composite genetic map of the parasitoid wasp *Trichogramma brassicae* based on RAPD markers. *Genetics* **150**: 275–282.
- Lee, A.B., and Cooper, T.A., (1995). Improved direct PCR screen for bacterial colonies: wooden toothpicks inhibit PCR amplification. *BioTechniques* **18**: 225–226.
- Lexa, M., Horak, J., and Brzobohaty, B., (2001). Virtual PCR. *Bioinformatics* **17**: 192–193.
- Lie, X., and Singh, R.P., (2001). A novel usage of random primers for multiplex RT-PCR detection of virus and viroid in aphids, leaves and tubers. *J. Virol. Methods* **91**: 37–49.
- Lo, Y.M.D., Lun, F.M.F, Chan, K.C.A., Tsui, N.B.Y., Chong, K.C., and Lau, T.K., et al. (2007). Digital PCR for the molecular detection of fetal chromosomal aneuploidy. *Proc. Natl. Acad. Sci. USA* **104**: 1316–13121.
- Lunt, D.H., Hutchinson, W.F., and Carvalho, G.R., (1999). An efficient method for PCR-based isolation of microsatellite arrays (PIMA). *Mol. Ecol.* **8**: 891–894.

- MacPherson, J.M., Eckstein, P.E., Scoles, G.J., and Gajadhar, A.A., (1993). Variability of the random amplified polymorphic DNA assay among thermal cyclers, and effects of primer and DNA concentration. *Mol. Cell. Probes* 7: 293–299.
- Massung, R.F., Slater, K., Owens, J.H., Nicholson, W.L., Mather, T.N., and Solberg, V.B., et al. (1998). Nested PCR assay for detection of granulocytic Ehrlichiae. *J. Clin. Microbiol.* 36: 1090–1095.
- Mael, M.J., Carlton, S.J., and Mather, T.N., (1999). Polymerase chain reaction detection efficiency of the human granulocytic ehrlichiosis agent (Rickettsiaceae: Ehrlichiae) in ticks (Acari: Ixodidae) is dependent on the DNA extraction method. *J. Med. Entomol.* 36: 649–652.
- Meeus, L., deGraff, J.C., Jans, K., and Smagghe, G., (2010). Multiplex PCR detection of slowly-evolving trypanosomatids and neogregarines in bumblebees using broad-range primers. *J. Appl. Microbiol.* 109: 107–115.
- McClelland, M., and Welsh, J., (1994). DNA fingerprinting by arbitrarily primed PCR. *PCR Methods Appl.* 4: 559–565.
- McClelland, M., Mathieu-Daude, F., and Welsh, J., (1995). RNA fingerprinting and differential display using arbitrarily primed PCR. *Trends Genet.* 11: 242–245.
- McEwan, N.R., Bakht, J., Cecchini, E., Coulart, A., Geri, C., and McDonald, F.M., et al. (1998). Examination of the pipetting variation between individuals performing random amplified polymorphic DNA (RAPD) analysis. *J. Microbiol. Methods* 32: 213–215.
- McPherson, M.J., Quirke, P., and Taylor, G.R., Eds., (1991). *PCR. A Practical Approach*, IRL, Oxford, UK.
- McPherson, M.J., Jones, K.M., and Gurr, S.J., (1991). PCR with highly degenerate primers. In: McPherson, M.J., Quirke, P., and Taylor, G.R., Eds., *PCR. A Practical Approach*, IRL, Oxford, UK, pp. 171–186.
- Mertz, L.M., and Rashtchian, A., (1994). Nucleotide imbalance and polymerase chain reaction: effects on DNA amplification and synthesis of high specific activity radiolabeled DNA probes. *Analyt. Biochem.* 221: 160–165.
- Meunier, J.R., and Grimont, P.A.D., (1993). Factors affecting reproducibility of random amplified polymorphic DNA fingerprinting. *Res. Microbiol.* 144: 373–379.
- Micheli, M.R.R., Bova, P., Calissano, and D'Ambrosio, E., (1993). Randomly amplified polymorphic DNA fingerprinting using combinations of oligonucleotide primers. *BioTechniques* 15: 388–390.
- Micheli, M.R., Bova, R., Pascale, E., and D'Ambrosio, E., (1994). Reproducible DNA fingerprinting with the random amplified polymorphic DNA (RAPD) method. *Nucleic Acids Res.* 22: 1921–1922.
- Miller, W., Drautz, D.I., Ratan, A., Pusey, B., Qi, J., and Lesk, A.M., et al. (2008). Sequencing the nuclear genome of the extinct wooly mammoth. *Nature* 456: 387–390.
- Moreira, D., (1998). Efficient removal of PCR inhibitors using agarose-embedded DNA preparations. *Nucleic Acids Res.* 26: 3309–3310.
- Morlais, I., Grebaut, P., Bodo, J.M., Djoha, S., and Cuny, G., (1998). Characterization of trypanosome infections by polymerase chain reaction (PCR) amplification in wild tsetse flies in Cameroon. *Parasitology* 116: 547–554.
- Mueller, U.G., and Wolfenbarger, L.L., (1999). AFLP genotyping and fingerprinting. *Trends Ecol. Evol.* 14: 389–394.
- Mullis, K. B. (1987). Process for amplifying nucleic acid sequences. U.S. Patent 4,683,202. Filed 10-25-85; Issued 7-28-87.
- Mullis, K.B., and Falloona, F., (1987). Specific synthesis of DNA *in vitro* via a polymerase-catalyzed chain reaction. *Methods Enzymol.* 155: 335.
- Mullis, K., Falloona, F., Scharf, S., Saiki, R., Horn, G., and Erlich, H., (1986). Specific enzymatic amplification of DNA *in vitro*: The polymerase chain reaction. *Cold Spring Harbor Symp. Quant. Biol* 51: 263.

- Nelson, W.S., Prodohl, P.A., and Avise, J.C., (1996). Development and application of long-PCR for the assay of full-length animal mitochondrial DNA. *Mol. Ecol.* **5**: 807–810.
- Nevett, C.L., and Louwrier, A., (2000). A formulation of multiple reverse transcriptases for improved RT-PCR. *Biotechnol. Lett.* **22**: 1774–1778.
- Ochman, H., Medhora, M.M., Garza, D., and Hartl, D.L., (1990). Amplification of flanking sequences by inverse PCR, pp. 219–227. In: *PCR Protocols: A Guide to Methods and Applications*. Academic, New York.
- Olsen, D.B., Wunderlich, G., Uy, A., and Eckstein, F., (1993). Direct sequencing of polymerase chain reaction products. In: Wu, R., Ed., *Methods in Enzymology. Volume 218. Recombinant DNA. Part I*. Academic, San Diego, CA, pp. 79–92.
- Osborne, B.I., (1992). HyperPCR: a Macintosh Hypercard program for the determination of optimal PCR annealing temperature. *CABIOS* **8**: 83. Oxford University Press.
- Paabo, S. (1990). Amplifying ancient DNA, pp. 159–166. In: *PCR Protocols: A Guide to Methods and Applications*. Academic, New York.
- Paabo, S., (1991). Amplifying DNA from archeological remains: a meeting report. *PCR Methods Appl.* **1**: 107–110.
- Panelli, S., Damiani, G., Espen, L., Micheli, G., and Sgaramella, V., (2006). Towards the analysis of the genomes of single cells: further characterisation of the multiple displacement amplification. *Gene* **372**: 1–7.
- Pao, C.C., Hor, J.J., Tsai, P.L., and Horng, M.Y., (1993). Inhibition of *in vitro* enzymatic DNA amplification reaction by ultra-violet light irradiation. *Mol. Cell. Probes* **7**: 217–239.
- Paquin, P., and Vink, C.J., (2009). Testing compatibility between molecular and morphological techniques for arthropod systematics: a minimally destructive DNA extraction method that preserves morphological integrity, and the effect of lactic acid on DNA quality. *J. Insect Conserv.* **13**: 453–457.
- Park, D.J., Ed., 2010. *PCR Protocols*, 3rd ed. Humana Press, Totowa, NJ.
- Pekin, D., Skhiri, Y., Baret, J.C., Le Corre, D., Mazutis, L., and Salem, C.B., et al. (2011). Quantitative and sensitive detection of rare mutations using droplet-based microfluidics. *Lab Chip* **11**: 2156–2166.
- Perez, T., Albornoz, J., and Dominguez, A., (1998). An evaluation of RAPD fragment reproducibility and nature. *Mol. Ecol.* **7**: 1347–1357.
- Perler, F.B., Kumar, S., and Kong, H., (1996). Thermostable DNA polymerases. *Adv. Protein Chem.* **48**: 377–435.
- Persing, D.H., Telford III, S.R., Rys, P.N., Dodge, D.E., White, T.J., and Malawista, S.E., et al. (1990). Detection of *Borrelia burgdorferi* DNA in museum specimens of *Ixodes dammini* ticks. *Science* **249**: 1420–1423.
- Peter, T.F., Barbet, A.F., Alleman, A.R., Simbi, B.H., Burridge, M.J., and Mahan, S.M., (2000). Detection of the agent of heartwater, *Cowdria ruminantium*, in *Amblyomma* ticks by PCR: validation and application of the assay to field ticks. *J. Clin. Microbiol.* **38**: 1535–1544.
- Plantard, O., Bouju-Albert, A., Malard, M.A., Hermouet, A., Capron, G., and Verheyden, H., (2012). Detection of *Wolbachia* in the tick *Ixodes ricinus* is due to the presence of the Hymenoptera endoparasitoid *Ixodiphagus hookeri*. *PLoS ONE* **7**: e30692.
- Poinar, H.N., and Stankiewicz, B.A., (1999). Protein preservation and DNA retrieval from ancient tissues. *Proc. Natl. Acad. Sci. USA* **96**: 8426–8431.
- Priest, F.G., (1995). Age of bacteria from amber. *Science* **270**: 2015.
- Quicke, D.L.J.R., Belshaw, and Lopez-Vaamonde, C., (1999). Preservation of hymenopteran specimens for subsequent molecular and morphological study. *Norsk Acad. Sci. Lett. Zool. Scripta* **28**: 261–267.
- Rabinow, P., (1996). *Making PCR. A Story of Biotechnology*. University of Chicago Press, Chicago, IL.

- Rabinow, L., Barrera, A., and Hyde, J., (1993). A simple method for isolation of very high molecular weight DNA from *Drosophila* embryos. *Nucleic Acids Res.* **21**: 4985–4986.
- Rameckers, J., Hummel, S., and Herrmann, B., (1997). How many cycles does a PCR need? Determinations of cycle numbers depending on the number of targets and the reaction efficiency factor. *Naturwissenschaften* **84**: 259–262.
- Rao, V.B., (1994a). Direct sequencing of polymerase chain reaction-amplified DNA. *Analyt. Biochem.* **216**: 1–14.
- Rao, V.B., (1994b). Strategies for direct sequencing of PCR-amplified DNA. *PCR Methods Appl.* **4**: 515–523.
- Ravikumar, G., Urs, S.R., Prakash, N.B.B., Rao, C.G.P., and Vardhana, K.V., (2011). Development of a multiplex polymerase chain reaction for the simultaneous detection of microsporidians, nucleopolyhedrovirus and densovirus infecting silkworms. *J. Invertebr. Pathol.* **107**: 193–197.
- Reineke, A., Kaarlovsky, P., and Zebitz, C.P.W., (1998). Preparation and purification of DNA from insects for AFLP analysis. *Insect Mol. Biol.* **7**: 95–99.
- Roehrdanz, R.L., (1995). Amplification of complete insect mitochondrial genome in two easy pieces. *Insect Mol. Biol.* **4**: 169–172.
- Rohland, N., Siedel, H., and Hofreiter, M., (2004). Nondestructive DNA extraction method for mitochondrial DNA analyses of museum specimens. *BioTechniques* **36**: 814–821.
- Rollo, F., (1998). Ancient DNA: problems and perspectives for molecular microbial paleoecology. In: Carvalho, G.R., Ed., *Advances in Molecular Ecology*. IOS Press, Amsterdam, The Netherlands, pp. 133–149.
- Roux, K.H., (1995). Optimization and troubleshooting in PCR. *PCR Methods Appl.* **4**: S185–S194.
- Rowley, D.L., Coddington, J.A., Gates, M.W., Norrbom, A.L., Ochoa, R.A., and Vandenberg, N.J., et al. (2007). Vouchering DNA-barcoded specimens; test of a nondestructive extraction protocol for terrestrial arthropods. *Mol. Ecol. Notes* **7**: 915–924.
- Ruano, G., Brash, D.E., and Kidd, K.K., (1991). *PCR: The first few cycles. Amplifications. A Forum for PCR Users*. Perkin Elmer Cetus, Norwalk, CT.
- Rugman-Jones, P.F., Forster, L.D., Guerrieri, E., Luck, R.F., Morse, J.C., and Monti, M.M., et al. (2011). Taxon-specific multiplex-PCR for quick, easy, and accurate identification of encyrtid and aphelinid parasitoid species attacking soft scale insects in California citrus groves. *Biocontrol* **56**: 265–275.
- Rychlik, W.W.J., Spencer, and Rhoads, R.E., (1990). Optimization of the annealing temperature for DNA amplification *in vitro*. *Nucleic Acids Res.* **18**: 6409–6412.
- Saiki, R.K., Scharf, S.J., Faloona, F., Mullis, K.B., Horn, G.T., and Erlich, H.A., et al. (1985). Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* **230**: 1350–1354.
- Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., and Horn, G.T., et al. (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239**: 487–491.
- Sall, T., Lind-Halldén, C., and Halldén, C., (2000). Primer mixtures in RAPD analysis. *Hereditas* **132**: 203–208.
- Sambrook, J., and Russell, D.W., (2001)., 3rd Ed. *Molecular Cloning: A Laboratory Manual*, Vol. 2 Cold Spring Harbor Laboratory Press, Cold Sprig Harbor, NY.
- Savelkoul, P.H.M., Aarts, H.J.M., De Haas, J., Dijkshoorn, L., Duim, B., and Otsen, M., et al. (1999). Amplified-fragment length polymorphism analysis: the state of the art. *J. Clin. Microbiol.* **37**: 3083–3091.
- Scharf, S.J., (1990). Cloning with PCR, pp. 84–91. In: *PCR Protocols: A Guide to Methods and Applications*. Academic, New York.
- Schierwater, B., (1995). Arbitrarily amplified DNA in systematics and phylogenetics. *Electrophoresis* **15**: 1643–1647.

- Schierwater, B., and Ender, A., (1993). Different thermostable DNA polymerases may amplify different RAPD products. *Nucleic Acids Res.* **21**: 4647–4648.
- Schriefer, M.E., Sacci Jr., J.B., Wirtz, R.A., and Azad, A.F., (1991). Detection of polymerase chain reaction-amplified malarial DNA in infected blood and individual mosquitoes. *Exp. Parasitol.* **73**: 311–316.
- Schwartz, I., Varde, S., Nadelman, R.B., Wormser, G.P., and Fish, D., (1997). Inhibition of efficient polymerase chain reaction amplification of *Borrelia burgdorferi* DNA in blood-fed ticks. *Am. J. Trop. Med. Hyg.* **56**: 339–342.
- Schweder, M.E., Shatters Jr., R.G., West, S.H., and Smith, R.L., (1995). Effect of transition interval between melting and annealing temperatures on RAPD analyses. *BioTechniques* **19**: 38–42.
- Shahjahan, R.M., Hughes, K.J., Leopold, R.A., and DeVault, J.D., (1995). Lower incubation temperature increases yield of insect genomic DNA isolated by the CTAB method. *BioTechniques* **19**: 333–334.
- Siebert, P.D., and Lerrick, J.W., (1992). Competitive PCR. *Nature* **359**: 557–558.
- Silver, J., (1991). Inverse polymerase chain reaction. In: McPherson, M.J., Quirke, P., and Taylor, G.R., Eds., *PCR. A Practical Approach*, IRL, Oxford, UK, pp. 137–146.
- Simon, C., Frati, F., Beckenbach, A., Crespi, B., Liu, H., and Flook, P., (1994). Evolution, weighting and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Ann. Entomol. Soc. Am.* **87**: 651–701.
- Singer-Sam, J., Tanguay, R.L., and Riggs, A.D., (1989). Use of Chelex to improve the PCR signal from a small number of cells. *Amplifications* **1989** (3): 11.
- Smyth, R.P., Schlub, T.E., Grimm, A., Venturi, V., Chopra, A., and Mallal, S., et al. (2010). Reducing chimera formation during PCR amplification to ensure accurate genotyping. *Gene* **469**: 45–51.
- Steiner, J.J., Poklemba, C.J., Fjellstrom, R.G., and Elliott, L.F., (1995). A rapid one-tube genomic DNA extraction process for PCR and RAPD analyses. *Nucleic Acids Res.* **23**: 2569–2570.
- Suazo, A., and Hall, H.G., (1999). Modification of the AFLP protocol applied to honey bee (*Apis mellifera* L.) DNA. *BioTechniques* **26**: 704–708.
- Sykes, B., (1997). Lights turning red on amber. *Nature* **386**: 764–765.
- Tadmer, A.D., Ottesen, E.A., Leadbettor, J.R., and Phillips, R., (2011). Probing individual environmental bacteria for viruses using microfluidics. *PCR Science* **333**: 58–62.
- Taylor, G.R., (1991). Polymerase chain reaction: basic principles and automation. In: McPherson, M.J., Quirke, P., and Taylor, G.R., Eds., *PCR. A Practical Approach*, IRL, Oxford, UK, pp. 1–14.
- Thomsen, P.F., Elias, S., Gilbert, M.T.P., Haile, J., Munch, K., and Kuzmina, S., et al. (2009). Non-destructive sampling of ancient insect DNA. *PLoS One* **4** (4): e5048.
- Tingey, S.V., and del Tufo, J.P., (1993). Genetic analysis with random amplified polymorphic DNA markers. *Plant Physiol.* **101**: 349–352.
- Townson, H., Harbach, R.E., and Callan, T.A., (1999). DNA identification of museum specimens of the *Anopheles gambiae* complex: an evaluation of PCR as a tool for resolving the formal taxonomy of sibling species complexes. *Syst. Entomol.* **24**: 95–100.
- Tranah, G.J., Lescault, P.J., Hunter, D.J., and De Vivo, I., (2003). Multiple displacement amplification prior to single nucleotide polymorphism genotyping in epidemiologic studies. *Biotechnol. Lett.* **25**: 1031–1036.
- Vanlerberghe-Masutti, F., (1994). Molecular identification and phylogeny of parasitic wasp species (Hymenoptera: Trichogrammatidae) by mitochondrial DNA RFLP and RAPD markers. *Insect Mol. Biol.* **3**: 229–237.
- Vogler, A.P., and Pearson, D.L., (1996). A molecular phylogeny of the tiger beetles (Cicindelidae): congruence of mitochondrial and nuclear rDNA data sets. *Mol. Phyl. Evol.* **6**: 321–338.
- Vos, P., Hogers, R., Bleeker, M., Reijans, M., van de Lee, T., and Horne, M., et al. (1995). AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res.* **23**: 4407–4414.

- Wada, M., Klein, C., Schell, J., and Reiss, B., (1994). A functional assay for *Taq* DNA polymerase in PCR. *BioTechniques* **16**: 28–30.
- Walden, K.K.O., and Robertson, H.M., (1997). Ancient DNA from amber fossil bees? *Mol. Biol. Evol.* **14**: 1075–1077.
- Waters, L.C., Jacobson, S.C., Kroutchinina, N., Khandurina, J., Foote, R.S., and Ramsey, J.M., (1998). Multiple sample PCR amplification and electrophoretic analysis on a microchip. *Anal. Chem.* **70**: 5172–5176.
- Watts, P.C.D.J., Thompson, K.A., Allen, and Kemp, S.J., (2007). How useful is DNA extracted from the legs of archived insects for microsatellite-based population genetic analyses. *J. Insect Conserv.* **11**: 195–198.
- Weill, M., Berticat, C., Raymond, M., and Chevillon, C., (2000). Quantitative polymerase chain reaction to estimate the number of amplified esterase genes in insecticide-resistant mosquitoes. *Analyt. Biochem.* **285**: 267–270.
- Welsh, J., and McClelland, M., (1990). Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Res.* **18**: 7213–7218.
- Welsh, J., Petersen, C., and McClelland, M., (1990). Polymorphisms generated by arbitrarily primed PCR in the mouse; application to strain identification and genetic mapping. *Nucleic Acids Res.* **19**: 303–306.
- Welsh, J., Chada, K., Dalal, S.S., Cheng, R., Ralph, D., and McClelland, M., (1992). Arbitrarily primed PCR fingerprinting of RNA. *Nucleic Acids Res.* **20**: 4965–4970.
- Whale, A.S., Huggett, J.F., Coneen, S., Speirs, V., Shaw, J., and Ellison, S., et al. (2012). Comparison of microfluidic digital PCR and conventional quantitative PCR for measuring copy number variation. *Nucleic Acids Res.* **40** (11): e82.
- Wheeler, E.K., Hara, C.A., Frank, J., Deotte, J., Hall, S.B., and Benett, W., et al. (2011). Under-three minute PCR: probing the limits of fast amplification. *Analyst* **136**: 3707–3712.
- White, T.J., Arnheim, N., and Erlich, H.A., (1989). The polymerase chain reaction. *Trends Genetics* **5**: 185–189.
- White, R.A., Quake, S.R., and Curr, K., (2012). Digital PCR provides absolute quantitation of viral load for an occult RNA virus. *J. Virol. Methods* **179**: 45–50.
- Williams, J.G.K., Kubelik, A.R., Livak, K.J., Rafalski, J.A., and Tingey, S.V., (1990). DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Res.* **18**: 6531–6535.
- Williams, J.G.K., Hanafey, M.K., Rafalski, J.A., and Tingey, S.V., (1993). Genetic analysis using random amplified polymorphic DNA markers. *Methods Enzymol.* **218**: 704–740.
- Wise, D.L., and Weaver, T.L., (1991). Detection of the Lyme disease bacterium, *Borrelia burgdorferi*, by using the polymerase chain reaction and a nonradioisotopic gene probe. *J. Clin. Microbiol.* **29**: 1523–1526.
- Wright, D.K., and Manos, M.M., (1990). Sample preparation from paraffin-embedded tissues, pp. 153–158. In: *PCR Protocols: A Guide to Methods and Applications*, Academic, New York.
- Wu, K., and Hoy, M.A., (2012). Extended starvation reduced and eliminated *Wolbachia*, but not *Cardinium*, from *Metaseiulus occidentalis* females (Acari: Phytoseiidae): A need to reassess *Wolbachia*'s status in this predatory mite? *J. Invert. Path.* **109**(1): 20–26.
- Wu, K., and Hoy, M.A., (2012). *Cardinium* is associated with reproductive incompatibility in the predatory mite *Metaseiulus occidentalis* (Acari: Phytoseiidae). *J. Invert. Path.* **110**(3): 359–365.
- Youryev, A. Ed., (2007). *PCR Primer Design (Methods in Molecular Biology)*. Humana Press, Totowa, NJ.
- Yousten, A.A., and Rippere, K.E., (1997). DNA similarity analysis of a putative ancient bacterial isolate obtained from amber. *FEMS Microbiol. Lett.* **152**: 345–347.
- Zaspel, J.M., and Hoy, M.A., (2008). Comparison of short-term preservation and assay methods for the molecular detection of *Wolbachia* in the Mediterranean flour moth *Ephestia kuehniella*. *Fla. Entomol.* **91**: 494–497.

# Transposable-Element Vectors and Other Methods to Genetically Modify *Drosophila* and Other Insects

## Chapter Outline

- 9.1 Overview 374
- 9.2 Introduction 375
- 9.3 *P* Elements and Hybrid Dysgenesis 376
- 9.4 *P*-Element Structure Varies 377
- 9.5 Transposition Method of *P* Elements 378
- 9.6 Origin of *P* Elements in *D. melanogaster* 379
- 9.7 *P* Vectors and Germ-Line Transformation 381
  - 9.7.1 Protocols 381
  - 9.7.2 Characterizing Transformants 385
- 9.8 Using *P*-Element Vectors 386
  - 9.8.1 Transposon Tagging 386
  - 9.8.2 Expressing Exogenous Genes 387
  - 9.8.3 Evaluating Position Effects 387
  - 9.8.4 Targeted Gene Transfer 388
- 9.9 Transformation of Other Insects with *P* Vectors 390
- 9.10 Evolution of Resistance to *P* Elements 390
- 9.11 Using *P* to Drive Genes into Populations 392
- 9.12 Relationship of *P* to Other Transposable Elements (TEs) 392
- 9.13 Other TEs Can Transform *D. melanogaster* 393
- 9.14 Improved Transformation Tools for *Drosophila* 393
- 9.15 TE Vectors to Transform Insects other than *Drosophila* 395
  - 9.15.1 *piggyBac* 395
  - 9.15.2 *Hermes* and *Herves* 395
  - 9.15.3 *Minos* 396
  - 9.15.4 *mariner* 396
  - 9.15.5 *hobo* 398
- 9.16 Cross Mobilization of TE Vectors 398

9.17 Conversion of Inactive TE Vectors to Activity	398
9.18 Suppression of Transgene Expression	399
9.19 Other Transformation Methods	399
9.19.1 JcDNV Gene Vectors for Somatic Transformations	401
9.19.2 RNAi for <i>Drosophila</i>	401
9.19.3 Zinc-Finger Nucleases (ZFNs)	403
9.19.4 Transcription Activator-Like Effector Nucleases (TALENs)	404
9.19.5 Meganucleases (or Homing Endonucleases)	405
9.19.6 Cell-Penetrating Peptides	406
9.19.7 Nanotechnology Approaches	406
9.20 Conclusions	407
General References	407
References Cited	407

## 9.1 Overview

*P* transposable elements (TEs) were genetically modified to serve as vectors for inserting exogenous DNA into *Drosophila* in 1982 and this tool revolutionized research on *D. melanogaster*. For the first time, geneticists could insert genes to study insect development, gene structure, function, regulation and position effects. *P* elements are found in certain strains (called P) of *D. melanogaster*, but are lacking in others (M). When P males and M females are crossed, their progeny exhibit a condition called “hybrid dysgenesis” because the *P* elements present in the chromosomes of the F<sub>1</sub> progeny are no longer prevented from transposing (or moving). The resultant insertions of *P* elements into new chromosomal sites result in mutations and sterility (=hybrid dysgenesis). *P* elements appear to have invaded *D. melanogaster* ≈60 years ago from another *Drosophila* species (an example of horizontal transmission). A mite (Acarina) could have been the mechanism by which *P* was transferred into *D. melanogaster*. It was hypothesized that a parasitic mite obtained *P* elements from the eggs of one *Drosophila* species during feeding. Subsequent feeding by this “infected” mite on *D. melanogaster* eggs might have resulted in the mechanical transfer of *P* elements to *D. melanogaster* and their subsequent spread in field populations around the world. Alternatively, rare interspecies mating could have allowed *P* to invade *D. melanogaster*.

When *P*-element vectors containing cloned genes are microinjected into early-stage *Drosophila* embryos (before cellular blastoderm), some of the *P* vectors integrate into the chromosomes in germ-line tissues. If the newly inserted DNA is transmitted to the progeny of the injected embryos, stable transformation has occurred. Multiple lines of putative transformants are produced and evaluated for stability and expression levels of the inserted gene. The location of the insertion may affect expression levels. Transposon tagging, which occurs when

a single *P* inserts into a gene and causes a visible mutation, facilitates the identification and cloning of genes from *Drosophila*. *P*-element vectors repair gaps left in chromatids when *P* elements excise, which offers the possibility of inserting exogenous DNA into targeted, rather than random, sites in the *Drosophila* chromosome.

Many drosophilid species have inactive forms of *P* in their genomes, which are suppressed by several mechanisms to reduce the deleterious effects transposition imposes on individuals and populations. The long-term survival of TEs such as *P* may require that it move horizontally into new species.

*P*-element vectors cannot be used to transform insects other than *Drosophila*, but other TEs, including *Hermes*, *minos*, *hobo*, *piggyBac* and *mariner*, have been engineered as vectors. They can transform a variety of insect species, providing tools with which to insert exogenous genes into both pest and beneficial insects with the goal of improving pest-management programs. Risk issues relating to use of TE vectors include the possibility of horizontal transfer to nontarget species and a lack of stability of the transgene. Also, transformation frequencies are usually <5% and insertions can occur in regions of the genome that result in low levels of expression.

Recently, other techniques have been developed to insert genes into specific sites within the genome or to replace a genome sequence with an inserted one or to inhibit expression of endogenous genes. These include zinc-finger nucleases (ZFNs) and transcription activator-like effectors (TALEs). Homing endonucleases (or meganucleases) also can result in gene mutations/insertions. RNA interference (RNAi) can be used to reduce transcription of specific genes, at least temporarily, and nanotechnology methods are being studied as tools to deliver RNAi or to insert novel genes. It remains to be determined whether these new methods will result in more-stable and effective transformations of insects for pest-management programs.

## 9.2 Introduction

The *P* element first was genetically modified to serve as a vector of exogenous DNA in 1982 (Rubin and Spradling 1982, Spradling and Rubin 1982). A variety of different *P*-element vectors are used routinely to introduce exogenous DNA into *D. melanogaster*. *P*-mediated transformation of *D. melanogaster* has revolutionized how geneticists study gene structure, function, regulation, position effects, dosage compensation, and development. *P*-mediated transformation allows geneticists to decipher the genetic basis of behavior, development, and sex determination in *Drosophila*, as is described in Chapters 10 and 11.

This chapter describes *P* elements and hybrid dysgenesis, and the methods used in introducing *P*-element vectors into the germ line of *D. melanogaster*. This approach to germ-line transformation inspired entomologists to attempt to engineer insects other than *Drosophila* using TE vectors with a broader host range.

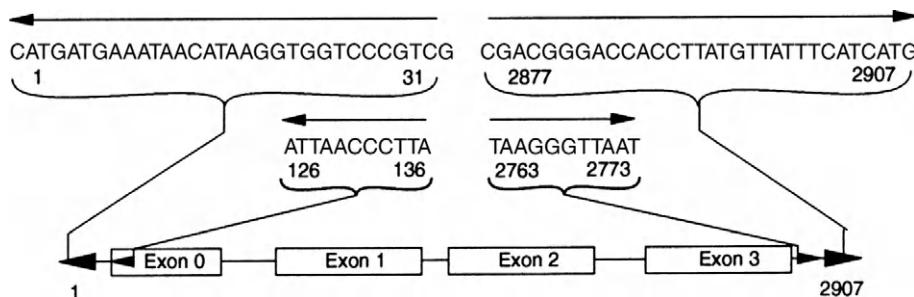
An analysis of the spread of *P* elements into natural populations of *D. melanogaster* previously lacking these TEs allows us to learn more about the evolution of *P* and other TEs. Because TEs have been proposed as possible “drive mechanisms” for the genetic modification of wild insect populations, the study of *P*-element invasion into *D. melanogaster* may serve as a model to understand the potential for using this type of drive mechanism for other insects. Analyses of *P*-element invasions also may provide clues to the evolution of resistance (suppressive factors) to TEs.

Finally, several new tools have been developed that allow genetic modification of *Drosophila* and other insects (see Section 9.19). These methods may result in higher rates of transformation, increased stability and more effective levels of gene expression.

### 9.3 *P* Elements and Hybrid Dysgenesis

Intact *P* elements are 2907 bp long and encode a single polypeptide that has transposase activity (Figure 9.1). There are four exons (numbered from 0 to 3) flanked by inverted repeats that are 31 bp long. The presence of intact inverted repeats is required if the *P* element is to transpose (move).

Multiple copies of *P* (30–60) are dispersed throughout the genome of *P* strains of *D. melanogaster*, but their activity is suppressed by multiple mechanisms. Many *P* elements in *D. melanogaster*, and other *Drosophila* species, have some sequences deleted, which makes them incapable of transposing.



**Figure 9.1** Structure of an intact *P* element. There are four exons (0–3), separated by short introns (thin line). The 31-bp inverted terminal repeats (sequences 1–31 and 2877–2907) are indicated by the filled arrows. There are also inverted repeats at sequences 126–136 and 2763–2773. (Modified from Engels 1989.)

Movements of *P* elements cause mutations by inactivating genes, altering rates of transcription or developmental or tissue-specific gene expression. *P*-element movements break chromosomes and cause nondisjunction during meiosis that can lead to chromosome rearrangements and germ-cell death. Transposition of *P* elements in somatic cells reduces the life span of *D. melanogaster* males, as well as reducing fitness, mating activity, and locomotion (Woodruff et al. 1999).

A syndrome called **hybrid dysgenesis** is induced in *D. melanogaster* when males from a strain that contains *P* (*P* males) are mated with females lacking *P* (*M* females) (Bingham et al. 1982). Their *F*<sub>1</sub> progeny exhibit a high rate of mutation, chromosomal aberrations and, sometimes, complete sterility, caused by transposition of *P* in their germ-line chromosomes. The reciprocal cross does not exhibit these negative effects because the *P* female's cytotype suppresses movement of *P*. *P*-*M* hybrid dysgenesis disrupts piwi-interacting RNAs (piRNAs), allowing the *P* elements to function (Khurana et al. 2011). However, as dysgenic flies age, fertility is restored, *P* elements are silenced and *P*-element piRNAs are produced, which also silence other resident TEs.

When transposition of *P* elements occurs in germ-line tissues in *D. melanogaster*, three short introns have to be cut out of the original RNA transcript to produce the mRNA that codes for a functional 87-kilodalton (kDa) transposase (Laski et al. 1986, Rio et al. 1986, Kobayashi et al. 1993). In somatic cells, a 66-kDa protein is produced that can function as a repressor of *P* activity (Lemaitre et al. 1993).

#### 9.4 *P*-Element Structure Varies

Many *P* elements in the *D. melanogaster* genome are defective. Some have internal deletions and are unable to produce their own transposase but if they retain their 31-bp terminal repeats, they can move if supplied with transposase by intact elements. An early report by Eggleston et al. (1988) suggested that *P* elements are unable to mobilize other TE families in *D. melanogaster*. However, that is not true because *P* elements mobilize the *Tc1* element from the nematode *Caenorhabditis elegans* (Szekely et al. 1994).

Cross mobilization may be common. For example, four different TEs, *Ulysses*, *Penelope*, *Paris* and *Helena*, can be mobilized in dysgenic crosses between strains of *D. virilis* despite the fact that the TEs are structurally diverse (Petrov et al. 1995). *Ulysses* is a retroelement related to the *Ty3-gypsy* superfamily and *Penelope* is similar to another class of retroelements. *Paris* is in the *mariner/Tc1* superfamily that transposes without an RNA intermediate, whereas

*Helena* is a LINE-like element. The transposition mechanisms used by these elements are different and the mutants examined all showed evidence that transposition occurred in the appropriate manner. The four TEs appear to have been mobilized due to “genomic stress” brought about by the dysgenic cross (Petrov et al. 1995). The stress could have been caused by breakage of double-stranded DNA, and can be caused by exposure to UV light and other agents. This can increase transcription and/or mobilization of retroelements. Thus, the production of ds breaks from the mobilization of a single TE might induce a cellular response that releases other TEs from repression. *P* elements with defective 31-bp terminal repeats are unable to transpose because these repeats are the site of action of the transposase. The location of *P* in the chromosome is important in determining the frequency of transposition. Although transposition is more-or-less “random” at the genome scale, *P* elements containing specific gene sequences show some specificity by frequently inserting near the parent gene (which is called “homing”) (Taillebourg and Dura 1999). *P* elements also tend to insert into upstream promoter regions of genes (Spradling et al. 1995).

## 9.5 Transposition Method of *P* Elements

*P* elements move from site to site in the genome (jump) by a “cut-and-paste” method (Engels et al. 1990, Gloor et al. 1991, Sentry and Kaiser 1992, Engels 1997). When a *P* jumps, it leaves behind a ds gap in the DNA. A matching sequence is then used as a template to repair the gap. This matching sequence can occur on the sister chromatid or elsewhere in the genome. If the transposition occurs in an individual that is *heterozygous* for the *P* insertion, and the matching site on the homologous chromosome is used as the template for DNA replication and repair, there can be a precise *loss* of the *P* sequence in the original site, although there is no *net loss* in the genome because the *P* element has simply changed locations. However, if a *P* jumps after the chromosomes have duplicated (when there are two chromatids per chromosome), but before the cell divides, one of the sister chromatids will still have a *P* in its original position. In this situation, the homologous *P* may serve as the template for filling in the gap left when the *P* moved to a new position in the genome. Under these circumstances, the number of *P* elements in the genome is *increased* by one. The *P* element is *replaced* in its original site by gap repair and also is present in a new site in the genome.

The cut-and-paste mechanism of transposition implies that *P* elements don’t have to confer an advantage on the organism to invade and persist in the genome. In fact, a mathematical simulation model indicates that *P* elements can become fixed in populations even when fitness is reduced by 50% (Hickey

1982) and many laboratory studies have shown that colonies can change rapidly from M to P strains. The cut-and-paste model has been the conceptual basis of **targeted gene replacement** (see Section 9.8.4) in *Drosophila* (Engels et al. 1990, Gloor et al. 1991, Sentry and Kaiser 1992, Rong and Golic 2000).

Guerreiro (2011) reviewed factors that make TEs move in the *Drosophila* genome, which is important if transformations obtained by the use of TE vectors are to be stable. Activation of TEs by both biotic and abiotic stresses occurs in many organisms. In *Drosophila*, thermal stresses are associated with movement of TEs, as are dysgenic crosses, hybrid crosses, and colonization events. However, many TE movements occur without any obvious cause. Clearly, there are mechanisms in the genome that subdue the movement of TEs because TE movement rates typically are low. Furthermore, although heat shocks may initiate activity, in other cases it does not; likewise, X-irradiation may activate *P* elements, and gamma irradiation may activate *412* but not *hobo*. Injection of viruses may have an effect or not, depending on the TE tested. The host genetic background and environmental conditions seem to be important in activating TEs in *Drosophila* (and probably in other arthropods). Guerreiro (2011) suggests that TEs increase the evolutionary potential of species, leading to new phenotypes on which selection can act and suggests, "The coevolution of TEs and host genomes may constitute a way to diminish the detrimental effect of transpositions by the silencing mechanisms. In this way, bursts of TE activity followed by calm periods could occur during evolution ... These alternating periods of low and high rate of transposition may be crucial for the generation of genetic variability, and as an effective way to avoid the complete elimination of TEs from host genomes."

## 9.6 Origin of *P* Elements in *D. melanogaster*

*P* elements are relatively new in the genome of *D. melanogaster*. Laboratory strains of *D. melanogaster* collected before 1950 lacked *P*, but most colonies collected from the wild within approximately the past 60 years have *P* elements (Anxolabehere et al. 1988, Engels 1989, 1992, Powell and Gleason 1996).

*P* elements are relatively common in other species of *Drosophila*. Surveys indicated that closely related, full-sized, and potentially active *P* elements occur in *D. willistoni*, *guanche*, *bifasciata*, and *Scaptomyza pallida* (Hagemann et al. 1996). A *P* element isolated from *S. pallida*, a drosophilid distantly related to *D. melanogaster*, is able to transpose in *D. melanogaster* and to mobilize a defective *melanogaster* *P* element (Simonelig and Anxolabehere 1991).

Phylogenetic analyses of DNA sequences from *P* elements in 17 *Drosophila* species in the *melanogaster* species group within the subgenus *Sophophora*

show that sequences from the *P* family fall into distinct subfamilies or clades that are characteristic for particular species subgroups (Clark and Kidwell 1997, Clark et al. 1998). These clades indicate that vertical transmission of *P* elements has occurred, but in some cases the *P* phylogeny is *not* congruent with species phylogeny, indicating horizontal movement has occurred, as well. More than one subfamily of *P* elements may exist within a group, with sequences differing by as much as 36%. In fact, horizontal transfer may be essential to the long-term survival of TEs (Clark and Kidwell 1997).

*P* probably invaded *D. melanogaster* within the past 60 years. The donor species that provided a *P* to *D. melanogaster* is thought to be in the *willistoni* group, which is not closely related to *D. melanogaster* (Daniels et al. 1984, Lansman et al. 1985, Daniels and Strausbaugh 1986, Engels 1997). Because these species diverged from each other  $\approx$ 60 million years ago, there should have been sufficient time for considerable sequence divergence in the *P* elements if they had been present in both genomes before divergence (and subsequently transmitted vertically). However, *P*-element sequences from *melanogaster* and *willistoni* are nearly identical, supporting the hypothesis of horizontal transfer. Engels (1997) speculated that the invasion of *D. melanogaster* by *P* occurred after *D. melanogaster* was introduced into the Americas and that invasion by TEs could be a “general hazard associated with the expansion of any species into a new ecosystem.” Such TE invasions potentially could provide genetic variation that contributes to postrelease adaptations that occurs in some species subsequent to their invasion into new environments.

Two mechanisms were proposed to explain how *P* could have infected *D. melanogaster*. One mechanism involves horizontal transfer and the other mechanism involves interspecific crosses. Both *D. melanogaster* and *D. willistoni* now overlap in their geographical ranges in Florida and in Central and South America, but they apparently are unable to interbreed. Horizontal transfer could have been achieved by a viral, bacterial, fungal, protozoan, spiroplasmal, mycoplasmal, or small-arthropod vector (Hymenoptera or Acari [mites]). One candidate for horizontal vector may be a semiparasitic mite, *Proctolaelaps regalis*, which is associated with both *Drosophila* species (Houck et al. 1991, Kidwell 1992, Engels 1997). This mite has been found in laboratory colonies and in the field associated with fallen or rotting fruit, which is the natural habitat for these species. Laboratory observations indicate that *P. regalis* feeds on fly eggs, larvae, and pupae and can make rapid thrusts of its mouthparts into a series of adjacent hosts. This brief feeding on multiple hosts might allow it to pick up DNA from one egg and inject it into another. Mites obtained from *Drosophila* colonies with *P* elements in their genome were analyzed by the PCR and Southern

blot. These analyses indicated the mites carried both *P* and *Drosophila* ribosomal DNA sequences. Mites isolated from M colonies (which lack *P*) lacked the *P* sequences.

For *P. regalis* to have transferred *P* elements from *D. willistoni* to *D. melanogaster*, several conditions had to occur in the proper sequence (Houck 1993). Females of *D. melanogaster* and *D. willistoni* had to deposit their eggs in proximity and mites had to feed sequentially on one and then the other, in the correct order. The recipient egg had to be <3 hours old, the germ line of the recipient embryo had to incorporate a complete copy of the exogenous *P*, the transformed individual had to survive to adulthood, and the adult had to reproduce.

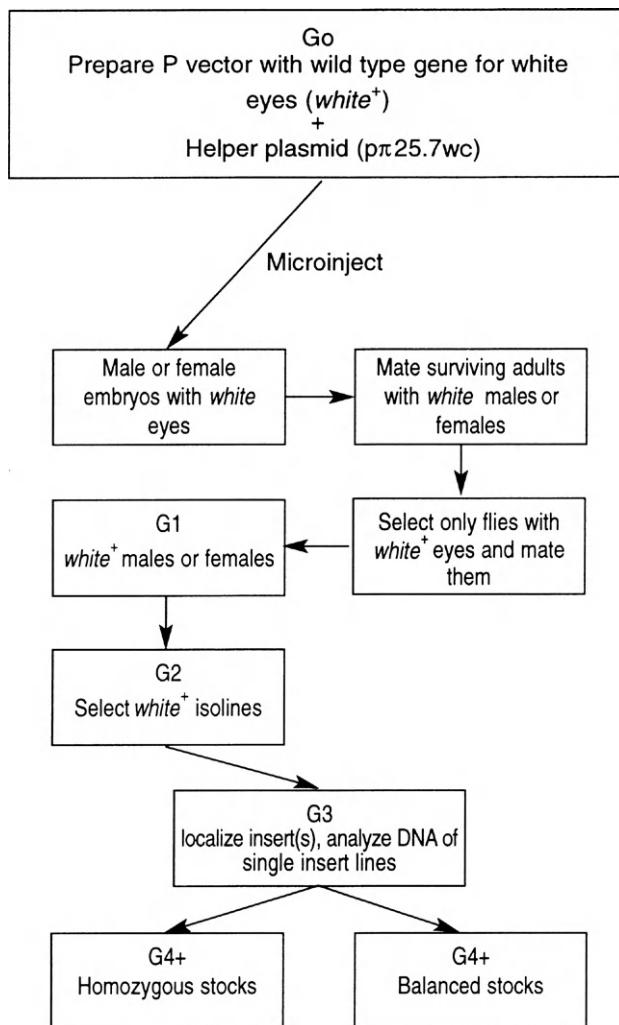
A second potential mechanism for horizontal transfer of *P* involves interspecific crosses. Crosses between the sibling species *Drosophila simulans* and *D. mauritiana* produce sterile males, but fertile females. When F<sub>1</sub> females are backcrossed to males of either species, a few fertile males are produced. To determine whether interspecific transmission of *P* might occur, the two species were crossed and the hybrid progeny were evaluated by *in situ* hybridization of larval salivary glands and Southern blot (Montchamp-Moreau et al. 1991). The results indicated that the *P* element is able to pass from one species to another when the postmating sterility barrier is incomplete. Hybridization, although rare, occurs between some *Drosophila* species.

*P* elements have been found in other Diptera, including Opomyzidae and Trixoscelididae (Anxolabehere and Periquet 1987). Inactive *P* elements were found in the sheep blowfly *Lucilia cuprina* (Calliphoridae) and the house fly *Musca domestica* (Muscidae) (Lee et al. 1999). The *P* elements in *M. domestica* differed from those in *D. melanogaster* by having two introns in exon 2 (as does the *P* from *L. cuprina*). The lack of a functional exon 3 in the house fly *P* likely is the basis for the element's inactivity. *P* elements found in families other than the Drosophilidae suggests that *P* elements are not limited to this lineage (Lee et al. 1999).

## 9.7 *P* Vectors and Germ-Line Transformation

### 9.7.1 Protocols

After *P* elements were cloned (Rubin et al. 1982), they were genetically engineered to serve as vectors to insert exogenous DNA into the germ line of *D. melanogaster* (Spradling and Rubin 1982). Many different vectors with different characteristics have been produced subsequently (Fujioka et al. 2000). The following example outlines the procedures involved in *P*-mediated transformation of *Drosophila* (Figure 9.2):



**Figure 9.2** Steps in transforming *D. melanogaster* with *P*-element vectors. See text for details.

(A) Construct or choose an appropriate *P* vector containing the exogenous DNA and appropriate regulatory elements, as well as a marker gene(s) of interest. In this example, the marker gene is a wild-type version (*white*<sup>+</sup> or *w*<sup>+</sup>) of the mutant gene for *white* (*w/w*) eyes. The wild-type allele (producing red eyes) is dominant over *white*, so if a single copy of *w*<sup>+</sup> is present, the fly will have red eyes. This vector is unable to insert into the chromosome because it can't produce its own transposase. The transposase gene has been eliminated in this vector.

Select a **helper plasmid**, such as pπ25.7wc, that contains a complete DNA sequence coding for transposase. This vector is unable to insert into chromosomes

by the normal transposition method because it lacks 23 bp of one terminal repeat, hence the designation *wc* for “wings clipped.” The inverted repeats are necessary for transposition.

(B) Microinject both the vector and helper plasmids into embryos ( $G_0$ ) from an appropriate host strain with *white* (*ww*) eyes. Embryos should be in the preblastoderm stage, when the cells in the embryo are still in a syncytium.

(C) Mate the  $G_0$  individuals that survive to adulthood after injection with males or females that are homozygous for *white* (*w/w*). If the wild-type gene (*w<sup>+</sup>*) inserted into the chromosome of the injected embryo, then the progeny ( $G_1$ ) will have red eyes but is heterozygous (*w/w<sup>+</sup>*).

(D) Mate  $G_1$  red-eyed progeny (*w/w<sup>+</sup>*) with wild-type (*w<sup>+/w<sup>+</sup></sup>*) flies to produce the next generation ( $G_2$ ) with wild-type eyes. These flies may be transgenic.

(E) Select individual  $G_2$  lines with wild-type (red) eyes.

(F) Identify possible transformants containing single insertions at unique sites (single-insert lines) and verify insert structure.

(G) Analyze the properties of the transformed lines, including the level of expression of the inserted DNA and the stability of the transformed line.

(H) Cross the most useful lines to balancer stocks to enable the lines to be maintained in a stable condition.

Insertion of *P*-element vectors DNA into germ-line chromosomes is enhanced if preblastoderm embryos are microinjected. At that stage, the cleavage nuclei are in a **syncytium** (lacking nuclear membranes), and the *P* elements can more easily be inserted into chromosomes of the **pole cells** that will give rise to the ovaries and testes. Insertion of exogenous DNA into the chromosomes of the germ line results in stable transformation. If only somatic cells obtain the exogenous DNA, the flies cannot transmit the desired trait to their progeny. Such adult  $G_0$  flies may exhibit the trait, but they are only transiently transformed. During embryogenesis and development of the larvae hatching from the injected eggs ( $G_0$  indicates the generation that is injected,  $G_1$  indicates their progeny, and so on), transcription and translation of the wild-type gene (*w<sup>+</sup>*) can produce sufficient xanthine dehydrogenase to produce a red eye color in the adults.  $G_0$  flies with a red eye color do not necessarily have the *white<sup>+</sup>* gene inserted into the germ-line chromosomes because the injected DNA may be transcribed and translated while in the cytoplasm, or the DNA may only be inserted into the chromosomes of somatic cells (=transient transformation). Only a portion of the *P* vector should insert into the chromosome. The DNA

inserted should consist of the sequences contained within the inverted terminal repeats of the *P* element. The plasmid DNA outside the inverted repeats should not insert and should be lost during subsequent development.

The next generation of flies ( $G_1$ ) is the crucial generation to be screened for transformation, because these flies should have red eyes *only* if the *white<sup>+</sup>* gene did insert into the germ-line chromosomes. The presence of one or more progeny with red eyes in the  $G_1$  indicates **stable transformation** occurred. Once transformed fly lines are obtained, the lines should be stable unless transposase is provided in some manner. If an experimenter wants to induce movements of inserted DNA, **secondary transposition** can be induced if transposase is introduced by injecting helper elements into a preblastoderm embryo.

Individual  $G_1$  flies may contain multiple insertions of the *P* element and the *P* element may have inserted into different sites in different  $G_1$  flies. As a result, colonies derived from single flies must be screened to identify colonies with a single insert (Figure 9.2). To determine how many *P* elements inserted into the chromosomes of each colony and their location, DNA from  $G_2$  adult flies is prepared from each isolate and evaluated by Southern-blot analysis. (See Chapter 5 for a description of Southern-blot analysis, which can document that the gene inserted is in the chromosomes of the fly. Analysis by the PCR only indicates that the target DNA is present; it could be present extrachromosomally. It does not prove that the gene inserted into the chromosomes unless the surrounding DNA is amplified by inverse PCR and shown to be *Drosophila* DNA.) In Southern-blot analysis, DNA is cut with restriction endonucleases and probed with labeled *P* sequences to determine the number of insertions. Lines containing multiple insertions should be discarded because these lines will be difficult to analyze.  $G_3$  lines with single inserts are then crossed to *Drosophila* stocks containing appropriate **balancer chromosomes**. Balancer chromosomes prevent crossing over between homologous chromosomes and thus help to maintain stable stocks. The location of the transposon in each single-insert line can be determined by *in situ* hybridization to salivary gland chromosomes.

Transformation rates vary, but may often be <5% of the embryos injected. Furthermore, of those that are transformed, variability in expression of the transgene can be extreme due to position effects. It usually is desirable to obtain at least 10 single-insert lines containing a transgene of interest. This may require microinjecting 600, or more, embryos because survival of embryos after microinjection averages 30 to 70% and, of these, only 50–60% survive to

adulthood ( $G_0$ ). Even after  $G_0$  adults are obtained, damage caused by microinjection may result in early death or sterility in 30–50%.

Transformation does not take place in all germ-line cells in an injected embryo. Usually, only a small fraction of the germ-line cells of a  $G_0$  individual produces transformed  $G_1$  progeny. Thus, it is important to maximize the recovery of  $G_1$  progeny from each  $G_0$  individual injected to increase the probability of detecting progeny in which integration of  $P$  elements occurred. The size of the introduced  $P$  element is another factor that may influence transformation success; the larger the construct, the less frequent the insertion.

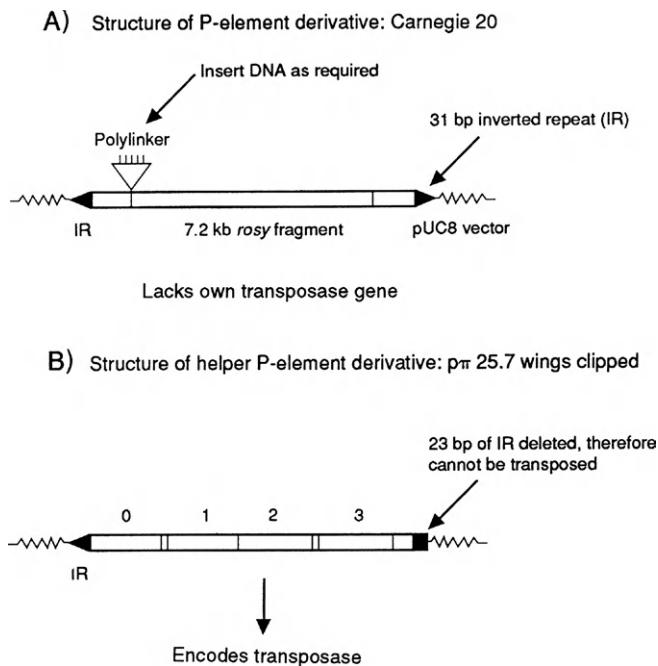
Detailed information on the life history and culture of *Drosophila* are available (Roberts 1986, Ashburner 1989, Matthews 1994, Horn and Wimmer 2000), as are detailed protocols for transforming *Drosophila* with  $P$  vectors (Spradling 1986, Karess 1987). The protocols provide complete information on the appropriate equipment for microinjection, how to stage and dechorionate embryos, align them on slides, desiccate them so that the eggs do not burst upon injection of the plasmid DNA, and inject them in the region that contains the pole cells. Once eggs have been injected, they need to be held under conditions of high relative humidity to prevent death by desiccation. Directions for preparing the DNA for injection and for pulling the required very fine glass needles are provided.

$P$  vectors have been engineered with different characteristics and functions (for examples, see Rubin and Spradling 1983, Karess and Rubin 1984, Cooley et al. 1988, Handler et al. 1993b, Horn and Wimmer 2000, Horn et al. 2000) (Figure 9.3). Generally, the vectors contain restriction sites for cloning, and usually contain one or more selectable marker gene(s).

### 9.7.2 Characterizing Transformants

Identification of transformed flies is achieved in several ways. If a visible marker, such as an eye color gene, is included in the vector then putatively transformed *D. melanogaster* can be determined visually.

Ideally, DNA from putatively transformed lines will be extracted and analyzed by Southern blot to confirm the number of insertions in each line. If large numbers of fly lines need to be characterized, dot-blot analysis can be done. *In situ* hybridization of larval salivary-gland chromosomes will allow a determination to be made of the number of insertions and their location(s). It is desirable to identify lines that carry only a single insertion if the timing and level of expression is



**Figure 9.3** Examples of modified *P*-element vectors. A) The Carnegie 20 vector contains a 7.2-kb segment of DNA coding for the *rosy* gene. It contains a polylinker for inserting exogenous DNA and retains the 31-bp inverted repeats (IR, dark arrows). This vector cannot transpose without a helper element because it cannot make transposase. B) The helper element, *pπ*25.7 wings clipped, produces transposase, but 23 bp of inverted repeat has been deleted at one end so this vector cannot insert into the chromosome.

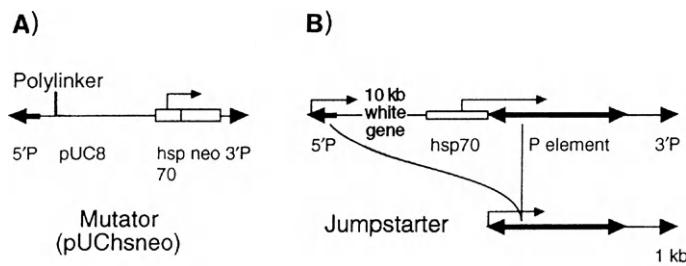
to be determined. Different lines are likely to have different levels of expression because of position effects (Spradling and Rubin 1983, Levis et al. 1985).

## 9.8 Using *P*-Element Vectors

### 9.8.1 Transposon Tagging

The insertion of *P* into a gene allows the isolation and cloning of that gene if the altered gene results in altered phenotype in *D. melanogaster*. Because many *P* strains contain 30–50 copies of *P*, transposon tagging should be carried out in *D. melanogaster* strains lacking endogenous *P* elements.

**Transposon tagging** relies on the development of two specially designed *P* vectors (Cooley et al. 1988). The goal is to introduce a *single P* into the germ line of flies *lacking P*. One vector, called “jumpstarter,” encodes transposase and mobilizes a second vector, called “mutator,” to transpose (Figure 9.4). The



**Figure 9.4** Two *P*-element vectors, mutator (A) and jumpstarter (B), were developed to facilitate insertion of a single *P* element to identify and clone genes in *D. melanogaster*. Jumpstarter encodes transposase and can therefore mobilize mutator. Mutator is able to transpose and carries ampicillin and neomycin resistance genes to facilitate identification and subsequent cloning of the *Drosophila* gene into which it has inserted. (Modified from Cooley et al. 1988.)

structure of the mutator element facilitates identifying and cloning genes because it carries two selectable markers.

### 9.8.2 Expressing Exogenous Genes

Genetic engineering techniques permit the expression of exogenous genes in a variety of organisms, and the availability of a transformation method for *Drosophila* makes it possible to express interesting genes in this insect. For example, Rancourt et al. (1990) obtained expression in *D. melanogaster* of two antifreeze-protein genes isolated from the Atlantic wolffish, *Anarhichas lupus*. The two genes were cloned into a *P* vector with *Drosophila* yolk-protein gene promoters. These highly active promoters were expressed in *Drosophila* females shortly after eclosion and remained active for several weeks. Transformed adult *Drosophila* females produced 1.5–5 mg/ml of antifreeze protein in their hemolymph. The antifreeze activity of the purified protein was determined by measuring freezing point depression and had full biological activity.

### 9.8.3 Evaluating Position Effects

**Transposon jumping** can be used to move stably inserted *P* elements lacking transposase to other sites within the genome. This allows researchers to explore the effects of position on gene expression. To induce jumping, embryos from a transformed strain are injected with helper plasmids that transcribe transposase. The transposase interacts with the terminal repeats of the stably inserted *P* construct, causing it to transpose to a new site. The gene located within the *P* vector experiences a new genomic environment that may modify expression levels. The helper element does not integrate, so the new strain should be stable until transposase is again supplied.

#### 9.8.4 Targeted Gene Transfer

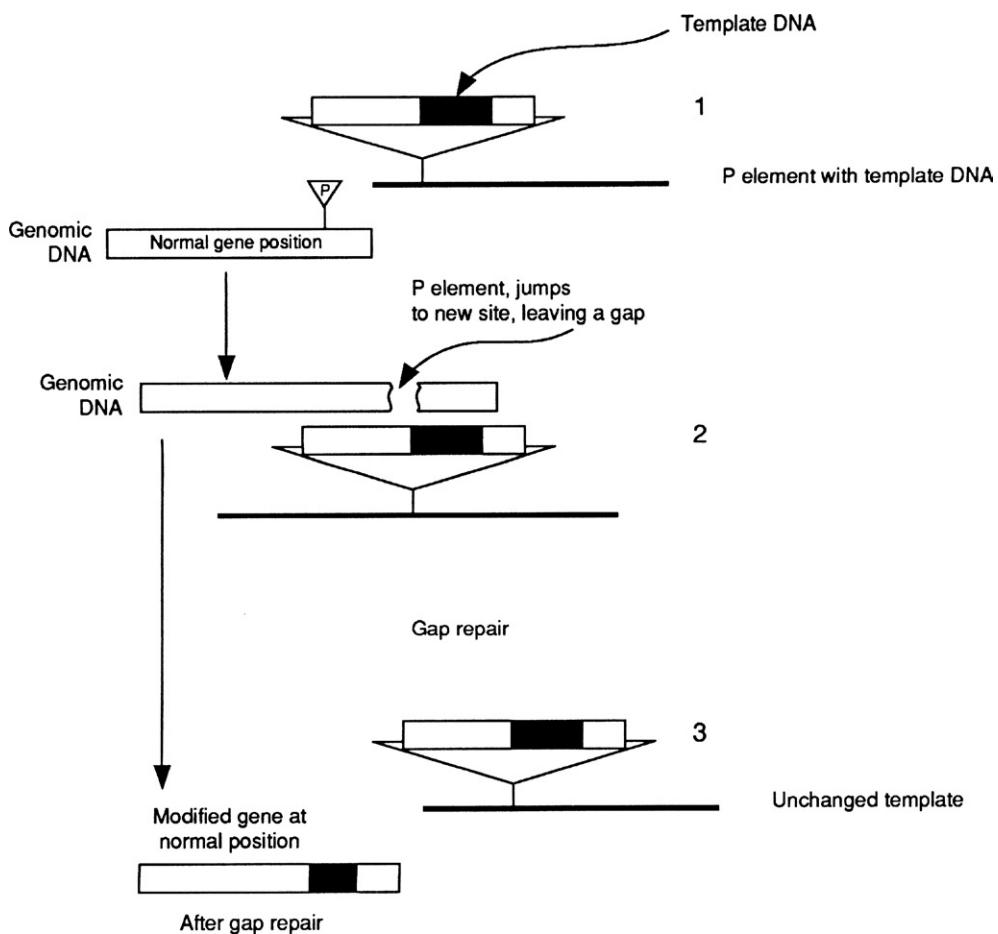
The ability to replace or modify genes in their normal chromosomal locations, **targeted gene transfer**, is a very valuable genetic tool (Ballinger and Benzer 1989, Kaiser and Goodwin 1990, Gloor et al. 1991, Sentry and Kaiser 1992, Lankenau 1995, Siegal and Hartl 1996, Golic et al. 1997, Rong and Golic 2000). Several methods have been evaluated to achieve this goal.

The cut-and-paste mechanism of *P* transposition provided a model for inserting a gene into the gap left behind by a *P* (Figure 9.5). As noted above, *P* transposition leaves a ds gap in the original insertion site, and this gap may be repaired using a template provided by a sister chromatid, or by a homologous chromosome containing a homologous DNA sequence, or by an extrachromosomal element. If the sister chromatid or homologous chromosome has a second copy of the *P*, the *P* sequences will be restored in the gap, giving the impression that transposition has been replicative.

Engels et al. (1990) proposed a method for **site-directed mutagenesis** (=targeted mutagenesis or targeted-gene replacement). The first step is to insert a *P* into the gene of interest, preferably close to the site to be modified. This is feasible because many different colonies of *Drosophila* have been developed that contain *P* elements in known locations. The next step is to transfer the desired *replacement* gene into a second colony with a *P* vector. Then, individuals from the first colony are crossed with the second. A source of transposase is added to promote transposition and targeted gene replacement. In some cases, the replacement gene serves as the template to fill in the gap left when the *P* transposes. The result is that the original site is converted to the desired introduced sequence.

The X-linked *white* locus was modified by targeted-gene replacement (gene conversion) with a success rate of  $\approx 1\%$  (Gloor et al. 1991). Changes ranged from a few base pairs to alterations of at least 2790 bp. A 1% gene conversion rate is sufficiently frequent to make targeted gene transfer a practical method for systematically altering genes in their normal locations to see how their function is modified. An advantage to targeted gene transfer is that it is possible to insert genes longer than 40 kb by this method. *P*-mediated transposition is limited to inserting DNA segments less than 40 kb in length.

Nassif and Engels (1993) investigated the length and stringency of homology required for repair of ds-DNA breaks in *Drosophila* germ cells using the targeted-gene transfer system. They found that a relatively short match (of a few hundred base pairs) of homologous sequence on either side of the target



**Figure 9.5** Targeted gene mutation in *D. melanogaster* is based on the gap repair hypothesis. If a *P* element jumps out of a normal gene, it will leave a gap that must be repaired. Repair is thought to involve using DNA with homologous ends from within the genome as a template for DNA repair. If a new *P* element with a modified gene structure is present, the sequence in the gap can be filled in using the modified gene as the template, leading to a targeted gene alteration. (Modified from Gloor et al. 1991.)

is sufficient to promote gap repair. However, gap repair was sensitive to single-base mismatches within the homologous regions. Interestingly, the data suggested that the ends of a broken chromosome could locate a single homologous template anywhere in the genome using a short stretch of closely matching sequence. How this occurs remains mysterious, but the search is sufficiently efficient that up to several percent of the progeny exhibited targeted-gene replacement at the *white* locus. This high rate of gene conversion is considered to be

unlikely if the process were dependent upon *random* collisions between homologous DNA sequences.

Targeted-gene insertion makes it possible to introduce new genetic information into specific chromosomal sites, or to modify existing genes in directed ways. Random insertions can cause lethality if they insert into essential genes, or can result in poor levels of expression if insertion occurs into heterochromatic regions.

## 9.9 Transformation of Other Insects with *P* Vectors

DNA from *Drosophila melanogaster* has been introduced into *D. simulans* and the more distantly related *D. hawaiiensis* with *P*-element vectors, producing hybrid dysgenesis similar to that found in *D. melanogaster* (Brennan et al. 1984, Daniels et al. 1989). Thus, *P* vectors can integrate and transpose in several *Drosophila* species. Rio et al. (1988) suggested that *P* transposase was active in mammalian cells and yeast, which elicited optimism about the possibility of using *P* elements for genetically engineering other arthropods. Unfortunately, efforts to use *P* vectors to transform arthropod species outside the genus *Drosophila* failed (O'Brochta and Handler 1988, Handler and O'Brochta 1991, Handler et al. 1993a, Handler and James 2000). However, transformation of insects other than *Drosophila* has been accomplished with other vectors (Fraser 2012), as described in Section 9.15.

## 9.10 Evolution of Resistance to *P* Elements

The spread of *P* elements into populations of *D. melanogaster* occurred worldwide since the 1930s. This invasion was remarkable because intact autonomous *P* elements induce severe disadvantages in the newly invaded populations. If *P* elements invade a small population, that population usually is lost (Engels 1997). If evolution of repression systems (resistance to transposition) fails to occur quickly enough, the invaded populations go extinct (Corish et al. 1996).

In fact, several *P* repressor systems (resistance mechanisms) have been identified. The repressors either are transmitted cytoplasmically (maternally inherited) or through the nuclear genome, in which case the transmission is biparental. The repressor systems have been classified as P, M', or Q (Corish et al. 1996, Badge and Brookfield 1998, French et al. 1999).

*P* fly strains have a strong maternally inherited repression system called *P* cytotype (Engels 1989). **P cytotype** is mediated by a 66-kDa protein produced by differential splicing of the complete element's transcript (Laski et al. 1986).

When P females are crossed to a strong P line <10% of the ovaries are dysgenic, indicating that P strains strongly repress hybrid dysgenesis. If P males are crossed to M females (which lack a repression system), >90% of the ovaries are dysgenic in their progeny. P strains are strong inducers of transposition.

M' strains also contain repressor elements. Transposition repression in M' strains is due to the KP element (French et al. 1999). M' strain females display intermediate levels of repression of dysgenesis when crossed to P males. Both males and females from M' strains are able to pass the repressing factor to their progeny.

Q strains strongly repress transposition, allowing a low induction of transposition. Some Q strains show a maternal mode of inheritance of repression while others have a biparental mode of inheritance. It is thought that a repressor (SR) results from a 309-bp deletion at the 3' end of the P element. The SR repressor cannot produce functional transposase but can produce the 66-kDa repressor and a novel 75-kDa protein, both of which may be involved in Q-type repression (French et al. 1999).

Evolution of resistance to P elements can develop rapidly, as demonstrated by two surveys of *D. melanogaster* along a 2900-km cline along the eastern coast of Australia. The first occurred in 1983 and the second in 1993. In 1983, P populations were found in the north, Q populations at central locations, and M' populations in the south (French et al. 1999). After 10 years, Q and M' populations had increased their range at the expense of P lines. French et al. (1999) speculated that the P and M' mechanisms of repression may be early, emergency, responses to the harmful effects of transposition by P. The surviving *D. melanogaster* populations then evolve a superior mechanism by acquiring the biparentally transmitted Q-repression system.

In species of *Drosophila* in which P elements have been present for a long time, no complete functional P has been found (French et al. 1999). Instead, many populations contain tandem repeats of elements with degenerate fourth exons, which might encode some repressor activity. In *D. nebulosa*, a complete element was isolated but the element contained many base changes in all four exons and was nonfunctional. These results reinforce the notion that active transposition of P is detrimental to species of *Drosophila* in the wild.

TEs consist of ≈12% of the genome of *D. melanogaster* and are responsible for ≈80% of the spontaneous mutations found (Guerreiro 2011). However, the activity of TEs is not related to their abundance in the genome. Research to understand the triggers of transposition has focused on biotic and abiotic

stresses and environmental changes. In *D. melanogaster*, a number of experiments suggest temperature and thermal stresses can increase transposition rates, at least some of the time. Dysgenic and hybrid crosses, as well as colonization events, can increase transposition rates of some TEs. X-irradiation can initiate *P*-element transposition, and infections by viruses or other pathogens have initiated transposition in *Drosophila*. The mutagen ethyl methanesulfonate (EMS) also can activate *P* elements. Gamma irradiation appears to induce transposition of the *412* retrotransposon, but not of *hobo*, in *D. melanogaster*. Transposition increases the evolutionary potential of species, possibly leading to new phenotypes and contributing to rapid evolution of species.

### 9.11 Using *P* to Drive Genes into Populations

Using TEs, such as *P*, as drivers for inserting exogenous genes (transgenes) into natural populations for pest control has been analyzed. Computer simulations and empirical studies have used *D. melanogaster* as a model system (Hastings 1994, Carareto et al. 1997). Several suggest that TEs may be used successfully to drive specific genes into pest populations, including populations with different sizes, reproductive rates, density dependence, and transposition frequency. Typically equilibrium was reached within 50 generations, especially if 5 or 10% of the population initially carried the TE. However, there are concerns that horizontal transfer of TEs and the transgenes they carry could affect nontarget species. This topic is discussed further in Chapter 14.

### 9.12 Relationship of *P* to Other Transposable Elements (TEs)

The *Tc1* transposable element from the nematode *Caenorhabditis elegans* is part of a class of TEs that are structurally similar to the *P*, *HB*, and *Uhu* elements of *Drosophila* (Szekely et al. 1994). Both *P* and *Tc1* have perfect inverted repeats and contain transposases. Both excise at a high frequency, and the mechanisms appear similar (Szekely et al. 1994). Thus, it should be no surprise that *D. melanogaster* could be transformed with the *Tc1* element from *C. elegans* (Szekely et al. 1994), which supports the hypothesis that *Tc1*-like elements have a wide distribution within eukaryotes (Avancini et al. 1996). The ability of the *Tc1* element to use *P* transposase indicates that TEs introduced into new species could be mobilized by endogenous TEs. This raises the question as to whether genetically engineered insects transformed with an inactive TE (lacking transposases) would be unstable if endogenous transposases allowed the inserted TE vector and the genes it contains to move. This issue will have to be addressed in risk assessments conducted before the release of transgenic insects; even those transformed with disabled TE vectors, because “conversion” could occur (see Section 9.17).

### 9.13 Other TEs Can Transform *D. melanogaster*

Several types of TEs have been used to transform *D. melanogaster*, including *piggyBac* (Lobo et al. 1999), *hobo* (Ladeveze et al. 1998), *mariner* (Garza et al. 1991), and *Minos* (Loukeris et al. 1995). The *hobo* element occurs naturally in populations of *D. melanogaster*, so it is not surprising that it can serve as a transformation vector. *Minos* was discovered in a related species, *D. hydei*. A TE vector derived from *mariner*, found originally in *D. mauritiana*, also is effective in transforming *D. melanogaster* (Garza et al. 1991). *piggyBac*, isolated from a nuclear polyhedrosis virus infecting a *Trichoplusia ni* cell line (Lepidoptera), is able to transpose in *D. melanogaster*. The finding of *piggyBac* within a virus suggests one mechanism by which TEs could move horizontally between insects. *piggyBac* is related to class II short-inverted-repeat elements, which includes *hobo*, *Minos*, *Hermes*, *mariner*, *P*, *Tc1* (found in nematodes), and *Ac* (found in plants).

### 9.14 Improved Transformation Tools for *Drosophila*

The efficiency of TE-mediated germ-line transformation is dependent both upon the efficiency of the vector and the ability to detect (select) the transformed progeny. Benedict et al. (1994) reported that a parathion hydrolase gene (*opd*), isolated from bacteria, under the control of a *Drosophila hsp70* promoter, conferred resistance to paraoxon in *D. melanogaster* and suggested that this gene could serve as a semidominant selectable marker to detect transformation. Benedict et al. (1994, 1995) also suggested the *opd* gene could be inserted into beneficial arthropods (parasitoids and predators) for improved pest management. The resistant natural enemies could survive the treatment with organophosphate insecticides, whereas the target pests could not. However, due to their long residual activity and high toxicity to mammals, the Food Quality Protection Act has eliminated the use of organophosphates in the United States. The potential horizontal transfer of pesticide-resistance genes to pest species could be detrimental, so this selectable marker should be used only for laboratory studies and should be eliminated before any releases into the environment.

The use of the green fluorescent protein (GFP) as a selectable marker in transformed *D. melanogaster* and other insects has improved detection (Brand 1995, Yeh et al. 1995, Plautz et al. 1996, Tsien 1998, Hazelrigg 2000, Pinkerton et al. 2000). Horn et al. (2000) reported that eye-specific expression of GFP outperforms the mini-white marker in *Drosophila* germ-line transformation experiments. A red fluorescent protein is also useful as a marker for identifying transgenic insects (Horn et al. 2002).

A method of gene targeting involves the insertion of the FLP-FRT system into the chromosomes of *D. melanogaster* via *P*-mediated transformation (Golic et al. 1997). *P*-mediated transformation is essentially a random event and transgenes end up scattered throughout the genome in multiple copies, which results in position-effect variation in expression and gene silencing. The ability to target a transgene to a single site makes comparison of transgenes and their regulatory elements simpler because the various constructs can all be evaluated within a single chromosomal environment.

The FLP-FRT site-specific recombination system is based on a target site (FRT) and the FLP site-specific recombinase enzyme system of yeast (Kilby et al. 1993). This system has two 34-bp target sites for recombination. The recombinase can mediate site-specific recombination resulting in inversions or deletions between the recombination sites. Or, it could result in integration of exogenous DNA into these sites, supplied by injected plasmids containing the gene of interest.

Another approach to gene targeting uses *D. melanogaster*'s endogenous DNA-repair machinery and recombination to substitute one allele for another at a targeted gene or to integrate DNA at a target site as determined by DNA sequence homology (Rong and Golic 2000). This provides *Drosophila* geneticists with the ability to do 'reverse genetics' (Engels 2000). **Reverse genetics** is the induction of a mutation in a gene in order to determine its phenotype.

*P*-element transformation is limited by the size of the DNA that can be inserted, so large genes or gene complexes cannot be inserted. Many *P* elements insert into regulatory elements of genes (possibly disrupting expression), and expression may be affected by their position in the chromosome. To resolve these issues, Venken et al. (2006) constructed a bacterial artificial chromosome (BAC) that can contain DNA inserts as large as 133 kb. BACs insert large DNA fragments into specific sites in the genome, allowing functional analysis of *Drosophila* genes.

Because of the concern that TE vectors could be remobilized by endogenous transposases or by insects mating with insects containing functional transposases, possibly resulting in horizontal transfer of the transgene to other species or loss of the transgene in the modified population, methods were developed by Nimmo et al. (2006), Schetelig et al. (2011), and Tkachuk et al. (2011) to stabilize transgenic lines. Nimmo et al. (2006) modified the genome of *Aedes aegypti*, Schetelig et al. (2009) modified the Mediterranean fruit fly, and Tkachuk et al. (2011) modified *D. melanogaster*. Tkachuk et al. (2011) used homing endonucleases to delete transposon and marker sequences while retaining the transgene and its regulatory elements in the genome. See Section 9.19.5 for a description of homing endonucleases.

## 9.15 TE Vectors to Transform Insects other than *Drosophila*

The genetic modification of pest and beneficial insects by recombinant DNA methods to reduce their pest status or to improve their beneficial effects, respectively, is a goal of many scientists (see Chapter 14, [Ashburner et al. 1998](#), [Fraser 2012](#)). A mechanism with which to reliably insert exogenous DNA into the genome of the target insect and to have it transmitted in a stable manner in the germ line is required to achieve such goals. The *P*-element model dominated the efforts of many scientists initially, but other TE-insertion vectors were developed from *piggyBac*, *mariner*, *Hermes*, *Minos*, and *hobo*. [Palazzoli et al. \(2010\)](#) discuss the scientific and legal issues surrounding the patenting of these transposon tools.

### 9.15.1 *piggyBac*

*piggyBac* is a class II TE found in many insect and other eukaryotic genomes ([Fraser 2000](#)). The *piggyBac* vector can transpose in a *Spodoptera frugiperda* cell line ([Fraser et al. 1995](#), [Elick et al. 1996](#)), as well as in embryos of *D. melanogaster*, *Aedes aegypti*, and *Trichoplusia ni* ([Lobo et al. 1999](#)). The *piggyBac* vector has transformed agricultural pests such as *Ceratitis capitata* ([Handler et al. 1998](#)), and *Pectinophora gossypiella* ([Peloquin et al. 2000](#)). It can transform *Bombyx mori* ([Tamura et al. 2000](#)). Thus, *piggyBac* appears to have a broad host range, has a transformation rate averaging 2–5%, and the insertions appear to be precise (producing a characteristic TTAA duplication at the insertion site) and stable, suggesting that this vector is useful for transforming diverse insects ([Fraser 2000](#), [Handler 2002](#)). It does not, however, function in all insects, perhaps due to repression by endogenous elements ([Fraser 2012](#)).

### 9.15.2 *Hermes* and *Herves*

The *Hermes* TE was discovered in the house fly *Musca domestica* ([O'Brochta et al. 1996](#)). *Hermes* transposed in embryos of *Aedes aegypti* ([Sarkar et al. 1997b](#)). *Hermes* is functional in several dipteran families (Drosophilidae, Calliphoridae, Tephritidae, and Muscidae) ([Atkinson and O'Brochta 2000](#), [O'Brochta et al. 2000](#)). Interestingly, the two strains of *M. domestica* tested exhibited the lowest rates of transformation ([Sarkar et al. 1997a](#)), perhaps because *Hermes* is present in them and some form of resistance to *Hermes* has been selected for. Likewise, the *Lucilia cuprina* strain tested exhibited low levels of transposition, perhaps because a *Hermes*-like element called *hermit* is present in the genome that elicited a partial resistance.

*Aedes aegypti* ([Jasinskiene et al. 1998, 2000](#)) was transformed with *Hermes*, but integrations did not occur precisely at the end of the terminal inverted

repeats and were accompanied by small deletions in the plasmids. These abnormal integrations also did not produce the typical 8-bp duplications at the insertion sites, suggesting that the vector could have integrated into the genome by general recombination or through a partial replicative transposition. As a result, Jasinskiene et al. (2000) concluded this insertion mechanism "precludes its immediate use in experiments that involve field release of transformed animals into the field" although *Hermes* remains useful for laboratory experiments.

*Herves* was found in a population of *Anopheles gambiae* and appears to be an ancient TE that inserted into mosquitoes before the divergence of *Anopheles* species (O'Brochta et al. 2006).

### 9.15.3 *Minos*

*Minos* is related to *mariner* elements and was first found in *D. hydei*. It can transpose in *Drosophila*, *Bombyx mori*, *Tribolium castaneum*, and *Anopheles stephensi* cells and embryos (Catteruccia et al. 2000a, Klinakis et al. 2000a, Shimizu et al. 2000, Pavlopoulos et al. 2004) and also produce stable germ-line transformation (Catteruccia et al. 2000b). Surprisingly, *Minos* can transform human cell lines, making it a useful tool for mutagenesis and functional analysis (Klinakis et al. 2000b). The ability of *Minos* to transform human cells suggests that it potentially has a very wide host range, which could elicit concerns about risks if *Minos* is used to transform insects destined for release into the field in pest-management programs. To reduce potential risk with insects destined for field release, it would be necessary to eliminate the *Minos* element after transformation of the insect.

### 9.15.4 *mariner*

This TE (called *Mos1*) initially was isolated from *Drosophila mauritiana*, but it is extremely widespread among arthropods (Robertson 1995). The phylogeny of *mariner* elements from diverse organisms is not congruent with their host, indicating that *mariner* has moved horizontally across phyla (Robertson 1995, Robertson and Lampe 1995). For example, *mariner* has been found in the insect-parasitic nematode *Heterorhabditis bacteriophora* (Grenier et al. 1999), a root-knot soil nematode (Leroy et al. 2000), three flatworms (*Dugesia tigrina*, *Stylochus zebra*, and *Bdelloura candida*) (Garcia-Fernandez et al. 1995), and two hydras (*Hydra littoralis* and *H. vulgaris*) (Robertson 1997). The host range of *mariner* extends to mammals, including human, mouse, rat, Chinese hamster, sheep, and cow (Auge-Gouillou et al. 1995, Oosumi et al. 1995, Robertson and Martos 1997).

A *mariner* element can transform vertebrate cells, but the transposition rate can be low and variable (Delauriere et al. (2009)). Vertebrates transformed

include the chicken (Sherman et al. 1998) and the zebra fish (Fadool et al. 1998). It even transformed the flagellate protozoan *Leishmania major*, which indicates that *mariner* has a general ability to “parasitize the eukaryotic genome” (Gueiros-Filho and Beverley 1997), perhaps because host proteins are not required for successful integration.

At least two different subfamilies of *mariner* were isolated from the human genome, suggesting multiple horizontal transfers occurred, and Oosumi et al. (1995) suggested that *mariner* could be used as a transformation vector for humans. A *mariner* vector from fish was genetically engineered to make it more active in humans; through genetic recombination and site-directed mutagenesis of a *mariner*-like defective element, a new element called *Sleeping Beauty* was constructed that had 25-fold higher levels of activity in human cells than the “standard” *mariner* (Plasterk et al. 1999). *Sleeping Beauty* is active in tissue-culture cells, as well as the germ line of the mouse and zebrafish (Ivics et al. 2004).

So far, all *mariner* elements discovered in humans are “molecular fossils derived from a *mariner* that was long ago active in the genome of a human ancestor,” with each copy having multiple mutations (Robertson and Martos 1997). Robertson and Zumpano (1997) found that *mariner* sequences are present in all major primate lineages, and estimated that there are ≈200 copies of one (*Hsmar1*) in the human genome, as well as ≈2400 copies of a derived 80-bp inverted repeat structure and ≈46,000 copies of single inverted repeats, suggesting that *mariner* had “a considerable mutagenic effect on past primate genomes.” The human genome is estimated to have been invaded by at least 14 families of TEs and is estimated to have >100,000 degenerate copies of TEs (Smit and Riggs 1996). These include elements called *pogo* (originally discovered in *Drosophila*) and *Tigger*, which are related to *Tc1* and *mariner* (Robertson 1996).

Despite the successes in transforming chickens, fish and other organisms, rates of transformation of arthropods with *mariner* vectors have been low (Lampe et al. 2000). Coates et al. (1995) found *mariner* could excise in *D. melanogaster*, *D. mauritiana*, *Lucilia cuprina*, and *Bactrocera tryoni* embryos. Wang et al. (2000) showed that *mariner* could mediate excision and transposition in *Bombyx mori* tissue-culture cells. Later, *mariner* was shown to transform *Aedes aegypti* (Coates et al. 1998). Some mutants of the transposase gene from a *mariner* isolated from *Haematobia irritans* were found to have 4- to 50-fold increases in activity, indicating that *mariner* vectors could be developed that are more active in arthropods (Lampe et al. 1999). Because *mariner* can function in bacteria such as *Escherichia coli*, it is possible to study basic biochemistry of these elements and to improve them as genetic tools (Lampe 2010).

### 9.15.5 *hobo*

The *hobo* vector transposed in a plasmid-based excision assay in several drosophilid species (Handler and Gomez 1995), and in cells of *Trichoplusia ni* and *Helicoverpa zea* (DeVault et al. 1996), as well as in several tephritids, including *Anastrepha suspensa*, *Bactrocera dorsalis*, *Bactrocera cucurbitae*, *Ceratitis capitata*, and *Toxotrypana curvicauda* (Handler and Gomez 1996). Elements related to *hobo* were found in many of the tephritids. *hobo* mediated germ-line transformation of *D. virilis* (Lozovskaya et al. 1996).

Excision of *hobo* from *H. zea* was stimulated by heat shocks that presumably stimulated the production of an endogenous *hobo*-like transposase. The excision rate was 8- to 10-fold higher than that seen for the normal host or other dipteran species (Atkinson et al. 1993) and, in hindsight, could have been predicted because *hobo* had been found previously in *H. zea* (DeVault and Narang 1994). The instability indicates the importance of checking the target insect species' genome to be sure that endogenous elements related to the TE vector are lacking.

## 9.16 Cross Mobilization of TE Vectors

Laboratory assays were conducted to compare the ability of *Minos*, *piggyBac*, *mariner*, and *Hermes* vectors to **cross mobilize** each other (Sundararajan et al. 1999). The *hobo* transposase functioned equally well with *hobo* and *Hermes* substrates. Conversely, the *Hermes* transposase rarely was able to excise the *hobo* elements from plasmids.

The hAT family of elements (which includes TEs from widely divergent host taxa, including plants, fungi, fish, insects, and humans) appears able to function in novel hosts, and to move horizontally relatively easily (Kidwell and Lisch 1997, Kempken and Windhofer 2001). These attributes make them desirable as vectors for inserting transgenes into arthropods but could be considered negative attributes from the point of view of risk assessments if transgenic insects that contain them are being evaluated for release into the environment (Hoy 2000). The ability of different TEs to mobilize endemic (native) TEs (cross mobilization) is not limited to *Hermes* and *hobo* (Sundararajan et al. 1999).

## 9.17 Conversion of Inactive TE Vectors to Activity

The ability of disabled TE vectors to function in transgenic arthropods should be evaluated before transgenic arthropods are released into the environment (Hoy 2000) because an inactivated *P* vector was converted to activity in a process called **conversion** (Peronnet et al. 2000). The defective *P* vector was converted

into an active TE through a three-step process. The defective *P* (unable to move) used a remote template (another *P* that was itself unable to transpose because it lacked 21bp at its 5' end) for part of the template for the new element. The new element had a restored 5' end that allowed it to transpose, which it obtained from a third element. This provides strong evidence that the search for homology occurs during the DNA repair process after a ds break (Peronnet et al. 2000). Conversion to activity could, in some cases, make a transgene unstable within the transgenic arthropod's genome and could pose a potential risk for horizontal gene transfer.

Schetelig et al. (2011) review methods currently available for enhancing transgene stability. Several different transgene stabilization methods were developed for *Drosophila*. The use of site-specific recombination methods make it possible to integrate DNA cassettes into a specific genomic DNA locus and the elimination of the transposon sequences subsequently should increase their stability because they are unable to move by transposition (Handler et al. 2004, Tkachuk et al. 2011).

## 9.18 Suppression of Transgene Expression

Transgenic plants and mammals are known to inactivate multiple copies of genes if they overexpress a gene or exhibit abnormal transcription (Henikoff 1998). The inactivation phenomenon is thought to be due to a cellular defense mechanism that prevents high levels of expression of TEs or of viruses. In fungi and plants, **gene silencing** is associated with methylation of the DNA, or post-transcriptional and transcriptional processes. Transgene silencing has been described in *D. melanogaster* for the *white-alcohol dehydrogenase* transgenes (Pal-Bhadra et al. 1999). Transgene silencing in *Drosophila* also is associated with the production of heterochromatin (Dorer and Henikoff 1994, 1997). Pal-Bhadra et al. (2002) reported that RNAi mechanisms could affect transcriptional and posttranscriptional transgene silencing in *Drosophila*, especially when the transgenes are present in high copy number. Transgenes that are introduced into different locations in the genome may be silenced along with the endogenous genes. Methods to eliminate transgene silencing will be necessary, or this phenomenon could reduce the effectiveness of transgenic insects released for pest-management programs (Hoy 2000).

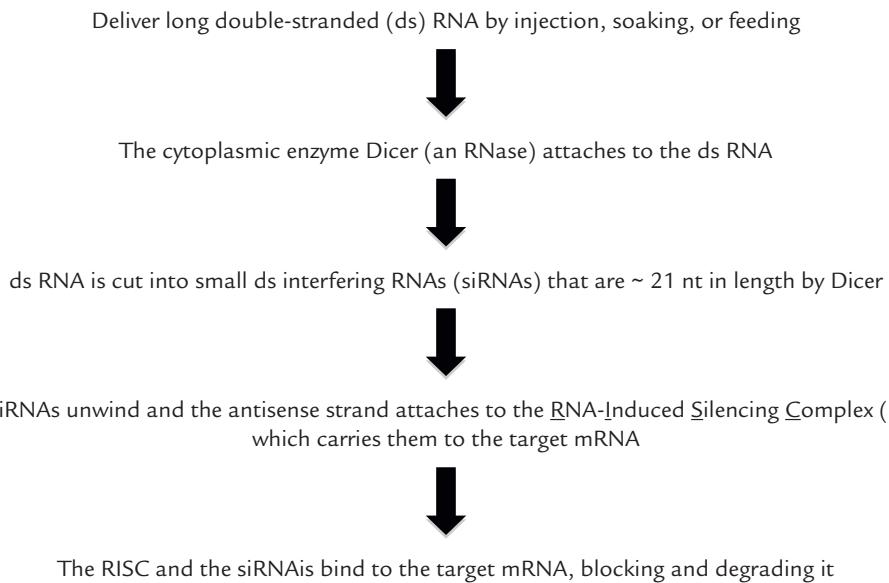
## 9.19 Other Transformation Methods

TE vectors have disadvantages. First, they usually insert randomly, which can result in position effects in the expression of the inserted transgene. Second, they have a relatively low insertion rate (often much <10%), requiring the

injection of hundreds or thousands of embryos to obtain multiple transformant lines for analysis of expression level and fitness. Third, the size of the construct that can be inserted is small (usually 10–15 kb, which is smaller than many genes). It also is unclear whether multiple genes can be inserted that could modify an entire metabolic pathway. Fourth, TE insertions are likely to be unstable for several reasons, which is undesirable if the transformed arthropod is to be released into the field.

In an effort to resolve some of these issues, other transformation methods are being developed, including RNAi (Box 9.1), Zinc-finger nucleases (ZFN), transcription activator-like effectors (TALEs), homing endonucleases (HEs) or meganucleases (Box 9.2), and nanotechnology approaches.

### Box 9.1 Steps Involved in the Basic Mechanism of RNAi and Categories of RNAi



#### Categories

RNAi can be **cell-autonomous** (silencing limited to cell in which the dsRNA is introduced/expressed). The diagram above refers primarily to cell-autonomous RNAi.

**Non-cell-autonomous RNAi** takes place in tissues or cells different from the location or the production of the dsRNA. This is divided into two types: environmental RNAi and systemic RNAi.

Environmental RNAi involves the uptake of dsRNA by a cell from the environment. **Systemic RNAi** takes place in multicellular organisms when a silencing signal is transported from one cell to another or from one tissue type to another.

Derived, in part, from Belles (2010) and Huvenne and Smagghe (2010).

**Box 9.2 Some Newer Tools for Modifying the Genome of *Drosophila* and Other Insects****Homing endonucleases (HEs) or meganucleases**

Homing nucleases, or meganucleases, are sequence-specific endonucleases with large cleavage sites (14–25 bp) that can create double-stranded breaks at specific locations in the genome. Meganucleases are encoded by mobile genetic elements that induce recombination in a process called homing. Several hundred naturally occurring HEs have been identified but the likelihood of finding a HE cleavage site in a particular gene can be low. Artificial meganucleases now can be produced with a specificity determined by the target gene (Smith et al. 2006, Fajardo-Sanchez 2008, Baker 2012, deSouza 2012).

**Transcription activator-like effector nucleases (TALENs)**

TALENs are artificial molecules that have a DNA-binding domain from TAL effector proteins, which are transcription factors from plant-pathogenic bacteria (*Xanthomonas*), and the *Fok I* endonuclease. The DNA-binding domain is a tandem array of amino-acid repeats, each ≈34 residues long. Each TALEN repeat binds to a single base pair. TALENs cleave DNA, have relatively long target sequences and are perhaps less difficult to use than ZFNs (Baker 2012, de Souza 2012).

**Zinc-finger nucleases (ZFNs)**

ZFNs are synthetic molecules containing three or four zinc fingers linked to the *Fok I* nuclease domain that can be used as gene-targeting tools. ZFNs induce double-stranded breaks that result in DNA repair processes that can produce both targeted mutagenesis and targeted gene replacement. ZFNs must be developed for each genomic target. The new DNA (exogenous gene) is inserted by homologous recombination between the original and exogenous gene copies. First, specific DNA sequences are cleaved by the ZFN, donor DNA is added, and homologous recombination will result in gene replacement. Or, nonhomologous end-joining can result in targeted mutagenesis, resulting in deletions, insertions or base substitutions. Because the cleavage domain has no specificity, recognition domains can be inserted that recognize specific domains that should result in targeted cleavages (Carroll 2011, Baker 2012, Isalan 2012). Unfortunately, a substantial proportion of ZFNs fail, so multiple sequences must be tested for a single target gene to identify the most effective. Nontarget cleavages also can occur so extensive tests are conducted to confirm efficacy. ZFNs are commercially available because they are difficult for the average research group to build on their own.

### **9.19.1 JcDNV Gene Vectors for Somatic Transformations**

A lepidopteran densovirus isolated from *Junonia coenia* (JcDNV) and modified to serve as a vector produces somatic transformations that can be used to examine the regulation of transgenes in insects (Bossin et al. 2007). Microinjection of these plasmids into pre-blastoderm embryos of *D. melanogaster*, *Plodia interpunctella*, *Ephestia kuhniella*, and *Trichoplusia ni* allowed transgene expression throughout development and evaluation of different tissue-specific promoters.

### **9.19.2 RNAi for *Drosophila***

RNA interference (RNAi) is an evolutionarily conserved cellular defense mechanism for controlling the expression of alien DNA in plants and animals. RNAi also is involved in controlling development. Fire et al. (1998) discovered RNAi when the application of exogenous ds RNA silenced the homologous mRNA

in the nematode *C. elegans*. The Nobel Prize in Physiology or Medicine was awarded to Andrew Fire and Craig Mello in 2006 for this discovery.

The integration of transposons or viruses results in the production of ds RNA, which activates sequence-specific degradation of homologous single-stranded mRNA or viral genomic RNA. This prevents expression or replication of the foreign genetic material. The ds RNA is used as the guide in this sequence-specific RNA degradation process. Often transgenes inserted into genomes will be silenced together with the endogenous gene from which the transgene was derived ([Pal-Bhadra et al. 2002](#)). RNAi can be used by molecular biologists to learn about gene function for which no phenotype is known ([Mohr et al. 2010](#)), which is useful because  $\approx 30\%$  of putative genes identified in insect genomes are so-called orphan genes, genes without a known homolog. Determining what these orphans do is a high priority if we are to understand their role in the biology of insects. Even for the genes that appear to be homologs to *Drosophila* genes, additional analysis often reveals different functions and pathways.

RNAi is divided into cell-autonomous and non-cell-autonomous RNAi ([Huvenne and Smagghe \(2010\)](#)). In **cell-autonomous RNAi**, the process is limited to the cell in which the ds RNA is present and the whole process occurs within individual cells. The mechanism of cell-autonomous RNAi involves the processing of long ds RNAs into short interfering RNA (siRNA) duplexes of 21–25 nucleotides by a ribonuclease III enzyme, called Dicer. The siRNAs are then incorporated in the **RNA-Induced Silencing Complex (RISC)**. After the siRNA is discarded the RISC binds to a homolog mRNA, cutting it so that translation cannot occur ([Box 9.1](#)).

**Non-cell-autonomous RNAi** can take place in tissues or cells different from the location of the application or production of the dsRNA. Non-cell-autonomous RNAi is further categorized as **environmental RNAi** and **systemic RNAi**. Environmental RNAi involves situations in which ds RNA is taken up by a cell from the environment by a single cell. Systemic RNAi takes place in multicellular organisms and involves the transport of a silencing signal from one cell to another or from one tissue to another. Systemic RNAi can follow environmental RNAi in multicellular organisms.

The production of ds RNA to purposefully interfere with translation of protein-coding genes can be achieved and allows geneticists to determine gene function. Genome-wide RNAi screening is possible using tissue-culture cells or live *Drosophila*. The *Drosophila* RNAi Screening Center at Harvard Medical School provides services and information on the extensive studies conducted using *Drosophila* to study genes relevant to cell viability, metabolism, cell death, transcription, translation, RNA biology, response to pathogen infection, neurobiology, neurodegenerative disease, signal transduction, import/

export/influx/efflux and transport, cell division, morphology, as well as circadian rhythms (Flockhart et al. 2006, Mohr et al. 2010). In addition, the center provides information on reagents, database and software tools for analysis of data, and the off-target effects of RNAi. RNAi using *Drosophila* can be conducted using high-throughput methods, which allows diverse processes to be studied and identification of new components of physiological processes (Mohr et al. 2010). Mohr et al. (2010) concluded RNAi is “a powerful method for genome-scale interrogation of gene function...” but “The results of RNAi high-throughput screens (HTSs) are acutely sensitive to assay design and are subject to significant rates of false discovery, which can be addressed using various statistical, bioinformatics and experimental approaches.” Mohr et al. (2010) also highlighted the need for improved methods, standardization of data reporting, and other methodological improvements.

Unfortunately, feeding ds RNA to *D. melanogaster* does not appear to work, as it does in *C. elegans*. RNAi reagents must be delivered by injection into pre-cellular-blastoderm embryos or as transgenes. RNAi analyses using injection of embryos limits the study of gene function to those genes active during embryonic development. Transgenic RNAi can be used to study gene function in somatic tissues, but the RNAi is cell-autonomous in *D. melanogaster*. So far, RNAi works well in somatic tissues, but does not work in the female germ line, for unknown reasons (Perrimon et al. 2010, Ni et al. 2011). However, Ni et al. (2011) are constructing RNAi tools for all 14,208 protein-coding genes in *Drosophila* so that the *Drosophila* community can conduct RNAi studies on these genes.

RNAi also is being studied in other insects and is considered a possible tool for obtaining control of pests (see Chapter 14).

### 9.19.3 Zinc-Finger Nucleases (ZFNs)

Zinc-finger nucleases (ZFNs) are artificial molecules composed of a Fok I type-II restriction endonuclease and zinc fingers. ZFNs are being studied as a mechanism to genetically modify arthropods without using TE vectors. ZFNs can be used to create targeted double-stranded breaks in the genome that result in targeted mutagenesis or in gene replacement (Kim et al. 1996, Carroll 2011). The enzymes consist of a DNA-binding region, which can be designed to bind to a particular sequence in the genome, and a Fok I endonuclease domain, which introduces the break (Carroll 2011; deSouza 2011, 2012; Baker 2012). The binding domain consists of a tandem array of Cys<sub>2</sub>His<sub>2</sub> zinc fingers, each of which recognizes three nucleotides in the target DNA sequence. By linking together multiple fingers (three to six), ZFN pairs can be designed to bind to genomic sequences 18–36 nt long. Each zinc finger contacts 3 bp of DNA through interactions in the major groove of the DNA and fingers that recognize many of the 64 triplets have been

isolated. In principle, zinc-finger domains can be designed to target a broad range of DNA sequences. Donor DNA can be supplied to allow repair of the break by a homology-driven process, so that a specific gene can be replaced with a desired sequence. Alternatively mutations can occur through nonhomologous end joining.

It is possible to introduce specific mutations in the genomes of vertebrates, plants, and fruit flies with ZFNs. So far, the design of ZFNs is difficult, expensive, and laborious and must be developed for each species (DeFrancesco 2011, Isalan 2012). ZFNs obtained from companies and the Zinc Finger Consortium cost US\$650–\$25,000 depending on source and whether they are custom made. Zinc-finger proteins must fit together in a specific way, so designing and testing ZFNs can be laborious. Information and other resources for ZFNs are available through the Zinc Finger Consortium (<http://zincfingers.org>).

Two genomic loci were modified in the germ line of *D. melanogaster* at a high frequency, which allowed geneticists to discover gene function (Beumer et al. 2006). Beumer et al. (2008) reported high frequencies of gene targeting in *Drosophila* by injecting embryos with ZFNs, resulting in replacement by homologous recombination in up to 10% of the targets at a given locus. More recently, Takasu et al. (2010) produced targeted somatic and germ-line mutations in *Bombyx mori* by injecting ZFNs. Despite these successes, the future use of ZFNs may be supplanted by a different new technology (TALENs, see Section 9.19.4) because TALENs may be easier to design and use. However, TALENs are less well studied, so it is still too early to determine whether TALENs are the better technology (DeFranisco 2011, Baker 2012).

#### 9.19.4 Transcription Activator-Like Effector Nucleases (TALENs)

Transcription Activator-Like Effector Nucleases (TALENs) are a synthetic molecule composed of Transcription Activator-Like Effectors (TALEs) and the *Fok I* endonuclease. They were discovered in 2007 as a tool for modifying genomes (DeFrancesco 2011). TALENs are less expensive and easier to design than ZFNs, and they can be used to upregulate or to edit genes. TALEs, when fused to the *Fok I* nuclease, cleave DNA when present as a dimer. Thus, TALENs function in pairs, binding opposing DNA targets across a spacer over which the *Fok I* domains come together to create a break in the DNA strand.

TAL effectors are a family of virulence factors produced by plant pathogens (*Xanthomonas* species) that can be imported into nuclei and act as transcriptional activators. When TALEs are injected into a host plant, they bind to specific host promoter sequences that regulate genes affecting the disease process. The proteins consist of 17 to 18 repeats of 34 amino acids, arranged side by side. DNA

binding is determined by the amino acids at positions 12 and 13 within each repeat. [Boch et al. \(2009\)](#) and [Moscou and Bogdanove \(2009\)](#) published independent papers in which the TALE code was broken so that specific TALEs can be designed for specific gene sequences ([Deng et al. 2012](#)). This technology is patented and is being developed commercially. Costs for a designer TALEN construct appears to range from US\$5000 to \$12,000. This technology has been used to modify yeast, plants, zebrafish and human cells. Will insects be next?

#### **9.19.5 Meganucleases (or Homing Endonucleases)**

**Meganucleases**, also called **homing endonucleases**, are sequence-specific endonucleases that are found in a variety of single-celled organisms such as Archaea, Eubacteria, yeast, algae, and some plant organelles. They are called “mega” because they recognize sites in the DNA ranging from 12 to 30 bp in length (recall that most endonucleases recognize sequences 4–6 bp in length). To date, ≈600 meganucleases from various unicellular organisms have been identified and sequenced. Meganucleases can identify a single relevant sequence within a genome because they have such large recognition sites. The length of their recognition site makes them highly specific and useful tools for genome modification.

Naturally occurring meganucleases function as parasitic elements that use the double-stranded DNA cell-repair mechanisms to multiply and spread, without damaging the genetic material of its host. A meganuclease binding to its specific DNA recognition site induces a double-stranded break at the unique site in the DNA of a living cell. Once DNA damage is sensed by living cells, DNA-repair mechanisms are initiated and result in homologous recombination. The specificity of meganucleases gives them a high degree of precision and lower cell toxicity than other naturally occurring restriction enzymes. If a homing endonuclease inserts on one homologous chromosome, it will induce a double-stranded break on the normal homolog. Repair of the break, using the chromosome containing the homing endonuclease as a template, results in the homing endonuclease being copied on to the wild-type chromosome.

Despite the large number of naturally occurring meganucleases, the probability of finding a homing site in a chosen gene is low ([Smith et al. 2006](#), [Fajardo-Sánchez 2008](#)). As a result, efforts have been made to design artificial meganucleases with specificities determined by the researcher. Commercial sources have modified the recognition site of natural meganucleases in order to target additional specific genomic DNA sequences. To create tailor-made meganucleases, two approaches were adopted: existing meganucleases were modified by introducing a small number of variations to the amino-acid sequence and then selecting the functional proteins on variations of the natural recognition site.

A second approach involved associating or fusing protein domains from different meganucleases, which results in chimeric meganucleases with a new recognition site composed of a half-site of meganuclease A and a half-site of B. These two approaches can be combined to increase the possibility of creating new enzymes, while maintaining a high degree of efficacy and specificity.

It has been suggested that these selfish elements could be used to genetically manipulate natural populations of pest insects by driving the homing endonuclease gene through the population even when the insert may have deleterious effects on the host (Deredec et al. 2008, Chan et al. 2011). A homing endonuclease was inserted into *D. melanogaster* and high rates of homing were achieved within spermatogonia and in the female germ line (Chan et al. 2011).

#### **9.19.6 Cell-Penetrating Peptides**

Large molecules, such as proteins, DNA or RNA, cannot penetrate cell membranes easily. However, cell-penetrating peptides are a group of short peptides that can cross the cell membrane and could potentially deliver genes into living cells (Chen et al. 2012). Chen et al. (2012) showed that three arginine-rich cell-penetrating peptides could form stable complexes with plasmid DNA and enter insect cells in cell culture. The transferred plasmids containing green fluorescent and red fluorescent protein genes were expressed. This gene-delivery method resulted in a low level of cell mortality (7–16%), suggesting the peptides are not toxic and might be useful for developing transgenic arthropods.

#### **9.19.7 Nanotechnology Approaches**

The definition of nanotechnology is difficult to resolve because it involves many techniques and has potential for use in many applications, including plant biotechnology, electronics, energy, and medicine (Nair et al. 2010). Furthermore, the field is changing rapidly and evolving in its potential techniques and applications. However, nanotechnology can deliver macromolecules in a targeted manner and has been used to deliver DNA into plants and insects (Nair et al. 2010, Zhang et al. 2010). To date, the risk issues and regulatory issues associated with nanotechnology remain to be resolved. Nanoparticles can be transported in the environment and can bioaccumulate in the food chain.

RNAi and nanoparticles were combined to silence chitin synthase genes in *Anopheles gambiae* (Zhang et al. 2010). The double-stranded RNA molecules (RNAi) were included in nanoparticles of chitosan, a nontoxic and biodegradable polymer, and the nanoparticles were fed to larvae. The results suggested that RNAi in *An. gambiae* is systemic (unlike the situation in *D. melanogaster*)

and that the nanoparticle delivery system may improve the stability of ds RNA, which typically has a very short half-life.

## 9.20 Conclusions

Despite the ability to transform a variety of arthropod species with diverse and new methods, considerable work remains to be completed before genetically modified arthropods can be used in practical pest-management programs, as is discussed in Chapter 14.

## General References

- Demerec, M., Ed., (1994). *Biology of Drosophila*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY. This is a facsimile edition of the original first published by John Wiley and Sons in 1950.
- Greenspan, R.J., (1997). *Fly Pushing. The Theory and Practice of Drosophila Genetics*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Hartenstein, V., (1993). *Atlas of Drosophila Development*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Sullivan, W., Ashburner, M., and Hawley, R.S., (2000). *Drosophila Protocols*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

## References Cited

- Anxolabehere, D., and Periquet, G., (1987). *P*-homologous sequences in Diptera are not restricted to the Drosophilidae family. *Genet. Iber.* **39**: 211–222.
- Anxolabehere, D., Kidwell, M.G., and Periquet, G., (1988). Molecular characteristics of diverse populations are consistent with the hypothesis of a recent invasion of *Drosophila melanogaster* by mobile *P* elements. *Mol. Biol. Evol.* **5**: 252–269.
- Ashburner, M., (1989). *Drosophila. A Laboratory Handbook*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Ashburner, M., Hoy, M.A., and Peloquin, J.J., (1998). Prospects for the genetic transformation of arthropods. *Insect Mol. Biol.* **7**: 201–213.
- Atkinson, P.W., Warren, W.D., and O'Brochta, D.A., (1993). The *hobo* transposable element of *Drosophila* can be cross-mobilized in houseflies and excises like the *Ac* element of maize. *Proc. Natl. Acad. Sci. USA* **90**: 9693–9697.
- Atkinson, P.W., and O'Brochta, D.A., (2000). *Hermes* and other *hAT* elements as gene vectors in insects. In: Handler, A.M., and James, A.A., Eds., *Insect Transgenesis*, CRC Press, Boca Raton, FL, pp. 219–235.
- Auge-Gouillou, C., Bigot, Y., Pollet, N., Hamelin, M.H., Meunier-Rotival, M., and Periquet, G., (1995). Human and other mammalian genomes contain transposons of the *mariner* family. *FEBS Lett.* **368**: 541–546.
- Avancini, R.M.P., Walden, K.K.O., and Robertson, H.M., (1996). The genomes of most animals have multiple members of the *Tc1* family of transposable elements. *Genetics* **98**: 131–140.
- Badge, R.M., and Brookfield, J.F.Y., (1998). A novel repressor of *P* element transposition in *Drosophila melanogaster*. *Genet. Res. Camb.* **71**: 21–30.
- Baker, M., (2012). Gene-editing nucleases. *Nature Methods* **9**: 23–26.
- Ballinger, D.G., and Benzer, S., (1989). Targeted gene mutations in *Drosophila*. *Proc. Natl. Acad. Sci. USA* **86**: 9402–9406.

- Belles, X., (2010). Beyond *Drosophila*: RNAi in vivo and functional genomics in insects. *Annu. Rev. Entomol.* **55**: 111–128.
- Benedict, M.Q., Scott, J.A., and Cockburn, A.F., (1994). High-level expression of the bacterial opd gene in *Drosophila melanogaster*: improved inducible insecticide resistance. *Insect Mol. Biol.* **3**: 247–252.
- Benedict, M.Q., Salazar, C.E., and Collins, F.H., (1995). A new dominant selectable marker for genetic transformation; *Hsp70-opd*. *Insect Biochem. Mol. Biol.* **25**: 1061–1065.
- Beumer, K., Bhattacharyya, G., Bibikova, M., Trautman, J.K., and Carroll, D., (2006). Efficient gene targeting in *Drosophila* with zinc-finger nucleases. *Genetics* **172**: 2391–2403.
- Beumer, K.J., Trautman, J.K., Bozas, A., Liu, J.L., Rutter, J., and Gall, J.G., et al. (2008). Efficient gene targeting in *Drosophila* by direct embryo injection with zinc-finger nucleases. *Proc. Natl. Acad. Sci. USA* **105**: 19821–19826.
- Bingham, P.M., Kidwell, M.G., and Rubin, G.M., (1982). The molecular basis of P-M hybrid dysgenesis: the role of the *P* element, a *P*-strain-specific transposon family. *Cell* **29**: 995–1004.
- Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., and Kay, S., et al. (2009). Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* **326**: 1509–1512.
- Bossin, H., Furlong, R.B., Gillett, J.L., Bergoin, M., and Shirk, P.D., (2007). Somatic transformation efficiencies and expression patterns using the *JcDNV* and *piggyBac* transposon gene vectors in insects. *Insect Mol. Biol.* **16**: 37–47.
- Brand, A., (1995). GFP in *Drosophila*. *Trends Genet.* **11**: 324–325.
- Brennan, M.D., Rowan, R.G., and Dickinson, W.J., (1984). Introduction of a functional *P* element into the germ-line of *Drosophila hawaiiensis*. *Cell* **38**: 147–151.
- Carareto, C.M.A., Kim, W., Wojciechowski, M.F., O'Grady, P., Prokchorova, A.V., and Silva, J.C., et al. (1997). Testing transposable elements as genetic drive mechanisms using *Drosophila P* element constructs as a model system. *Genetica* **101**: 13–33.
- Carroll, D., (2011). Genome engineering with zinc-finger nucleases. *Genetics* **188**: 773–782.
- Catteruccia, F., Nolan, T., Blass, C., Muller, H.M., Crisanti, A., and Kafatos, F.C., et al. (2000a). Toward *Anopheles* transformation: *Minos* element activity in anopheline cells and embryos. *Proc. Natl. Acad. Sci. USA* **97**: 2157–2162.
- Catteruccia, F., Nolan, T., Loukeris, T.G., Blass, C., Savakis, C., and Kafatos, F.C., et al. (2000b). Stable germline transformation of the malaria mosquito *Anopheles stephensi*. *Nature* **405**: 959–962.
- Chan, Y.S., Naujoks, D.A., Huen, D.S., and Russell, S., (2011). Insect population control by homing endonuclease-based gene drive: an evaluation in *Drosophila melanogaster*. *Genetics* **188**: 33–44.
- Chen, Y.J., Liu, B.R., Dai, Y.H., Lee, C.Y., Chan, M.H., and Chen, H.H., et al. (2012). A gene delivery system for insect cells mediated by arginine-rich cell-penetrating peptides. *Gene* **492**: 201–210.
- Clark, J.B., and Kidwell, M.G., (1997). A phylogenetic perspective on *P* transposable element evolution in *Drosophila*. *Proc. Natl. Acad. Sci. USA* **94**: 11428–11433.
- Clark, J.B., Kim, P.C., and Kidwell, M.G., (1998). Molecular evolution of *P* transposable elements in the genus *Drosophila*. III. The *melanogaster* species group. *Mol. Biol. Evol.* **15**: 746–755.
- Coates, C.J., Turney, C.L., Frommer, M., O'Brochta, D.A., Warren, W.D., and Atkinson, P.W., (1995). The transposable element *mariner* can excise in non-drosophilid insects. *Mol. Gen. Genet.* **249**: 246–252.
- Coates, C.J., Jasinskiene, N., Miyashiro, L., and James, A.A., (1998). *Mariner* transposition and transformation of the yellow fever mosquito, *Aedes aegypti*. *Proc. Natl. Acad. Sci. USA* **95**: 3748–3751.
- Cooley, L., Kelley, R., and Spradling, A., (1988). Insertional mutagenesis of the *Drosophila* genome with single *P* elements. *Science* **239**: 1121–1128.
- Corish, P., Black, D.M., Featherston, D.W., Merriam, J., and Dover, G.A., (1996). Natural repressors of *P*-induced hybrid dysgenesis in *Drosophila melanogaster*: a model for repressor evolution. *Genet. Res. Camb.* **67**: 109–121.

- Daniels, S.B., and Strausbaugh, L.D., (1986). The distribution of *P* element sequences in *Drosophila*: the *willistoni* and *saltans* species groups. *J. Mol. Evol.* **23**: 138–148.
- Daniels, S.B., Strausbaugh, L.D., Ehrman, L., and Armstrong, R., (1984). Sequences homologous to *P* elements occur in *Drosophila paulistorum*. *Proc. Natl. Acad. Sci. USA* **81**: 6794–6797.
- Daniels, S.B., Chovnick, A., and Kidwell, M.G., (1989). Hybrid dysgenesis in *Drosophila simulans* lines transformed with autonomous *P* elements. *Genetics* **121**: 281–291.
- DeFrancesco, L., (2011). Move over ZFNs. *Nature Biotech.* **29**: 681–684.
- Delauriere, L., Chenais, B., Hardivillier, Y., Gauvry, L., and Casse, N., (2009). *Mariner* transposons as genetic tools in vertebrate cells. *Genetica* **137**: 9–17.
- Deng, D., Yan, C., Pan, X., Mahfouz, M., Wang, J., and Zhu, J.K., et al. (2012). Structural basis for sequence-specific recognition of DNA by TAL effectors. *Science* **335**: 720–723.
- Deredec, A., Burt, A., and Godfray, H.C.J., (2008). The population genetics of using homing endonuclease genes in vector and pest management. *Genetics* **179**: 2013–2026.
- de Souza, N., (2011). Zinc-finger nucleases. *Nat. Methods* **8**: 43.
- de Souza, N., (2012). Primer: genome editing with engineered nucleases. *Nat. Methods* **9**: 27.
- DeVault, J.D., and Narang, S.K., (1994). Transposable elements in Lepidoptera: *hobo*-like transposons in *Heliothis virescens* and *Helicoverpa zea*. *Biochem. Biophys. Res. Commun.* **203**: 169–175.
- DeVault, J.D., Hughes, K.J., Leopold, R.A., Johnson, O.A., and Narang, S.K., (1996). Gene transfer into corn earworm (*Helicoverpa zea*) embryos. *Genome Res.* **6**: 571–579.
- Dorer, D.R., and Henikoff, S., (1994). Expansions of transgene repeats cause heterochromatin formation and gene silencing in *Drosophila*. *Cell* **77**: 993–1002.
- Dorer, D.R., and Henikoff, S., (1997). Transgene repeat arrays interact with distant heterochromatin and cause silencing in *cis* and *trans*. *Genetics* **147**: 1181–1190.
- Eggleston, W.B., Johnson-Schlitz, D.M., and Engels, W.R., (1988). P-M hybrid dysgenesis does not mobilize other transposable element families in *D. melanogaster*. *Nature* **331**: 368–370.
- Elick, T.A., Bauser, C.A., and Fraser, M.J., (1996). Excision of the *piggyBac* transposable element *in vitro* is a precise event that is enhanced by the expression of its encoded transposase. *Genetica* **98**: 33–41.
- Engels, W.R., (1989). *P*-elements in *D. melanogaster*. In: Berg, D.E., and Howe, M.M., Eds., *Mobile DNA*, ASM Press, Washington, DC, pp. 437–484.
- Engels, W.R., (1992). The origin of *P* elements in *Drosophila melanogaster*. *BioEssays* **14**: 681–686.
- Engels, W.R., (1997). Invasions of *P* elements. *Genetics* **145**: 11–15.
- Engels, W.R., (2000). Reversal of fortune for *Drosophila* geneticists. *Science* **288**: 1973–1974.
- Engels, W.R., Johnson-Schlitz, D.M., Eggleston, W.B., and Sved, J., (1990). High-frequency *P* element loss in *Drosophila* is homolog dependent. *Cell* **62**: 515–525.
- Fadool, J.M., Hartl, D.L., and Dowling, J.E., (1998). Transposition of the *mariner* element from *Drosophila mauritiana* in zebrafish. *Proc. Natl. Acad. Sci. USA* **95**: 5182–5186.
- Fajardo-Sanchez, E., Stricher, F., Paques, F., Isalan, M., and Serrano, L., (2008). Computer design of obligate heterodimer meganucleases allows efficient cutting of custom DNA sequences. *Nucleic Acids Res.* **36**: 2163–2173.
- Fire, A., Xu, S.Q., Montgomery, M.K., Kostas, S.A., Driver, S.E., and Mello, C.C., (1998). Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391**: 806–811.
- Flockhart, I., Booker, M., Kiger, A., Boutros, M., Armknecht, S., and Ramadan, N., et al. (2006). FlyRNAi: The *Drosophila* RNAi screening center database. *Nucleic Acids Res.* **34**: D489–494.
- Fraser Jr., M.J., (2000). The TTAA-specific family of transposable elements: identification, functional characterization and utility for transformation of insects. In: Handler, A.M., and James, A.A., Eds., *Insect Transgenesis*, CRC Press, Boca Raton, FL, pp. 249–268.
- Fraser Jr., M.J., (2012). Insect transgenesis: Current applications and future prospects. *Annu. Rev. Entomol.* **57**: 267–289.

- Fraser, M.J., Cary, L., Boonvisudhi, K., and Wang, H.H., (1995). Assay for movement of lepidopteran transposon IFP2 in insect cells using a baculovirus genome as a target DNA. *Virology* **221**: 397–407.
- French, D.J., Corish, P., Shi, M., and Dover, G.A., (1999). Evolutionary flux of *P* element regulation in a *Drosophila melanogaster* hybrid dysgenesis cline. *Genet. Res. Camb.* **73**: 205–216.
- Fujioka, M.J.B., Jaynes, A., Bejsovec, and Weir, M., (2000). Production of transgenic *Drosophila*. In: Tuan, R.S., and Lo, C.W., Eds., *Methods in Molecular Biology*. Vol. 126: *Developmental Biology Protocols*, Humana Press, Totowa, NJ, pp. 353–363. Vol. II.
- Garcia-Fernandez, J., Bayascas-Ramirez, J.R., Marfany, G., Munoz-Marmol, A.M., Casali, A., and Baguna, J., et al. (1995). High copy number of highly similar *mariner*-like transposons in planarian (Platyhelminthes): Evidence for a trans-phyla horizontal transfer. *Mol. Biol. Evol.* **12**: 421–431.
- Garza, D., Medhora, M., Koga, A., and Hartl, D.L., (1991). Introduction of the transposable element *mariner* into the germline of *Drosophila melanogaster*. *Genetics* **128**: 303–310.
- Gloor, G.B., Nassif, N.A., Johnson-Schlitz, D.M., Preston, C.R., and Engels, W.R., (1991). Targeted gene replacement in *Drosophila* via *P* element-induced gap repair. *Science* **253**: 1110–1117.
- Golic, M.M., Rong, Y.S., Petersen, R.B., Lindquist, S.L., and Golic, K.G., (1997). FLP-mediated DNA mobilization to specific target sites in *Drosophila* chromosomes. *Nucleic Acids Res.* **25**: 3665–3671.
- Grenier, E., Abadon, M., Brunet, F., Capy, P., and Abad, P., (1999). A *mariner*-like transposable element in the insect parasite nematode *Heterorhabditis bacteriophora*. *J. Mol. Evol.* **48**: 328–336.
- Gueiros-Filho, F.J., and Beverley, S.M., (1997). Trans-kingdom transposition of the *Drosophila* element *mariner* within the protozoan *Leishmania*. *Science* **276**: 1716–1719.
- Guerreiro, M.P.G., (2011). What makes transposable elements move in the *Drosophila* genome? *Heredity* **108**: 461–468.
- Hagemann, S., Haring, E., and Pinsker, W., (1996). Repeated horizontal transfer of *P* transposons between *Scaptomyza pallida* and *Drosophila bifasciata*. *Genetica* **98**: 43–51.
- Handler, A.M., (2002). Use of the *piggyBac* transposon for germ-line transformation of insects. *Insect Biochem. Mol. Biol.* **32**: 1211–1220.
- Handler, A.M., and Gomez, S.P., (1995). The *hobo* transposable element has transposase-dependent and -independent excision activity in drosophilid species. *Mol. Gen. Genet.* **247**: 399–408.
- Handler, A.M., and Gomez, S.P., (1996). The *hobo* transposable element excises and has related elements in tephritid species. *Genetics* **143**: 1339–1347.
- Handler, A.M., and James, A.A., Eds., (2000). *Insect Transgenesis. Methods and Applications*, CRC Press, Boca Raton, FL.
- Handler, A.M., and O'Brochta, D.A., (1991). Prospects for gene transformation in insects. *Annu. Rev. Entomol.* **36**: 158–183.
- Handler, A.M., Gomez, S.P., and O'Brochta, D.A., (1993a). A functional analysis of the *P*-element gene-transfer vector in insects. *Arch. Insect Bioch. Physiol.* **22**: 373–384.
- Handler, A.M., Gomez, S.P., and O'Brochta, D.A., (1993b). Negative regulation of *P* element excision by the somatic product and terminal sequences of *P* in *Drosophila melanogaster*. *Mol. Gen. Genet.* **237**: 145–151.
- Handler, A.M., McCombs, S.D., Fraser, M.J., and Saul, S.H., (1998). The lepidopteran transposon vector, *piggyBac*, mediates germ-line transformation in the Mediterranean fruit fly. *Proc. Natl. Acad. Sci. USA* **95**: 7520–7523.
- Handler, A.M., Zimowska, G.J., and Horn, C., (2004). Post-integration stabilization of a transposon vector by terminal sequence deletion in *Drosophila melanogaster*. *Nat. Biotechnol.* **22**: 1150–1154.

- Hastings, I.M., (1994). Selfish DNA as a method of pest control. *Proc. Trans. R. Soc. Lond. B* **344**: 313–324.
- Hazelrigg, T., (2000). GFP and other reporters. In: Sullivan, W., Ashburner, M., and Hawley, R.S., Eds., *Drosophila Protocols*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 313–343.
- Henikoff, S., (1998). Conspiracy of silence among repeated transgenes. *Bioessays* **20**: 532–535.
- Hickey, D.A., (1982). Selfish DNA: a sexually transmitted nuclear parasite. *Genetics* **101**: 519–535.
- Horn, C., and Wimmer, E.A., (2000). A versatile vector set for animal transgenesis. *Dev. Genes Evol.* **210**: 630–637.
- Horn, C., Jannich, B., and Wimmer, E.A., (2000). Highly sensitive, fluorescent transformation marker for *Drosophila* transgenesis. *Dev. Genes Evol.* **210**: 623–629.
- Horn, C., Schmid, B.G.M., Pogoda, F.S., and Wimmer, E.A., (2002). Fluorescent transformation markers for insect transgenesis. *Insect Biochem. Mol. Biol.* **32**: 1221–1235.
- Houck, M.A., (1993). Mites as potential horizontal transfer vectors of eukaryotic mobile genes: *Proctolaelaps regalis* as a model. In: Houck, M.A., Ed., *Mites: Ecological and Evolutionary Analyses of Life-History Patterns*. Chapman and Hall, New York, pp. 45–69.
- Houck, M.A., Clark, J.B., Peterson, K.R., and Kidwell, M.G., (1991). Possible horizontal transfer of *Drosophila* genes by the mite *Proctolaelaps regalis*. *Science* **253**: 1125–1129.
- Hoy, M.A., (2000). Deploying transgenic arthropods in pest management programs: Risks and realities. In: Handler, A.M., and James, A.A., Eds., *Insect Transgenesis. Methods and Applications*, CRC Press, Boca Raton, FL, pp. 335–337.
- Huvenne, H., and Smagghe, G., (2010). Mechanisms of dsRNA uptake in insects and potential for RNAi for pest control: a review. *J. Insect Physiol.* **56**: 227–235.
- Isalan, M., (2012). Zinc-finger nucleases: how to play with two good hands. *Nat. Methods* **9**: 32–34.
- Ivics, Z., Kaufman, C.D., Zayed, H., Misley, C., Walisko, O., and Izsvak, Z., (2004). The *Sleeping Beauty* transposable element: evolution, regulation and genetic applications. *Curr. Issues Mol. Biol.* **6**: 43–56.
- Jasinskiene, N., Coates, C.J., Benedict, M.Q., Cornel, A.J., Rafferty, C.S., and James, A.A., et al. (1998). Stable transformation of the yellow fever mosquito, *Aedes aegypti*, with the *Hermes* element from the housefly. *Proc. Natl. Acad. Sci. USA* **95**: 3743–3747.
- Jasinskiene, N., Coats, C.J., and James, A.A., (2000). Structure of *Hermes* integrations in the germline of the yellow fever mosquito, *Aedes aegypti*. *Insect Mol. Biol.* **9**: 11–18.
- Kaiser, K., and Goodwin, S.F., (1990). "Site-selected" transposon mutagenesis of *Drosophila*. *Proc. Natl. Acad. Sci. USA* **87**: 1686–1690.
- Karess, R.E., (1987). *P* element mediated germline transformation of *Drosophila*. In: Glover, D.M., Ed., *DNA Cloning Volume II. A Practical Approach*. IRL Press, Oxford, UK, pp. 121–141.
- Karess, R.E., and Rubin, G.M., (1984). Analysis of *P* transposable element functions in *Drosophila*. *Cell* **38**: 135–146.
- Kempken, F., and Windhofer, F., (2001). The hAT family: a versatile transposon group common to plants, fungi, animals, and man. *Chromosoma* **110**: 1–9.
- Khurana, J.S., Wang, J., Xu, J., Koppetsch, B.S., Thomson, T.C., and Nowosielska, A., et al. (2011). Adaptation to *P* element transposon invasion in *Drosophila melanogaster*. *Cell* **147**: 1551–1563.
- Kidwell, M.G., (1992). Horizontal transfer of *P* elements and other short inverted repeat transposons. *Genetica* **86**: 275–286.
- Kidwell, M.G., and Lisch, D., (1997). Transposable elements as sources of variations in animals and plants. *Proc. Natl. Acad. Sci. USA* **94**: 7704–7711.
- Kilby, N.J., Snaith, M.R., and Murray, J.A.H., (1993). Site-specific recombinases: tools for genome engineering. *Trends Genet.* **9**: 413–421.

- Kim, Y.G., Cha, J., and Chandrasegaran, S., (1996). Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proc. Natl. Acad. Sci. USA* **93**: 1156–1160.
- Klinakis, A.G., Loukeris, T.G., Pavlopoulos, A., and Savakis, C., (2000a). Mobility assays confirm the broad host-range activity of the *Minos* transposable element and validate new transformation tools. *Insect Mol. Biol.* **9**: 269–275.
- Klinakis, A.G., Zagoraiou, L., Vassilatis, D.K., and Savakis, C., (2000b). Genome-wide insertional mutagenesis in human cells by the *Drosophila* mobile element *Minos*. *EMBO Rep.* **1**: 416–421.
- Kobayashi, S., Kitamura, T., Sasaki, H., and Okada, M., (1993). Two types of pole cells are present in the *Drosophila* embryo, one with and one without splicing activity for the third *P*-element intron. *Development* **117**: 885–893.
- Ladeveze, V., Galindo, I., Chaminade, N., Pascual, L., Periquet, G., and Lemeunier, F., (1998). Transmission pattern of *hobo* transposable element in transgenic lines of *Drosophila melanogaster*. *Genet. Res. Camb.* **71**: 97–107.
- Lampe, D.J., (2010). Bacterial genetic methods to explore the biology of mariner transposons. *Genetica* **138**: 499–508.
- Lampe, D.J., Akerley, B.J., Rubin, E.J., Mekalanos, J.J., and Robertson, H.M., (1999). Hyperactive transposase mutants of the *Himar1* mariner transposon. *Proc. Natl. Acad. Sci. USA* **96**: 11428–11433.
- Lampe, D.J., Walden, K.K.O., Sherwood, J.M., and Robertson, H.M., (2000). Genetic engineering of insects with *mariner* transposons. In: Handler, A.M., and James, A.A., Eds., *Insect Transgenesis*, CRC Press, Boca Raton, FL, pp. 237–248.
- Lankenau, D.H., (1995). Genetics of genetics in *Drosophila*: *P* elements serving the study of homologous recombination, gene conversion and targeting. *Chromosoma* **103**: 659–668.
- Lansman, R.A., Stacey, S.M., Grigliatti, T.A., and Brock, H.W., (1985). Sequences homologous to the *P* mobile element of *D. melanogaster* are widely distributed in the subgenus *Sophophora*. *Nature* **318**: 561–563.
- Laski, F.A., Rio, D.C., and Rubin, G.M., (1986). Tissue specificity of *Drosophila P* element transposition is regulated at the level of mRNA splicing. *Cell* **44**: 7–17.
- Lee, S.H., Clark, J.B., and Kidwell, M.G., (1999). A *P* element-homologous sequence in the house fly, *Musca domestica*. *Insect Mol. Biol.* **8**: 491–500.
- Lemaitre, B., Ronsseray, S., and Coen, D., (1993). Maternal repression of the *P* element promoter in the germline of *Drosophila melanogaster*: a model for the *P* cytotype. *Genetics* **135**: 149–160.
- Leroy, H., Leroy, F., Auge-Gouillou, C., Castagnone-Sereno, P., Vanlerberghe-Masutti, F., Bigot, Y., and Abad, P., (2000). Identification of *mariner*-like elements from the root-knot nematode *Meloidogyne* spp. *Mol. Biochem. Parasitol.* **107**: 181–190.
- Levis, R., Hazelrigg, T., and Rubin, G.M., (1985). Effects of genomic position on the expression of transduced copies of the *white* gene of *Drosophila*. *Science* **229**: 558–561.
- Lobo, N., Li, X., and Fraser Jr., M.J., (1999). Transposition of the *piggyBac* element in embryos of *Drosophila melanogaster*, *Aedes aegypti* and *Trichoplusia ni*. *Mol. Gen. Genet.* **261**: 803–810.
- Loukeris, T.G., Arca, B., Livadaras, I., Dialetaki, G., and Savakis, C., (1995). Introduction of the transposable element *Minos* into the germ line of *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* **92**: 9485–9489.
- Lozovskaya, E.R., Nurminsky, D.I., Hartl, D.L., and Sullivan, D.T., (1996). Germline transformation of *Drosophila virilis* mediated by the transposable element *hobo*. *Genetics* **142**: 173–177.
- Matthews, K.A., (1994). Care and feeding of *Drosophila*, *Methods In Cell Biology*, Vol. **44**. Academic, New York, 13–31.
- Mohr, S., Baka, C., and Perrimon, N., (2010). Genomic screening with RNAi: results and challenges. *Annu. Rev. Biochem.* **79**: 37–64.
- Moscou, M.J., and Bogdanove, A.J., (2009). A simple cipher governs DNA recognition by TAL effectors. *Science* **326**: 1501.

- Montchamp-Moreau, C., Periquet, G., and Anxolabehere, D., (1991). Interspecific transfer of *P* elements by crosses between *Drosophila simulans* and *Drosophila mauritiana*. *J. Evol. Biol.* **4**: 131–140.
- Nair, R., Varghese, S.H., Nair, B.G., Maekawa, T., Yoshida, Y., and Kumar, D.S., (2010). Nanoparticulate material delivery to plants. *Plant Sci.* **179**: 154–163.
- Nassif, N., and Engels, W., (1993). DNA homology requirements for mitotic gap repair in *Drosophila*. *Proc. Natl. Acad. Sci. USA* **90**: 1262–1266.
- Ni, J.Q., Zhou, R., Czech, B., Liu, L.P., Holderbaum, L., and Yang-Zhou, D., et al. (2011). A genome-scale shRNA resource for transgenic RNAi in *Drosophila*. *Nature Methods* **8**: 405–409.
- Nimmo, D.D., Alphey, L., Meredith, J.M., and Eggleston, P., (2006). High efficiency site-specific genetic engineering of the mosquito genome. *Insect Mol. Biol.* **15**: 129–136.
- O'Brochta, D.A., and Handler, A.M., (1988). Mobility of *P* elements in drosophilids and nondrosophilids. *Proc. Natl. Acad. Sci. USA* **85**: 6052–6056.
- O'Brochta, D.A., Warren, W.D., Saville, K.J., and Atkinson, P.W., (1996). *Hermes*, a functional non-drosophilid insect gene vector from *Musca domestica*. *Genetics* **142**: 907–914.
- O'Brochta, D.A., Atkinson, P.W., and Lehane, M.J., (2000). Transformation of *Stomoxys calcitrans* with a *Hermes* gene vector. *Insect Mol. Biol.* **9**: 531–538.
- O'Brochta, D.A., Subramanian, R.A., Orsetti, J., Peckham, E., and Nolan, N., et al. (2006). *hAT* element population genetics in *Anopheles gambiae* s.l. in Mozambique. *Genetica* **127**: 185–198.
- Oosumi, T., Belknap, W.R., and Garlick, B., (1995). *Mariner* transposons in humans. *Nature* **378**: 672.
- Pal-Bhadra, M., Bhadra, U., and Birchler, J.A., (1999). Cosuppression of nonhomologous transgenes in *Drosophila* involves mutually related endogenous sequences. *Cell* **99**: 35–46.
- Pal-Bhadra, M., Bhadra, U., and Birchler, J.A., (2002). RNAi related mechanisms affect both transcriptional and posttranscriptional transgene silencing in *Drosophila*. *Mol. Cell* **9**: 315–327.
- Palazzoli, F., Testu, F.X., Merly, F., and Bigot, Y., (2010). Transposon tools: worldwide landscape of intellectual property and technological developments. *Genetica* **138**: 285–299.
- Pavlopoulos, A., Berghammer, A.J., Averof, M., and Klingler, M., (2004). Effective transformation of the beetle *Tribolium castaneum* using the *Minos* transposable element: quantitative and qualitative analysis of genomic integration events. *Genetics* **167**: 737–746.
- Peloquin, J.J., Thibault, S.T., Staten, R., and Miller, T.A., (2000). Germ-line transformation of pink bollworm (Lepidoptera: Gelechiidae) mediated by *piggyBac* transposable element. *Insect Mol. Biol.* **9**: 323–333.
- Perrimon, N., Ni, J.Q., and Perkins, L., (2010). In vivo RNAi: Today and tomorrow. *Cold Spring Harbor Perspect. Biol.* **2**: a003640.
- Peronnet, F., San Giorgioi, F., Lepesant, J.A., Deutsch, J.S., and Gonzy-Treboul, G., (2000). Three-partner conversion induced by the *P*-element transposase in *Drosophila melanogaster*. *Mol. Gen. Genet.* **262**: 1123–1131.
- Petrov, D.A., Schutzman, J.L., Hartl, D.L., and Lozovskaya, E.R., (1995). Diverse transposable elements are mobilized in hybrid dysgenesis in *Drosophila virilis*. *Proc. Natl. Acad. Sci. USA* **93**: 8050–8054.
- Pinkerton, A.C., Michael, K., O'Brochta, D.A., and Atkinson, P.W., (2000). Green fluorescent protein as a genetic marker in transgenic *Aedes aegypti*. *Insect Mol. Biol.* **9**: 1–10.
- Plasterk, R.H.A., Izsvák, Z., and Ivics, Z., (1999). Resident aliens. The *Tc1/mariner* superfamily of transposable elements. *Trends Genetics* **15**: 326–332.
- Plautz, J.D., Day, R.N., Dailey, G.M., Welsh, S.B., Hall, J.C., Halpain, S., and Kay, S.A., (1996). Green fluorescent protein and its derivatives as versatile markers for gene expression in living *Drosophila melanogaster*, plant and mammalian cells. *Gene* **173**: 83–87.
- Powell, J.R., and Gleason, J.M., (1996). Codon usage and the origin of *P* elements. *Mol. Biol. Evol.* **13**: 278–279.

- Rancourt, D.E., Peters, I.D., Walker, V.K., and Davies, P.L., (1990). Wolffish antifreeze protein from transgenic *Drosophila*. *BioTechnology* **8**: 453–457.
- Rio, D.C., Laski, F.A., and Rubin, G.M., (1986). Identification and immunochemical analysis of biologically active *Drosophila P* element transposase. *Cell* **44**: 21–32.
- Rio, D.C., Barnes, G., Laski, F.A., Rine, J., and Rubin, G.M., (1988). Evidence for *P* element transposase activity in mammalian cells and yeast. *J. Mol. Biol.* **200**: 411–415.
- Roberts, D.B., Ed., (1986). *Drosophila. A Practical Approach*. IRL Press, Oxford, UK.
- Robertson, H.M., (1995). The *Tc1-mariner* superfamily of transposons in animals. *J. Insect Physiol.* **41**: 99–105.
- Robertson, H.M., (1996). Members of the *pogo* superfamily of DNA-mediated transposons in the human genome. *Mol. Gen. Genet.* **252**: 761–766.
- Robertson, H.M., (1997). Multiple *mariner* transposons in flatworms and hydras are related to those of insects. *J. Hered.* **88**: 195–201.
- Robertson, H.M., and Lampe, D.J., (1995). Distribution of transposable elements in arthropods. *Annu. Rev. Entomol.* **49**: 333–337.
- Robertson, H.M., and Martos, R., (1997). Molecular evolution of the second ancient human *mariner* transposon, *Hsmar2*, illustrates patterns of neutral evolution in the human genome lineage. *Gene* **205**: 219–228.
- Robertson, H.M., and Zumpano, K.L., (1997). Molecular evolution of an ancient *mariner* transposon, *Hsmar1*, in the human genome. *Gene* **205**: 203–217.
- Rong, Y.S., and Golic, K.G., (2000). Gene targeting by homologous recombination in *Drosophila*. *Science* **288**: 2013–2018.
- Rubin, G.M., and Spradling, A.C., (1982). Genetic transformation of *Drosophila* with transposable element vectors. *Science* **218**: 348–353.
- Rubin, G.M., and Spradling, A.C., (1983). Vectors for *P* element-mediated gene transfer in *Drosophila*. *Nucleic Acids Res.* **11**: 6341–6351.
- Sarkar, A., Coates, C.J., Whyard, S., Willhoeft, U., Atkinson, P.W., and O'Brochta, D.A., (1997a). The *Hermes* element from *Musca domestica* can transpose in four families of cyclorrhaphan flies. *Genetica* **99**: 15–29.
- Sarkar, A., Yardley, K., Atkinson, P.W., James, A.A., and O'Brochta, D.A., (1997b). Transposition of the *Hermes* element in embryos of the vector mosquito, *Aedes aegypti*. *Insect Bioch. Mol. Biol.* **27**: 359–363.
- Schetelig, M.F., Scolari, F., Handler, A.M., Kittelmann, S., Gasperi, G., and Wimmer, E.A., (2009). Site-specific recombination for the modification of transgenic strains of the Mediterranean fruit fly *Ceratitis capitata*. *Proc. Natl. Acad. Sci. USA* **106**: 18171–18176.
- Schetelig, M.F., Gotschel, F., Viktorinova, I., Handler, A.M., and Wimmer, E.A., (2011). Recombination technologies for enhanced transgene stability in bioengineered insects. *Genetica* **139**: 71–78.
- Sentry, J.W., and Kaiser, K., (1992). *P* element transposition and targeted manipulation of the *Drosophila* genome. *Trends Genet.* **8**: 311–329.
- Sherman, A., Dawson, A., Mather, C., Gilhooley, H., Li, Y., Mitchell, R., Finnegan, D., and Sang, H., (1998). Transposition of the *Drosophila* element *mariner* into the chicken germ line. *Nat. Biotechnol.* **16**: 1050–1053.
- Shimizu, K., Kamba, M., Sonobe, H., Kanda, T., Klinakis, A.G., Savakis, C., and Tamura, T., (2000). Extrachromosomal transposition of the transposable element *Minos* occurs in embryos of the silkworm *Bombyx mori*. *Insect Mol. Biol.* **9**: 277–281.
- Siegal, M.L., and Hartl, D.L., (1996). Transgene coplacement and high efficiency site-specific recombination with the *Cre/loxP* system in *Drosophila*. *Genetics* **144**: 715–726.
- Simonelig, M., and Anxolabehere, D., (1991). A *P* element of *Scaptomyza pallida* is active in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* **88**: 6102–6106.

- Smit, A.F.A., and Riggs, A.D., (1996). Tiggers and other DNA transposon fossils in the human genome. *Proc. Natl. Acad. Sci. USA* **93**: 1443–1448.
- Smith, J., Grizot, S., Arnould, S., Duclert, A., Epinat, J.C., and Chames, P., et al. (2006). A combinatorial approach to create artificial homing endonucleases cleaving chosen sequences. *Nucleic Acids Res.* **34**: e149.
- Spradling, A.C., (1986). *P* element-mediated transformation. In: Roberts, D.B., Ed., *Drosophila. A Practical Approach*. IRL Press, Oxford, UK, pp. 97–175.
- Spradling, A.C., and Rubin, G.M., (1982). Transposition of cloned *P* elements into *Drosophila* germline chromosomes. *Science* **218**: 341–347.
- Spradling, A.C., and Rubin, G.M., (1983). The effect of chromosomal position on the expression of the *Drosophila* xanthine dehydrogenase gene. *Cell* **34**: 47–57.
- Spradling, A.C., Stern, D., Kiss, I., Roote, J., Laverty, T., and Rubin, G.M., (1995). Gene disruptions using *P* elements: An integral component of the *Drosophila* genome project. *Proc. Natl. Acad. Sci. USA* **92**: 10824–10830.
- Sundararajan, P., Atkinson, P.W., and O'Brochta, D.A., (1999). Transposable element interactions in insects: crossmobilization of *hobo* and *Hermes*. *Insect Mol. Biol.* **8**: 359–368.
- Szekely, A.A., Woodruff, R.C., and Mahendran, R., (1994). *P* element mediated germ line transformation of *Drosophila melanogaster* with the *Tc1* transposable DNA element from *Caenorhabditis elegans*. *Genome* **37**: 356–366.
- Taillebourg, E., and Dura, J.M., (1999). A novel mechanism for *P* element homing in *Drosophila*. *Proc. Natl. Acad. Sci. USA* **96**: 6856–6861.
- Takasu, Y., Kobayashi, I., Beumer, K., Uchino, K., Sezutsu, H., and Sajwan, S., et al. (2010). Targeted mutagenesis in the silkworm *Bombyx mori* using zinc finger nuclease mRNA injection. *Insect Biochem. Mol. Biol.* **40**: 759–765.
- Tamura, T., Thibert, C., Royer, C., Kanda, T., Eappen, A., and Kamba, M., et al. (2000). Germline transformation of the silkworm *Bombyx mori* L. using a *piggyBac* transposon-derived vector. *Nat. Biotechnol.* **18**: 81–84.
- Tkachuk, A., Kim, M., Kravchuk, O., and Savitsky, M., (2011). A new powerful method for site-specific transgene stabilization based on chromosomal double-strand break repair. *PLoS ONE* **6** (10): e26422.
- Tsien, R.Y., (1998). The green fluorescent protein. *Annu. Rev. Biochem.* **67**: 509–544.
- Venken, K.J.T., He, Y., Hoskins, R.A., and Bellen, H.J., (2006). P[acman]: a BAC transgenic platform for targeted insertion of large DNA fragments in *D. melanogaster*. *Science* **314**: 1747–1751.
- Wang, W., Swevers, L., and Ilatrou, K., (2000). *Mariner* (*Mos1*) transposase and genomic integration of foreign gene sequences in *Bombyx mori* cells. *Insect Mol. Biol.* **9**: 145–155.
- Woodruff, R.C., Thompson Jr., J.N., Barker, J.S.F., and Huai, H., (1999). Transposable DNA elements and life history traits: II. Transposition of *P* DNA elements in somatic cells reduces fitness, mating activity and locomotion of *Drosophila melanogaster*. *Genetica* **107**: 261–269.
- Yeh, E., Gustafson, K., and Boulian, G.L., (1995). Green fluorescent protein as a vital marker and reporter of gene expression in *Drosophila*. *Proc. Natl. Acad. Sci. USA* **92**: 7036–7040.
- Zhang, X., Zhang, J., and Zhu, K.Y., (2010). Chitosan/double-stranded RNA nanoparticles-mediated RNA interference to silence chitin synthase genes through larval feeding in the African malaria mosquito (*Anopheles gambiae*). *Insect Mol. Biol.* **19**: 683–693.

# Sex Determination in Insects

## Chapter Outline

- 10.1 Overview 420**
- 10.2 Introduction 420**
- 10.3 Costs and Benefits of Sexual Reproduction 421**
  - 10.3.1 Sexual Reproduction Has Costs 421
  - 10.3.2 Advantages of Sex Must Be Large 422
  - 10.3.3 Origin of Sex 423
- 10.4 Sex Determination Involves Soma and Germ-Line Tissues 424**
- 10.5 Sex Determination in *Drosophila melanogaster* 424**
  - 10.5.1 Dosage Compensation of X Chromosomes 425
  - 10.5.2 Somatic-Sex Determination 427
  - 10.5.3 Germ-Line Determination 430
- 10.6 Are Sex-Determination Mechanisms Diverse? 431**
  - 10.6.1 Intraspecific Variability 434
  - 10.6.2 Environmental Effects 436
  - 10.6.3 Postzygotic Sex Determination 437
- 10.7 A Single Model? 437**
- 10.8 Meiotic Drive Can Distort Sex Ratios 440**
  - 10.8.1 Segregation Distorter (SD) 440
  - 10.8.2 Distorter in Mosquitoes 441
  - 10.8.3 Female-Biased Sex Ratios in Stalk-Eyed Flies 441
  - 10.8.4 Meiotic Drive as a Pest-Management Tool? 441
- 10.9 Hybrid Sterility 442**
- 10.10 Medea in *Tribolium* 442**
- 10.11 Cytoplasmic Agents Distort Normal Sex Ratios 442**
  - 10.11.1 *Spiroplasma* Strains 445
  - 10.11.2 L-Form Bacteria 445
  - 10.11.3 *Rickettsia* 446
  - 10.11.4 *Wolbachia* 446
  - 10.11.5 *Cardinium* 449
- 10.12 Paternal Sex-Ratio Chromosomes and Cytoplasmic Incompatibility in *Nasonia* 450**
- 10.13 Male Killing in the Coccinellidae 450**

**10.14 Sex and the Sorted Insects 451**

10.14.1 Genetic Control 451

10.14.2 Genetic Improvement of Parasitoids 453

**10.15 Conclusion 454****References Cited 454**

## 10.1 Overview

Resolving the molecular genetic basis of sex determination in arthropods has applied applications that potentially could result in improved genetic-control programs for pest insects or in useful genetic modifications of beneficial species. Sex determination has been studied extensively in *Drosophila melanogaster*. Sex determination in *D. melanogaster* has three components: 1) dosage compensation, 2) somatic-sexual development, and 3) germ-line sexual development. The primary cue for determining sex in *D. melanogaster* is the number of X chromosomes relative to autosomes (A) in a cell (X:A ratio). This ratio determines somatic sex, germ-line sex, and dosage compensation by regulating the functions of sets of regulatory genes. One model suggested that sex determination in all insects is based on modifications of the *Drosophila* scheme in which a hierarchy of key regulatory genes control sexual development. At the top of the regulatory cascade is *Sex lethal<sup>+</sup>* (*Sxl<sup>+</sup>*), which must be ON to determine the female pathway. If *Sxl<sup>+</sup>* is OFF, the male pathway is the “default” developmental process. Sex subsequently is determined by the differential splicing of messenger RNAs. At the end of the pathway, a DNA-binding regulatory protein coded for by the *doublesex<sup>+</sup>* gene influences sex determination. Although sex determination in arthropods is clearly determined by chromosomal and genetic processes, environment also plays a role in some. Furthermore microorganisms, including *Wolbachia*, *Rickettsia*, *Cardinium*, *Spiroplasma*, and viruses, are able to modify sex determination or sex ratio in many arthropods.

## 10.2 Introduction

Sexual reproduction results in genetic variability through crossing over, as well as variation due to Mendel’s laws of Independent Assortment and Independent Segregation. As a result of meiosis, genes obtained from different parents can be combined in a single descendant (review the discussion of meiosis in Chapter 3). New genotypes thus are constructed from preexisting variability by the mechanisms of **segregation**, **assortment**, and **recombination** during meiosis. Homologous chromosomes separate (segregate) randomly to yield haploid gametes containing chromosomes derived from both the individual’s maternal and paternal genomes. Recombination occurs during crossing over between

paired homologous chromosomes during meiosis and results in new assortments of alleles. Genes assort independently from each other unless they are closely linked on the chromosome.

Understanding sex determination and sex allocation in insects has both fundamental and applied applications. The evolutionary advantages of sexual reproduction and the reasons why organisms vary the sex ratio of their progeny are among the most-discussed topics in evolutionary biology (see [Hamilton 1967](#), [Hartl and Brown 1970](#), [Maynard Smith 1978](#), [Charnov 1982](#), [Bull 1983](#), [Thornhill and Alcock 1983](#), [Lewis 1987](#), [Michod and Levin 1988](#), [Hamilton et al. 1990](#), [Wrensch and Ebbert 1993](#), [Crow 1994](#), [Barton and Charlesworth 1998](#), [Marin and Baker 1998](#), [Mittwoch 1996](#), [Partridge and Hurst 1998](#), [Werren and Beukeboom 1998](#), [Antolin 1999](#), [Keightley and Eyre-Walker 2000](#), [West et al. 2000](#), [Rice and Chippindale 2001](#), [Schurko et al. 2008](#)).

## 10.3 Costs and Benefits of Sexual Reproduction

Theories about the evolution of sex have focused on the advantages of the combined effects of segregation and recombination. In a sexual population, advantageous mutations that arise at two different loci in two parents can be combined in one individual in later generations.

### 10.3.1 Sexual Reproduction Has Costs

Despite the advantages of sex, sex also has costs. [Crow \(1994\)](#) summarized these costs as follows: 1) sex expends energy that could be used for other purposes; 2) males are expensive, with a 50% savings possible if males were eliminated (although [Lehtonen et al. \[2012\]](#) argue that the cost of males is not always two-fold); 3) sexual selection in sexual species often leads to maladapted traits and expensive competition for mates; 4) sexual species do not allow the perpetuation or fixation of novel genetic types because heterozygotes often are broken up by segregation and recombination. Changes in ploidy (such as triploids and aneuploids) cannot go through meiosis successfully; 5) sexual species have to find a mate, which can be a disadvantage in sparse populations, or species with limited motility, or colonizers of a new area; 6) sexual species are prone to sexually transmitted diseases and harmful transposons whose spread is facilitated by biparental inheritance; 7) short-term selection is slower in sexual than asexual species; and 8) sexual species cannot colonize microhabitats without the distinctive properties of these adapted colonies being swamped by hybridization. Another negative aspect is genome dilution, in which only half the parental genome is transferred to offspring ([Lehtonen et al. 2012](#)).

### 10.3.2 Advantages of Sex Must Be Large

At least 20 hypotheses have attempted to explain why sexual reproduction persists (Crow 1994, Schurko et al. 2008, Lehtonen et al. 2012). 1) One possibility is that sex provides an ability to **incorporate and accumulate favorable mutations**. Mutations that arise in an asexual species in different individuals cannot be combined in one individual easily; successive advantageous mutations would have to occur in the same asexual lineage, one after the other. 2) Sexual reproduction may allow the accumulation of favorable mutations even when deleterious mutations are present (whereas the maintenance of a favorable mutation in an asexual population is dependent upon the relative fitness of the individual in which the mutation occurs). Thus, the value of sex “lies more in the ability to reassort existing genes as the environment changes and in the elimination of harmful mutations” (Crow 1994).

Sexual reproduction allows harmful mutations to be eliminated. This effect is based on a concept termed **Muller’s ratchet**. Muller (1964) noted that in an asexual population, unless it is very large, it is unlikely that any individual is free of harmful mutations. In such a population, the most-fit individual is one that has only one mutation. In the next generation, mutations occur again and this time the most-fit individual may have one new mutation, or two in total, and the “ratchet” has turned another cog. In the absence of reverse mutation, such a population would accumulate more and more deleterious mutations. In a sexual population, a mutant-free type can be created by recombination.

Thus, sexual reproduction can reduce the mutational load. The deleterious effects of mutations are related to their frequency of occurrence, not to the magnitude of their effects. The smaller the effect of a mutation, the more individuals it will affect before it is eliminated from the population. Over time, the number of mutations removed per eliminated individual is much larger in a sexual than in an asexual population and the mutation load is reduced correspondingly. Sexual reproduction may provide a means by which some transposable elements (TEs) can be eliminated from the population, although it could require thousands of generations and would be affected by the rate of transposition and excision of TEs, strength of selection, and population size (Lehtonen et al. 2012).

Sexual reproduction and diploidy have evolutionary advantages by providing protection from somatic mutations (Crow 1994). Diploidy is common in higher organisms, which have the most extensive and highly differentiated soma. Each generation somatic tissues develop and are identical, except for somatic mutations, which are often recessive. Diploidy thus diminishes the deleterious effects of recessive somatic mutations, which could destroy essential cells or initiate

abnormal growth. Diploidy requires passing through a single cell each generation so the soma can begin anew without these mutations. Diploidy also permits the efficient repair of double-stranded breaks in DNA. However, although >20 hypotheses have been proposed as to why sexual reproduction is maintained, we still do not know why it is maintained by evolution (Schurko et al. 2008).

### 10.3.3 Origin of Sex

The reasons for maintaining sexual reproduction in current populations are “likely to be quite different from the mechanisms by which sex got started in the first place” (Crow 1994). It is generally accepted that sex was determined initially by an allelic difference at a gene located on a homologous pair of autosomes (Rice 1994, Lucchesi 1999). The two sexes thus consisted of individuals heterozygous or homozygous at this sex-determining locus. The transformation of autosomes bearing the sex-determining gene into heteromorphic sex chromosomes (such as X and Y) is thought to have occurred by the accumulation of mutations in the neighborhood of the sex-limited allele. The retention of such mutations could be facilitated by a reduction in the rate of recombination in the chromosome of the individuals bearing the sex-limited allele.

Understanding the mechanisms of sex determination in insects provides insights into the regulation of development of a significant character in eukaryotes. Such knowledge could provide useful tools for the genetic improvement of arthropod natural enemies of pest arthropods and weeds, genetic modification of pests, or improve the methods by which genetic-control programs are achieved (LaChance 1979, Shirk et al. 1988, Stouthamer et al. 1992, Grenier et al. 1998, Heinrich and Scott 2000, Robinson and Franz 2000). Also see Chapter 14 for a discussion of genetic manipulation of pest and beneficial arthropods.

Schurko et al. (2008) noted that it is not always easy to determine whether a species reproduces sexually. An example is the fungus-growing ant *Mycoceropurus smithii* that long has been thought to be asexual (thelytokous) (Rabeling et al. 2011). However, when 234 populations of *M. smithii*, from 39 localities ranging in distribution from Mexico to Argentina and some Caribbean islands, were evaluated using 12 microsatellite markers both thelytokous (89.7%) and sexual (10.3%) populations were found. Males appeared to be absent from thelytokous populations, but sperm was found in the reproductive tracts of queens in the sexual populations. Thelytoky was assumed to be occurring if all individuals in the population tested shared genotypes, whereas sexuality was assumed if there was an increase in unique multilocus genotypes, indicating genetic recombination had occurred by sexual reproduction. The genetic uniformity across all

loci within colonies suggests mitotic parthenogenesis (**apomixis** or reproduction without meiosis) is occurring. Four populations from the Amazonian area of South America appear to reproduce both sexually and asexually because the progeny were a mixture of recombinant and clonal genotypes. Spermathecae of four queens were dissected and haploid sperm were present, confirming sexual reproduction was occurring, although no males were found. Thus, this species appears to be a mosaic of asexual and sexually reproducing populations, with sex being lost repeatedly in different lineages. The sexual populations are common in the center of this species' geographic distribution where there are high local population densities. The asexual lineages have "rapidly dispersed throughout much of Latin America, leading to the current widespread geographic distribution." This ant has the most extensive geographic distribution of any fungus-growing species, indicating that both sexual and asexual populations have sufficient fitness to persist. The authors note that it is possible that the loss of a single recessive allele can cause thelytoky and suggest that the high propensity for switching from sexual to asexual reproduction in *M. smithii* may be due to a small number of genes. Alternatively, it could be due to parthenogenesis-inducing symbionts.

## 10.4 Sex Determination Involves Soma and Germ-Line Tissues

Sex determination involves both the soma and germ-line tissues (ovaries and testes). Sexual dimorphism in adult insects is often extreme, with differences in setal patterns, pigmentation, external genitalia, internal reproductive systems and behavioral patterns (Greenspan and Ferveur 2000, Kopp et al. 2000).

How do sexually determined differences in the soma and germ line arise? The details are pretty clear for *D. melanogaster*, and some information is available for other economically important insects such as fruit flies, moths, honey bees, and mosquitoes.

First, let's review the basic sex-determination system in *D. melanogaster*. Then, sex determination in other insects is described. Finally, examples are provided that illustrate the importance of extrachromosomal and microbial genes in modifying sex in many arthropods.

## 10.5 Sex Determination in *Drosophila melanogaster*

Developing an understanding of sex determination in *D. melanogaster* has relied on identifying a relatively few spontaneous mutants (Table 10.1), which indicated that the number of genes involved is relatively low (Belote et al. 1985, Slee and Bownes 1990, Cline and Meyer 1996). Sex determination in *D. melanogaster*

**Table 10.1: Some Genes Involved in Somatic Sex Determination and Dosage Compensation in *D. melanogaster*.**

Gene	Function
<b>Maternal genes</b>	
<i>daughterless</i> <sup>+</sup> <i>hermaphrodite</i> <sup>+</sup> <i>extramachrochaetae</i> <sup>+</sup> <i>groucho</i> <sup>+</sup>	Necessary for numerator genes to act appropriately; <i>da</i> <sup>+</sup> and <i>her</i> <sup>+</sup> activate <i>Sxl</i> <sup>+</sup> in female embryos. <i>emc</i> <sup>+</sup> and <i>gro</i> <sup>+</sup> negatively regulate <i>Sxl</i> <sup>+</sup> in female embryos.
<b>Numerator genes</b>	
<i>sisterless-A</i> <sup>+</sup> <i>sisterless-B</i> <sup>+</sup> <i>sisterless-C</i> <sup>+</sup> <i>runt</i> <sup>+</sup>	Communicate X-chromosome dose in dosage compensation. X-linked genes involved in activating <i>Sxl</i> <sup>+</sup> in females; they “count” the number of X chromosomes and turn on <i>Sex-lethal</i> <sup>+</sup> .
<b>Zygotic genes</b>	
<i>Sex-lethal</i> <sup>+</sup>	Major control gene; produces a full-length protein in females; no protein produced in males.
<i>transformer</i> <sup>+</sup>	Active, with <i>tra-2</i> <sup>+</sup> ; in regulating <i>dsx</i> <sup>+</sup> in females.
<i>transformer-2</i> <sup>+</sup>	Active in females to induce female-specific <i>dsx</i> <sup>+</sup> expression and repress male-specific <i>dsx</i> <sup>+</sup> expression. Needed for spermatogenesis.
<i>doublesex</i> <sup>+</sup>	Active in males to repress female differentiation; in females <i>dsx</i> <sup>+</sup> represses male differentiation; loss of function mutants result in intersexes in both males and females; a pivotal terminal-differentiation switch.
<i>intersex</i> <sup>+</sup>	Active in females with <i>dsx</i> <sup>+</sup> product to repress male differentiation, not needed in males.
<i>fruitless</i> <sup>+</sup>	Gene product is necessary in nervous system of males to elicit normal mating behavior and development of male muscle (muscle of Lawrence).
<i>male-specific lethal</i> <sup>+</sup> <i>msl-1</i> <sup>+</sup> <i>msl-2</i> <sup>+</sup> <i>msl-3</i> <sup>+</sup>	All four genes regulate X chromosome transcription in males; <i>msl-2</i> <sup>+</sup> has no function in females. Absence of the MSL-2 protein in females prevents formation of the compensasome.

(Derived from Bownes 1992, Cline and Meyer 1996, Li and Baker 1998, Marin and Baker 1998.)

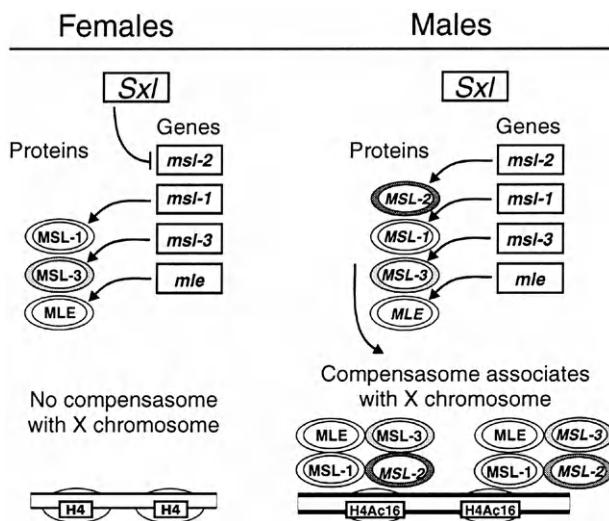
involves three major components: dosage compensation, somatic-cell differentiation, and germ-line differentiation. More is known about dosage compensation and somatic-cell differentiation than about germ-line differentiation.

### 10.5.1 Dosage Compensation of X Chromosomes

A basic aspect of sex determination in insects with an XY sex-determining system is dosage compensation of the X chromosomes because Y-chromosomes typically have few genes and consist primarily of heterochromatin (Baker et al. 1994). The mechanism of dosage compensation varies in arthropods. **Dosage compensation** equalizes the amount of gene products produced by individuals with an XX/XY

genetic system (males containing one and females two X chromosomes). Dosage compensation in *D. melanogaster* is achieved by **hypertranscription** of the single X chromosome in males (Marin et al. 2000, Lucchesi 2009). As a result, males produce **equivalent** amounts of gene product as females that have two X chromosomes. By contrast, dosage compensation in the mole cricket *Gryllotalpa fassor* is analogous to that in mammals; one of the two X chromosomes in females is transcriptionally **inactivated**, with the inactivation occurring randomly within each cell (Rao and Padmaja 1992). *Gryllotalpa fassor* males are XO and females are XX, and one of the two X chromosomes in female cells is late replicating and transcriptionally silent.

Males that are XY:AA are **aneuploid** for an X, which is a large fraction of the total genome (because the Y has only a few genes on it). (AA indicates that there are two sets of autosomes.) Aneuploidy (when the chromosomal composition in a cell is not an exact multiple of the haploid set) is normally lethal to an organism. In *Drosophila* males, **hypertranscription** of the single X chromosome requires the functions of autosomal genes, *male-specific lethal genes<sup>+</sup>* (*msl<sup>+</sup>*), which are under the control of *Sxl<sup>+</sup>* (Table 10.1, Figure 10.1) and some RNAs that are associated with the chromatin (Kelley and Kuroda 2000). The MSL proteins are assembled with the RNAs in a remodeling complex (called a **compensasome**) on  $\approx 100$  sites on the



**Figure 10.1** Model for the regulation of dosage compensation of X chromosomes in *D. melanogaster* males. MLE, MSL-I, and MSL-3 proteins are produced in both sexes. The *Sxl<sup>+</sup>* gene negatively regulates another gene so that no functional protein, probably MSL-2, is made in females. In the absence of this protein, the MLE, MSL-1, and MSL-3 proteins cannot associate stably as a “compensasome” with the X chromosome in females. The compensasome does associate with sites on the X chromosome in males that have the histone H4 acetylated at the lysine 16 position. As a result, the sole X chromosome is hypertranscribed in males.

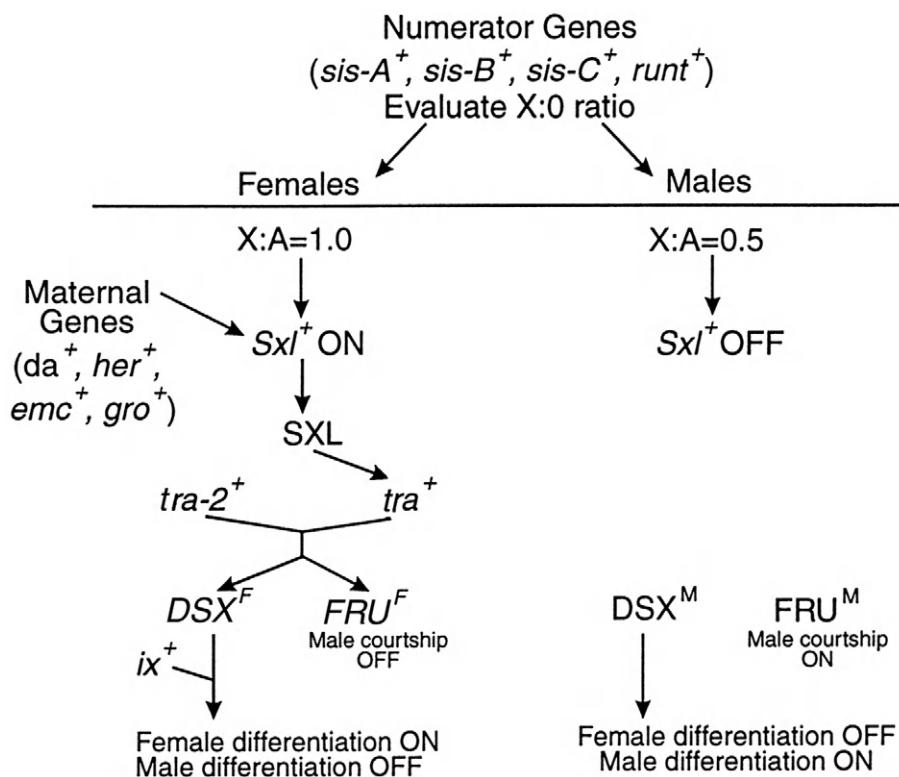
X chromosome in males (Figure 10.1). The resulting histone H4 acetylation leads to hypertranscription (Marin et al. 2000, Smith et al. 2000, Lucchesi 2009, Laverty et al. 2010). By contrast, the SXL protein in females prevents removal of a female-specific intron in the *msl-2<sup>+</sup>* mRNA; without this MSL-2 protein the other MSL proteins fail to assemble on the X chromosome and hyperactivation is prevented.

### 10.5.2 Somatic-Sex Determination

The relative number of X chromosomes and autosomes (A) in *D. melanogaster* was considered to be responsible for the primary step in sex determination immediately after fertilization. It was thought that cells with two X chromosomes and two sets of autosomes (2X:2A, or a ratio of 1.0) develop into females, while diploid cells (with 1X:2A, or a ratio of 0.5) develop into males (Figure 10.2). Flies with equal numbers of X chromosomes and autosomes (XX:AA, XXX:AAA, or X:A, or a ratio of 1.0) develop as females. Flies with an intermediate X:A ratio (XX:AAA) develop as intersexual flies that appear to be mosaics of discrete patches of male or female tissues. However, Erickson and Quintero (2007) suggest that X chromosome dose, not the X:A ratio, signals which sex will develop in *Drosophila*. Distinct boundaries between cells exist in *Drosophila* and sex determination takes place in individual cells (**cell-autonomous sex determination**). Cell-autonomous determination of sex has been assumed to be due to the lack of **sex hormones** in insects. However, DeLoof and Huybrechts (1998) note that hormones are involved in the development of secondary sexual characters in some insects.

The Y chromosome does not determine sex in *D. melanogaster*, although it is required for normal spermatogenesis and fertility. At least six genes on the Y chromosome are important in fertility, each performing a single, unique function (Hennig 1993, Hochstenbach et al. 1994). The fertility genes are >1000kb each and are highly susceptible to mutations (Hackstein and Hochstenbach 1995). The Y chromosome is important in sex determination in some insects (Marin and Baker 1998).

**Numerator genes** communicate the relative number or ratio of X chromosomes and autosomes in *D. melanogaster*. Numerator genes in *D. melanogaster* include *sisterless-A<sup>+</sup>* (*sis-A*), *sisterless-B<sup>+</sup>* (*sis-B*), *sisterless C<sup>+</sup>* (*SisC*), and *runt<sup>+</sup>* (Cline and Meyer 1996). Once the X:A ratio is assessed, the activities of a relatively small number of major regulatory genes are triggered that ultimately lead to male or female differentiation in the soma (Figure 10.2). Somatic sexual differentiation is regulated through a cascade of sex-specific events in which RNA transcripts are differentially spliced in males and females (Figure 10.2, Table 10.1). Note that in Figure 10.2 gene products (proteins) are capitalized (SXL), whereas the gene (*Sxl<sup>+</sup>*) is italicized.



**Figure 10.2** The general features of somatic sex determination in *D. melanogaster*. The ratio of X chromosomes to autosomes (A) determines whether *Sxl*<sup>+</sup> is ON. *Sxl*<sup>+</sup> produces a protein, *SXL*, that acts as a splicing factor on the RNA produced by the *tra*<sup>+</sup> gene, resulting in the production of active *TRA* protein. *TRA*, together with the product of the *tra-2*<sup>+</sup> gene, determine the female-specific splicing of the *dsx*<sup>+</sup> and *fruitless*<sup>+</sup> RNAs, which results in a cascade of genes functioning to produce a female. If *Sxl*<sup>+</sup> is OFF, the individual becomes a male because male-specific products (*DSX*<sup>M</sup> and *FRU*<sup>M</sup>) of the *dsx*<sup>+</sup> and *fru*<sup>+</sup> genes are produced.

*Sex-lethal*<sup>+</sup> is a key switch gene that, very early in development, affects both somatic sexual differentiation and dosage compensation in *D. melanogaster* (Figure 10.2). *Sex lethal*<sup>+</sup> codes for an RNA-splicing enzyme. Its action on the next gene in the cascade, *transformer*<sup>+</sup>, is restricted to females in its role in sexual differentiation. *Sex-lethal*<sup>+</sup> must be ON in females and OFF in males (Figure 10.2). Once the X:A ratio is read and the *Sxl*<sup>+</sup> gene is turned ON or OFF early in embryonic development, the developmental path chosen is stable (Cline and Meyer 1996).

*Sxl*<sup>+</sup> is transcribed in *D. melanogaster* females in a complex manner. Two different promoters function in somatic cells; one, the establishment promoter, *Sxl*<sup>Pe</sup>, acts very early and only for a brief period during nuclear cell cycle 12 to

early cell cycle 14, ending when somatic cells first form in the young embryo. As this promoter shuts off, the second promoter,  $SxI^{Pm}$ , comes on in both sexes. However, because the transcripts from this promoter require full-length SXL protein to remove a male-specific exon, only the expression of the  $SxI^{Pm}$  in females generates messenger RNAs (mRNAs) that encode the full-length SXL protein. Thus, the earliest  $SxI^+$  transcripts differ from later transcripts and male transcripts are inactive because they include an extra exon that stops the translation process. Initiation of  $SxI^+$  expression requires the action of genes from the mother (maternal genes such as  $da^+$ ,  $her^+$ ,  $emc^+$ , and  $gro^+$ ) (Figure 10.2). In males, the  $SxI^+$  master gene is OFF, and the four *male-specific-lethal<sup>+</sup>* autosomal genes are ON, a combination that leads to male somatic-sexual differentiation and hypertranscription of the single X chromosome (Figure 10.1).

Maternal-effect genes influence the development of progeny, as discussed in Chapter 4. Maternal-effect genes function in one of two ways: either the mother produces a gene product that is transferred into and stored in the egg, or the mother's messenger RNA is transferred into and stored in the eggs and subsequently is translated by the embryo. At least four maternal X:A signal-transduction genes have been found, including *daughterless<sup>+</sup>* ( $da^+$ ), *hermaphrodite<sup>+</sup>* ( $her^+$ ), *extramachrochaetae<sup>+</sup>* ( $emc^+$ ), and *groucho<sup>+</sup>* ( $gro^+$ ) (Figure 10.1, Table 10.1). Female progeny of mothers with mutant forms of  $da^+$  fail to activate the key master gene  $SxI^+$  and die as embryos. Male progeny of  $da$  mothers survive because they do not require  $SxI^+$ .

$SxI^+$  is the master switch gene involved in both sex determination and dosage compensation in *D. melanogaster*. It regulates pre-mRNA splicing for itself and for *transformer<sup>+</sup>* ( $tra^+$ ) and *male-specific-lethal-2<sup>+</sup>* ( $msl-2^+$ ). Once  $SxI^+$  is ON in females, a second series of regulatory genes are important in differentiating between the alternative pathways in somatic cell development. These secondary switch genes include *transformer<sup>+</sup>* ( $tra^+$ ), *transformer-2<sup>+</sup>* ( $tra-2^+$ ), *intersex<sup>+</sup>* ( $ix^+$ ), *doublesex<sup>+</sup>* ( $dsx^+$ ), and *fruitless<sup>+</sup>* ( $fru^+$ ) (Figure 10.2). Mutations of  $tra^+$ ,  $tra-2^+$ , and  $ix^+$  affect somatic-sex determination in females, but are not needed for male somatic differentiation (Table 10.1). In the absence of TRA proteins in males, the *fruitless<sup>+</sup>* gene transcript affects as many as 500 neurons in the brain (Figure 10.2), which regulates male sexual behavior and also affect the male-specific muscle (Muscle of Lawrence, MOL) used in mating. Although  $tra-2^+$  is not needed for male differentiation, it is critical for normal spermatogenesis in males.

The *doublesex<sup>+</sup>* locus is needed for differentiation of both male and female external morphology (Figure 10.2). The  $dsx^+$  gene is a **double switch**, with only one switch functioning in a particular sex. When  $dsx^+$  is active in males

(producing the male gene product,  $DSX^M$ ), it represses female differentiation. When *doublesex<sup>+</sup>* is active in females (producing  $DSX^F$ ), and the *intersex<sup>+</sup>* gene product is present, male development is suppressed. If *dsx<sup>+</sup>* is inactivated, both male and female genes are active within a cell, which results in an intersexual phenotype at the cellular level.

The determination of sex during embryogenesis in *D. melanogaster* is transmitted through a hierarchy of regulatory genes to the terminal differentiation genes, whose products are responsible for the sexually dimorphic traits of the adult fly (Bownes 1992). The different activities of the regulatory genes in males and females are largely due to sex-specific differences in RNA splicing that lead to the production of functionally different transcripts in the two sexes (Baker 1989). The individual genes in this regulatory hierarchy are not only themselves controlled at the level of RNA splicing but, in turn, specify the splicing pattern of the transcripts of genes down stream in the hierarchy, producing a cascade of RNA splicing reactions. Thus, RNA processing, involving alternative splicing, is a crucial regulatory mechanism in this developmental pathway.

In addition to the sex-determination genes, there are genes whose products are responsible for the structure and function of sexually dimorphic somatic tissues (Kopp et al. 2000). The sex-determination regulatory pathway regulates many structural genes, such as the yolk polypeptide genes, which are expressed in the fat body in a female-specific manner.

### 10.5.3 Germ-Line Determination

Sex determination in the development of germ-line tissues in *D. melanogaster* is different from that in the soma (Pauli and Mahowald 1990, Janzer and Steinmann-Zwicky 2001, Vincent et al. 2001). Pole cells in the embryo are segregated into the posterior pole of the insect embryo before cellular blastoderm, and they include the progenitors (stem cells) of the germ cells (Xie and Spradling 2000).

Components of the germ plasm (=pole plasm) are synthesized in the mother during oogenesis by a cluster of 15 nurse cells, which are connected to the oocyte at its anterior by cytoplasmic bridges. Pole plasm components are transported into the oocyte and translocated to the posterior pole of the egg. Maternally active genes important in the production of pole cells include *cappuccino<sup>+</sup>*, *spire<sup>+</sup>*, *staufen<sup>+</sup>*, *oskar<sup>+</sup>*, *vasa<sup>+</sup>*, *valois<sup>+</sup>*, *mago nashi<sup>+</sup>*, and *tudor<sup>+</sup>* (Ephrussi and Lehmann 1992). These genes also are important in the formation of normal abdomens in *Drosophila*.

During embryogenesis, prospective male- and female-germ cells are indistinguishable, but differentiation begins during the larval stage, when male gonads

grow larger than female gonads because they contain more germ cells. The sexual identity of germ cells is determined by both the X:A ratio of the germ cells and the X:A ratio of the surrounding soma (Cline and Meyer 1996). The expression of *Sxl<sup>+</sup>* in the soma is required in the female germ line. Three genes, *ovo<sup>+</sup>*, *ovarian tumor<sup>+</sup>*, and *sans fille<sup>+</sup>*, are important for growth and differentiation of female germ cells. Thus, activation and splicing of *Sex lethal<sup>+</sup>* in the ovary is regulated by a different set of proteins from those in the soma (Vincent et al. 2001).

*Drosophila* males have fewer abdominal segments than females, which is due to the sex-specific and segment-specific regulation of the Wingless morphogen. *Wingless<sup>+</sup>* expression is suppressed in the developing terminal male abdominal segment by a combination of the Hox protein Abdominal-B and the sex-determination regulator *doublesex<sup>+</sup>*, which results in cell death and suppression of cell proliferation of the terminal segment (Wang et al. 2011).

## 10.6 Are Sex-Determination Mechanisms Diverse?

Sex-determination mechanisms in insects appear to be diverse (White 1973, Lauge 1985, Retnakaran and Percy 1985, Wrensch and Ebbert 1993, Werren and Beukeboom 1998). Many insects have a **genetic sex-determination** system, with genetic differences determining maleness or femaleness. Others appear to have an **environmental sex-determination** system, in which there are no genetic differences between males and females but temperature or host conditions determine the sex. For example, in a few insects the hemolymph of the mother determines the sex of the offspring.

Ploidy levels sometimes are important in sex determination: both sexes of many arthropods are diploid ( $2n$ , diplo-diploidy), whereas others have haploid males and diploid females ( $n$  and  $2n$ , haplo-diploidy or **arrhenotoky**). Haplo-diploidy has apparently developed at least 10 times independently in insects (Normark 2004). Other species consist primarily of diploid females (**thelytoky**) and haploid males rarely are produced. In some species, haploid males are produced by the loss of paternally derived chromosomes after fertilization (known as **parahaploidy** or **pseudoarrhenotoky** or **male gamete loss**).

Males in apterygote and many pterygote insects are **heterogametic** (males are  $XO$ ,  $XY$ ,  $XXO$ ,  $XXY$ , or  $XYY$  and females are  $XX$ ), but in some higher pterygotes (Trichoptera, Lepidoptera) females may be the **heterogametic** ( $ZW$ ) sex. Mank (2009) noted that silk moths, in which females are heterogametic ( $ZW$ ), lack a global dosage-compensating mechanism for the  $Z$  chromosome, equalizing transcription for only some genes. It appears that other lepidopteran species also do not have globally dosage-compensated sex chromosomes (Mank 2009).

Sex-determining chromosomes (Y or W) differ from autosomes in that there is no genetic recombination, they are male- or female-specific, and usually contain a large amount of heterochromatic DNA and few functional genes (Kaiser and Bachtrog 2010). X chromosomes are female-biased and Z chromosomes are male-biased in their mode of inheritance and are hemizygous in the heterogametic sex.

Several different models have been proposed to explain sex determination in the haplo-diploid Hymenoptera (Cook 1993, Beukeboom 1995, Dobson and Tanouye 1998a, Heimpel and de Boer 2008). In the honey bee and the parasitoid *Bracon* (=*Habrobracon*) *hebetor*, sex is determined by a series of alleles at a single locus (**single-locus, multiple-allele model**) (Whiting 1943). In honey bees, the single locus (probably located on chromosome 8) has several alleles (19 so far). Individuals that are heterozygous for this locus are normal fertile (diploid) honey bee females, hemizygotes (unfertilized haploid eggs) became fertile drones (males), and homozygotes are sterile diploid males with degenerated testes containing reduced quantities of diploid sperm (Beye et al. 1996, 1999). Usually, worker bees eat any diploid honey bee males.

Under the single-locus, multiple-allele model, inbreeding should produce homozygous (diploid) males in the parasitoid *Bracon hebetor*. However, Heimpel et al. (1999) evaluated the diversity of sex-determining alleles in five U.S. populations and estimated that there were at least 20 alleles. This high allelic diversity suggests that the production of diploid males should be rare unless extreme inbreeding occurs.

In the parasitoid *Nasonia vitripennis*, which is arrhenotokous, no complementary sex-determination locus was found (Beukeboom et al. 2007). The *doublesex<sup>+</sup>* gene is homologous with *dsx<sup>+</sup>* from other insects, including *Apis*, and differential splicing occurs in males and females, indicating it is involved in somatic sex differentiation (Oliveira et al. 2009). However, maternal mRNA, in combination with zygotic transcription is essential for female development. Thus, males develop as a result of **maternal imprinting** that prevents zygotic transcription of the *transformer<sup>+</sup>* gene in unfertilized eggs. However, if fertilized, transcription of *tra<sup>+</sup>* is initiated, which autoregulates the female-specific *doublesex<sup>+</sup>* transcript, leading to female development (Verhulst et al. 2010a) (Table 10.2).

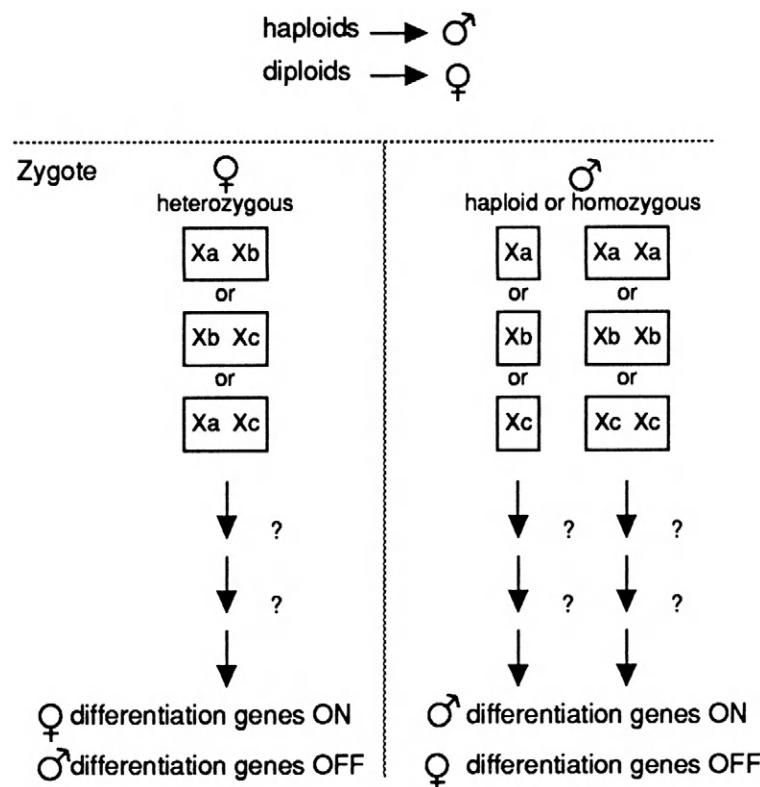
The haplo-diploid turnip sawfly, *Athalia rosae ruficornis*, when inbred, produces both diploid and triploid males, suggesting that sex in this hymenopteran is determined by the single-locus, multiple-allele system (Figure 10.3, Naito and Suzuki 1991). This sex-determination system also has been found in the

**Table 10.2: Some Mechanisms of Parthenogenesis in Insects and Mites.**

Type	Example: Symbiont type, if present	Reference(s)
<b>Arrhenotoky</b>		
Imprinting and <i>complementary sex determiner</i> gene ( <i>csd</i> ) in heterozygous state in females plus <i>feminizer</i> gene needed for females	<i>Apis mellifera</i> , No symbiont	Hasselmann et al. 2008
Imprinting prevents transcription of <i>transformer</i> <sup>+</sup> in unfertilized eggs resulting in males	<i>Nasonia vitripennis</i> , No symbiont	Verhulst et al. 2010a
Single-locus multiple-allele	<i>Athalia rosae</i> , No symbiont	Naito and Suzuki 1991
<b>Thelytoky</b>		
Apomixis	<i>Neochrysocharis formosa</i> , <i>Rickettsia</i>	Adachi-Hagimori et al. 2008
Automixis	<i>Leptopilina clavipes</i> , <i>Wolbachia</i>	Pannebakker et al. 2004
Feminization of males	<i>Trichogramma</i> , <i>Wolbachia</i> <i>Brevipalpus phoenicis</i> , <i>Cardinium</i> <i>Eurema hecate</i> (ZZ males feminized), <i>Wolbachia</i> <i>Zygindia pullata</i> (XO males feminized), <i>Wolbachia</i>	Huigens et al. 2000 Weeks et al. 2001 Hiroki et al. 2002, Narita et al. 2007 Negri et al. 2006, 2009
Male killing	<i>D. melanogaster</i> , <i>Spiroplasma</i>	Veneti et al. 2005
Maternal imprinting prevents female development in haploid eggs	<i>Adalia bipunctata</i> , <i>Rickettsia</i>	Werren et al. 1994
Single-locus recessive trait in nuclear genome restores diploidy by automixis	<i>Nasonia</i>	Verhulst et al. 2010b
Methylation and DNA amplification and genomic rearrangements	<i>Lysiphlebus fabarum</i> , No symbiont  <i>Apis mellifera capensis</i> <i>Bombus terrestris</i> , No symbiont	Sandrock and Vorburger 2011 Lattorff et al. 2005, 2007 Bigot et al. 2010

parasitoid *Diadromus pulchellus* (El Agoze et al. 1994); the bee *Apis cerana*, the sawfly *Neodiprion nigroscutum*, the red imported fire ant *Solenopsis invicta*, the stingless bee *Melipona quadrifasciata* (Cook 1993), and the parasitoid *Diadegma chrysostictos* (Butcher et al. 2000).

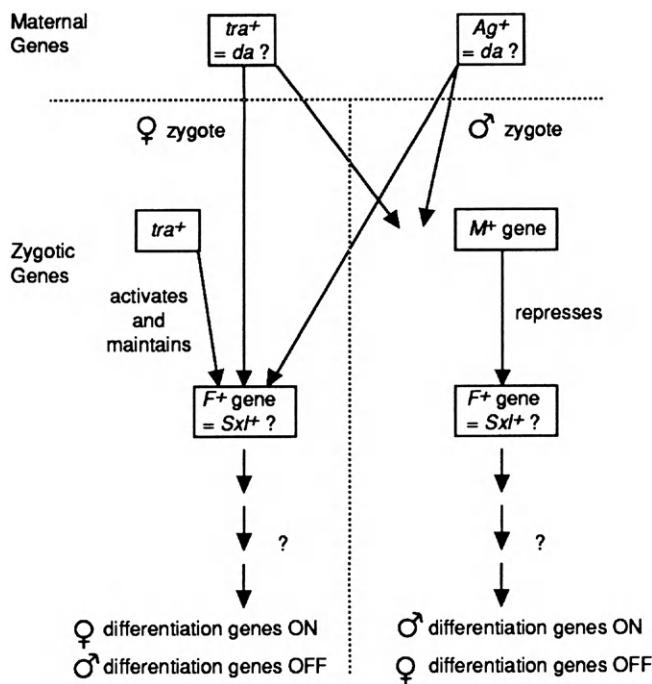
In other haplo-diploid Hymenoptera, sex is determined by a number of alleles at a series of loci (**multiple-locus, multiple-allele model**). According to this model, females must be heterozygous at one or more loci, whereas haploid males are hemizygous (Crozier 1971). After inbreeding, some diploid individuals are produced and these are males if they are homozygous for all loci. These two “multiple-allele” models can be combined if the assumption is made that the single-locus model is a special case of the multiple-locus, multiple-allele model. Under this assumption, only one locus has an effect in the first model.



**Figure 10.3** Multiple alleles at a single locus determine sex in the haplo-diploid hymenopteran *Athalia rosae*. Under normal conditions, males are haploid and females are diploid, with females heterozygous for the sex-determining locus (X), whereas males are hemizygous. If individuals become homozygous for an allele of X, perhaps through inbreeding, they become diploid males. (Redrawn from Bownes 1992.)

### 10.6.1 Intraspecific Variability

Within a single species, several different sex-determining mechanisms may occur (Dubendorfer et al. 2002). Many populations of the house fly, *Musca domestica*, have five pairs of autosomes and a pair of heterochromatic sex chromosomes; thus, females are XX and males are XY (Figure 10.4). In these populations, sex is determined by the presence or absence of the Y, which carries a male-determining factor *M*; the X plays no important role in sex determination. In other strains of *M. domestica*, both males and females are XX and have a special autosome that may carry a male-determining factor *A<sup>M</sup>* that determines sex. The *A<sup>M</sup>* component is located on different linkage groups (different chromosomes) in different populations. The presence or absence of the *M* factor seems to be the primary signal for sex determination in these strains. Interestingly, in



**Figure 10.4** Sex determination in the house fly, *Musca domestica*.  $tra^+$  may be equivalent to  $da^+$ .  $F^+$  may be equivalent to  $SxI^+$ .  $tra^+$  and  $Ag^+$  gene products are produced by the mother and stored in the egg. In the female zygote these products activate the  $F^+$  gene. The zygote's  $tra^+$  gene must be active to maintain the function of  $F^+$ . This leads to expression of female-differentiation genes, but the genes lower in the hierarchy are unidentified at present. The  $M^+$  gene product is present in males, which represses the  $F^+$  gene function, so that female-differentiation genes are repressed and unknown male-differentiation genes are activated. (Modified after Bownes 1992.)

other strains of *M. domestica*, both males and females have the  $M$  factors in the homozygous state, and the presence or absence of a female-determining dominant factor ( $F$ ) determines sex. Finally, a dominant maternal-effect mutation, *Arrhenogenous* ( $Ag$ ), has been found in *M. domestica* populations that cause female progeny to develop into fertile males. A recessive maternal-effect mutation, *transformer*, causes genotypic female progeny carrying no  $M$  factors to follow the male pathway of sexual development to varying degrees. This suggests that the normal  $tra^+$  gene product is necessary for female determination and/or differentiation and that the gene is expressed during oogenesis and in zygotes (Inoue and Hiroyoshi 1986). Experiments suggest that  $M$  acts early in embryogenesis to suppress a key gene, perhaps  $F$ , whose activity is required continuously for development of females, as is  $SxI^+$  in *Drosophila* (Hilfiker-Kleiner et al. 1993). Meise et al. (1998) found that *Sex-lethal<sup>+</sup>* is not sex-specifically regulated in *M. domestica*.

In the phorid fly *Megaselia scalaris*, the sex-determining linkage group is not fixed. Different chromosomes serve as the sex-determining pair in different populations (Traut 1994). Traut and Willhoft (1990) estimate that the male-determining factor moved to a different linkage group, thereby creating new Y chromosomes with a frequency of at least 0.06%, which is consistent with the hypothesis that the sex-determining factor is moving by transposition. An alternative explanation is that mutations at multiple sex loci in the genome result in males; however, the high rates of change (0.06%) are higher than expected if due to mutation. Analysis of the sex-determination cascade in *M. scalaris* indicates that *doublesex<sup>+</sup>* is highly conserved when compared with *dsx<sup>+</sup>* in *D. melanogaster* (Kuhn et al. 2000), but *Sex-lethal<sup>+</sup>* is not functionally conserved in *M. scalaris* (Sievert et al. 2000). Analyses of other insects also suggest that the base of the sex-determination cascade is more highly conserved in function than the upper level of the cascade (Figure 10.2).

In the blowfly *Chrysomya rufifacies* (Calliphoridae), females produce either female progeny only (**thelygenic** females) or male progeny only (**arrhenogenic** females) (Clausen and Ullerich 1990). Thelygenic females are heterozygous for a dominant female-determining maternal-effect gene (*F*), whereas arrhenogenic females and males are homozygous for the recessive allele (*f*). This species lacks differentiated sex chromosomes. DNA sequence homology between the *D. melanogaster* *da<sup>+</sup>* gene and a polytene band in the sex chromosomes of *C. rufifacies* was observed by *in situ* hybridization, suggesting that *F* in *C. rufifacies* and *da<sup>+</sup>* in *D. melanogaster* are equivalent (Clausen and Ullerich 1990). Muller-Holtkamp (1995) found that the *Sex-lethal<sup>+</sup>* gene homologue in *C. rufifacies* is highly conserved in sequence and exon–intron organization.

### 10.6.2 Environmental Effects

Environmental conditions can influence sex determination in some arthropods. Many haplo-diploid insects adjust the sex ratio of their progeny based on environmental factors. For example, females of species in the genus *Encarsia* (Hymenoptera: Aphelinidae) develop as **autoparasitoids** of whiteflies (which are considered the primary hosts). Males of the same *Encarsia* species develop as parasitoids of *Encarsia* female pupae, which are considered the secondary hosts. Virgin females deposit unfertilized eggs to produce haploid sons on secondary hosts (females of their own species), but typically do not oviposit in primary hosts (whiteflies), even if they are the only hosts available. When a virgin female does deposit haploid male eggs in a primary host (whiteflies), these eggs usually do not develop, for unknown reasons. An unusual population of *E. pergandiella* was found in which males could develop on the primary whitefly host. It appears that these haploid males started out as fertilized diploid eggs, but

become haploid males after the loss of the paternal set of chromosomes shortly after fertilization (Hunter et al. 1993).

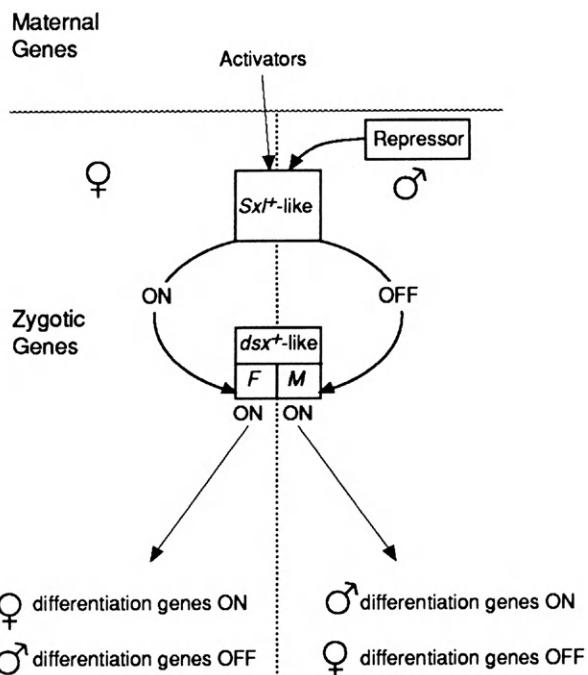
The haplo-diploid twospotted spider mite *Tetranychus urticae* (Acari: Tetranychidae) appears to modify its sex ratio based on the size of eggs in females before fertilization (Macke et al. 2011). Eggs that are larger before fertilization (which implies higher quality provisioning of eggs) produce a more female-biased sex ratio after fertilization. Unfertilized eggs produce males, which are smaller than females. Male eggs produced by mated females are smaller than male eggs produced by virgin females (who can only produce haploid male progeny). This suggests that virgin females have a male-fitness advantage over mated females, perhaps because larger males are more likely to outcompete smaller males to mate with females.

#### 10.6.3 Postzygotic Sex Determination

In several collembolans, including *Sminthurus viridis* and *Allacma fusca* (sub-order Symphleona) the two sexes differ by having 10 chromosomes in males and 12 in females (Dallai et al. 1999, 2000). Sex determination occurs after the zygote forms (rather than at syngamy). Two chromosomes are eliminated in male embryos in both the somatic and germ-line cells (Dallai et al. 2000). Oogenesis is normal but spermatogenesis is unusual; daughter cells of the first meiotic division have either six or four chromosomes. The cell receiving four chromosomes degenerates, but the cell with six completes meiosis and produces identical sperm. At fertilization, the pronuclei with six chromosomes fuse to form a zygote with 12 chromosomes. Male embryos then lose two sex chromosomes during the first mitosis, resulting in 10 chromosomes. The mechanism of chromosome elimination during early embryogenesis must be regulated by the genetic constitution of the mother, which means that females could regulate the sex ratio of their progeny. In fact, these species appear to have a female-biased sex ratio. Dallai et al. (2000) suggested that this aberrant meiosis and the large number of females in these species could be considered a step toward the evolution of parthenogenesis.

### 10.7 A Single Model?

Given the above-mentioned examples of the diverse sex-determining systems, is it likely that a single model can describe sex determination in all insects? Nothiger and Steinmann-Zwicky (1985) proposed that all the sex-determination mechanisms in insects are variations upon a theme (Figure 10.5). In their model, there is a gene equivalent to *Sxl<sup>+</sup>*, a repressor (R) that inactivates *Sxl<sup>+</sup>*, a gene that activates *Sxl<sup>+</sup>*, and a gene that is equivalent to *dsx<sup>+</sup>* that is expressed in two alternative forms to interact with one or the other of the two sets of male- and female-differentiation genes lower in the hierarchy.



**Figure 10.5** Is there a general model for sex determination in insects? This model assumes that activators are produced by the mother that activate an *Sx<sup>+</sup>*-like gene in the zygote. Other activators may be produced by the zygote, and the combination results in a functional *Sx<sup>+</sup>*-like product in females. Male zygotes produce a repressor of the *Sx<sup>+</sup>*-like gene and no functional product is produced in males. Next, a *dsx<sup>+</sup>*-like gene is turned on in both females and males but different products are produced in the two sexes. The different *dsx<sup>+</sup>* gene products turn on a subsequent series of genes that result in the differentiation of either males or females. Current data suggest the lower level of the cascade is more conserved. (Redrawn from Bownes 1992.)

Recent studies indicate that this model is partly correct, especially at the lower levels of the hierarchy. *Sx<sup>+</sup>* is apparently not as conserved as genes at the lower level of the hierarchy: *tra<sup>+</sup>*, *tra<sup>-2+</sup>*, *dsx<sup>+</sup>*, and *intersex<sup>+</sup>* are more conserved among the insects studied so far (Siegal and Baker 2005). Wilkins (1995) proposed the hypothesis that sex-determination pathways have evolved from the bottom up: the later a gene acts in the pathway, the more likely it will be conserved. Thus, a new genetic function inserted into the top of the pathway would be initiated by natural selection to maintain an optimal sex ratio. If the sex ratio deviates too much from a 1:1 ratio, there will be selective pressure to shift the ratio toward the less-frequent sex and this new gene at the top of the pathway would then affect all downstream genes. The *doublesex<sup>+</sup>* gene is functionally conserved in Diptera and Lepidoptera, and *tra<sup>+</sup>* appears to be at the top of the sex-determination hierarchy in the Mediterranean fruit fly and the honey bee (Siegal and Baker 2005).

Bownes (1992) used their model to compare the sex-determination system in the house fly (Figure 10.4). According to the Nothiger and Steinmann-Zwicky (1985) model, the male-determining factor (*M*) in *M. domestica* would correspond to the repressor (*R*) of *Sxl<sup>+</sup>*. The genes *tra<sup>+</sup>* and *Ag<sup>+</sup>* may be equivalent to *da<sup>+</sup>* in *Drosophila*. The *F* gene of *M. domestica* could be equivalent to *Sxl<sup>+</sup>*. *Sxl<sup>+</sup>* is involved in dosage compensation in *Drosophila*, but dosage compensation is not needed in species such as *M. domestica* with heterochromatic sex chromosomes (which usually contain few coding regions) or no sex chromosomes. As a result, insects with heterochromatic sex chromosomes or no sex chromosomes can survive mutations of *tra<sup>+</sup>*, *Ag<sup>+</sup>*, and *F*; such mutations can alter sex determination, but are not lethal to one sex.

Different genera of mosquitoes have several different sex-determination systems, but these systems still may conform to the Nothiger and Steinmann-Zwicky model (Bownes 1992). *Anopheles gambiae* and *Anopheles culicifacies* have XY males and XX females. Sex in *Aedes* is determined by a dominant male-determining factor. Intersex flies with phenotypes similar to the *ix*, *dsx*, and *tra* mutants of *Drosophila* have been found in *Aedes aegypti* and *Culex pipiens*. A single gene located on an autosome determines sex in *Culex*; *Culex* gynandromorphs have been found, suggesting that sex determination is cell autonomous, as it is in *Drosophila*. The sex of some northern strains of *Aedes* depends upon the temperature at which they are reared, with males transformed into intersexes at higher temperatures. This suggests that an allele equivalent to *ix<sup>+</sup>* is temperature sensitive in these populations. In *Cx. pipiens*, a sex-linked gene *cercus* (*c*) changes females into intersexes; these intersexes are sterile and fail to take blood meals. Bownes (1992) speculated that it is possible that *cercus<sup>+</sup>* is similar to *tra<sup>+</sup>*, *ix<sup>+</sup>*, or *dsx<sup>+</sup>* of *Drosophila*.

Sexual differentiation among different organisms (flies, nematodes, and mammals) has superficially similar patterns of hierarchical control (for reviews, see Marin and Baker 1998, McAllister and McVean 2000). Comparative genetic analysis suggests that the functions of *tra<sup>+</sup>*, *tra-2<sup>+</sup>*, and *dsx<sup>+</sup>* may be conserved throughout higher eukaryotes (Verhulst et al. 2010a). Sex-determining mechanisms are, however, variable and the function of *Sxl<sup>+</sup>* is not conserved among all arthropods. Verhulst et al. (2010a) evaluated what is known about insect sex determination and conclude

*"The primary signal that starts sex determination is processed by a cascade of genes ending with the conserved switch doublesex that controls sexual differentiation. Transformer is the doublesex splicing regulator and has been found in all examined insects, indicating its ancestral function as a sex-determining gene."*

Furthermore, sex determination in insects appears to be based on alternative splicing of *tra<sup>+</sup>* mRNAs, based on studies in Diptera and the Hymenoptera (Salz 2011).

## 10.8 Meiotic Drive Can Distort Sex Ratios

**Meiotic drive** alters the assortment of chromosomes during meiosis so that certain chromosomes are inherited more frequently than expected (>50%). Meiotic drive most frequently is observed when sex chromosome allocation is disrupted (sex-chromosome meiotic drive) so that the sex ratio is altered. Whether meiotic-drive mechanisms actually modify sex-chromosome distribution more frequently than autosomal chromosomes is unknown (Lyttle 1993).

Sex-chromosome meiotic drive has been found most often in the Diptera, including the *Drosophila obscura*, *melanica*, *tripunctata*, *testacea*, *melanogaster*, and *quinaria* groups, mosquitoes (*Aedes* and *Culex*), sciarid flies, and stalk-eyed flies (Diopsidae) (Jiggins et al. 1999). The frequency of meiotic drive in other insects is unclear (Jiggins et al. 1999, Jaenike 2001). Three examples of sex-chromosome meiotic drive in Diptera are described in Sections 10.8.1–10.8.3, including Segregation Distorter (SD) in *Drosophila*, male drive (MD) in the mosquitoes *Aedes aegypti* and *Culex quinquefasciatus*, and meiotic drive in stalk-eyed flies.

### 10.8.1 Segregation Distorter (SD)

In *Drosophila melanogaster*, the SD phenotype is present at low, but stable, frequencies in most field populations. Males of *D. melanogaster* that are heterozygous for the (SD) chromosome (SD/SD<sup>+</sup>) may produce only progeny with the SD chromosome, instead of half with SD and half with SD<sup>+</sup>, due to the failure of sperm with the SD<sup>+</sup> chromosome to mature (Ashburner 1989, Ganetzky 2000). Segregation distortion occurs because the nuclei of the sperm with the normal SD<sup>+</sup> chromosome fail to condense normally at sperm maturation. Another gene, the Enhancer locus of SD, *E(SD)*, is required for the full expression of meiotic drive.

The SD “locus” actually consists of two overlapping genes, HS2ST and RanGAP. RanGAP is an essential component of a system that transports proteins and RNA molecules into and out of the cell’s nucleus. Both HS2ST and RanGAP actually are present twice on the SD chromosome, as opposed to once on SD<sup>+</sup> chromosomes, and the tandem duplication is necessary for segregation distortion (Palopoli et al. 1994). Both genes appear normal in the right-hand copy, but the RanGAP gene on the left lacks the last 234 amino acids (Merrill et al. 1999, Ganetzky 2000).

### 10.8.2 Distorter in Mosquitoes

Meiotic drive has been described in the mosquitoes *Aedes aegypti* and *Culex quinquefasciatus*. In both species, a Y-linked gene results in excess males (Wood and Newton 1991). Excess males are produced because X chromosomes are broken during meiosis in males and thus fewer X than Y chromosomes are transmitted in the sperm, leading to the production of fewer female embryos. The *Distorter* gene (*D*) is linked closely to the sex locus *m/M* and causes chromosome breakage. Additional genes are involved, and sensitivity to *Distorter* is controlled by *m*, the female-determining locus. In some strains, sensitivity is influenced by a second sex-linked gene *t*. Yet another sex-linked gene *A* enhances the effect of *Distorter*. *Distorter* has been found in mosquito populations from Africa, America, Australia, and Sri Lanka, but resistance to it is widespread.

### 10.8.3 Female-Biased Sex Ratios in Stalk-Eyed Flies

Extreme female-biased sex ratios are found in two stalk-eyed fly species, *Cyrtodiopsis dalmanni* and *Cyrtodiopsis whitei* (Diopsidae), due to a meiotic-drive element on the X chromosome (Presgraves et al. 1997). Eyestalks are more exaggerated in males than in females (Wilkinson et al. 1998) and females prefer to mate with males with a long eye span. The long stalks appear to indicate to the female that the male either lacks meiotic-drive elements or can suppress meiotic drive, thereby increasing the female's fitness by avoiding a biased sex ratio in her progeny (Wilkinson and Reillo 1994). Apparently, there are both autosomal- and Y-linked polymorphisms for resistance to meiotic drive.

### 10.8.4 Meiotic Drive as a Pest-Management Tool?

Meiotic drive operates as an evolutionary force that can cause an increase in the population frequency of the allele or chromosome that is favored in transmission, even if it confers a disadvantage on its carriers. It has been proposed that meiotic drive might be used to introduce new genes (such as cold-sensitive lethal genes, insecticide-susceptibility genes, or behavior-altering genes that would reduce the negative effects of mosquitoes on humans) into natural populations as a method to achieve control of these important vectors of disease (Wood and Newton 1991). However, much remains to be learned about the stability of meiotic-drive mechanisms and the conditions under which they might function in pest-management programs. The fact that resistances can develop to suppress meiotic drive could limit the effectiveness of meiotic drive in pest-management programs.

## 10.9 Hybrid Sterility

When different species are crossed, hybrid progeny sometimes are produced. However, the progeny may have altered sex ratios, with one sex absent, rare, or sterile. The missing or sterile sex is usually the heterogametic sex (Laurie 1997). This phenomenon is known as **Haldane's Rule**, which is "one of a few general rules in evolutionary biology" (Sawamura 1996). However, Sawamura (1996) indicates some exceptions do exist to Haldane's rule, caused by maternal effects. Hurst and Pomiankowski (1991) suggest that Haldane's Rule only occurs in taxa with sex chromosome-based meiotic drive, such as the Lepidoptera and Diptera. Thus, Haldane's rule may be accounted for in some insects by a loss of suppression of sex-ratio distorters when in the novel nuclear cytotype of the hybrid.

Sex-ratio distorters that result in unisexual sterility in crosses between different species have been found in many species of *Drosophila*, the dipterans *Musca domestica* and *Glossina morsitans*, the hemipteran *Tetraneura ulmi*, and Lepidoptera (*Acraea encedon*, *Maniola jurtina*, *Danaus chrysippus*, *Philodoria potatoria*, *Mylothris spica*, *Abraxas grossulariata*, and *Talaeporia tubulosa*) (Hurst and Pomiankowski 1991).

## 10.10 Medea in *Tribolium*

A class of selfish genes, *Medea*, was found in the red flour beetle, *Tribolium castaneum* (Beeman et al. 1992). *Medea* causes a "Maternal-Effect Dominant Embryonic Arrest" that results in the death of zygotes that do not carry it. If a mother carries *Medea*, any of her offspring that lack this gene die before they pupate. Females who are heterozygous for *Medea* lose half their progeny if they mate with a wild-type male and  $\frac{1}{4}$  of their progeny when mated to a heterozygous male. It was hypothesized that *Medea* could lead to reproductive isolation and speciation in *T. castaneum* populations.

A survey of wild populations of *T. castaneum* from Europe, North and South America, Africa, and Southeast Asia showed that four different *Medea* alleles were widespread, but absent or rare in Australia and the Indian subcontinent (Beeman and Friesen 1999). Thomson and Beeman (1999) suggest that *Medea* factors are absent from India because a hybrid incompatibility factor (H) is found in the *T. castaneum* populations in India. Apparently, H and *Medea* strains of *T. castaneum* are incompatible due to suppression by the H factor of the self-rescuing activity of the lethal *Medea* genes.

## 10.11 Cytoplasmic Agents Distort Normal Sex Ratios

Many cytoplasmically transmitted organisms (bacteria, viruses, and protozoa) alter the "normal" sex-determining mechanism(s) in arthropods (Table 10.3). Most are inherited primarily through the oocyte of the mother (cytoplasmically

**Table 10.3: Examples of Nonnuclear Influences on Sex Determination or Sex Ratio in Arthropods.<sup>a</sup>**

Microorganism	Arthropod species (order)	Effect reference(s)
<i>Arsenophonus nasoniae</i> gram-negative bacterium	<i>Nasonia vitripennis</i> (Hymenoptera)	Kills male eggs Ghera et al. 1991
<b>Bacterial male-killing in Coleoptera</b>		
Several, including <i>Rickettsia</i> <i>Spiroplasma</i> <i>Wolbachia</i> (two types)	<i>Adalia bipunctata</i> (Coleoptera)	Males killed Werren et al. 1994
Flavobacterium	<i>Coleomegilla maculata</i>	Hurst et al. 1999a
<i>Rickettsia</i>	<i>Adalia decempunctata</i>	Hurst et al. 1999b
<i>Spiroplasma</i>	<i>Harmonia axyridis</i>	Hurst et al. 1997
<i>Wolbachia</i>	<i>Tribolium madens</i>	von der Schulenburg et al. 2001
<i>Rickettsia</i>	<i>Brachys tessellates</i>	Majerus et al. 1999
Maternal sex ratio (MSR)	<i>Nasonia vitripennis</i> (Hymenoptera)	Stevens 1993, Fialho and Stevens 2000
Paternally transmitted sex ratio factor (PSR), a supernumerary B chromosome that is mostly heterochromatic	<i>Nasonia vitripennis</i> (Hymenoptera)	Lawson et al. 2001 Results in nearly all female progeny Beukeboom and Werren 1992 Fertilized eggs lose paternal chromosomes; females converted to males; transmitted only via sperm
Sex-ratio condition viral?	<i>Drosophila bifasciata</i> (Diptera)	Werren et al. 1987, Nur et al. 1988, Beukeboom and Werren 1993
Sex-ratio condition	<i>D. willistoni</i> and related neotropical species, <i>D. melanogaster</i> (Diptera)	Death of male embryos Leventhal 1968
Spiroplasmas	<i>Oncopeltus fasciatus</i> (Hemiptera)	Males die as embryos
Sex-ratio condition	<i>Drosophila paulistorum</i> (Diptera)	Ebbert 1991, Martins et al. 2010
Sterility in male progeny streptococcal L-form bacteria	<i>Encarsia formosa</i>	Leslie 1984 Induction of semispecies? Somerson et al. 1984
Thelytoky		Males produced after antibiotic treatment; microorganisms restore diploidy to unfertilized eggs, resulting in all-female progeny (thelytoky).
<i>Wolbachia</i>	<i>Ostrinia scapulalis</i> (Hymenoptera)	Zchori-Fein et al. 1992 All-female progeny produced due to death of males; cured females produce all-male progeny indicating <i>Wolbachia</i> is necessary for female development.
<i>Cardinium</i>	<i>Encarsia pergandiella</i> (Lepidoptera)	Kageyama and Traut 2004 Bacterium causes thelytoky and alters host-selection behavior
<i>Rickettsia</i>	<i>Neochrysocharis formosa</i>	Zchori-Fein et al. 2001 Female diploidy maintained by meiotic cells undergoing a single equational division

(Continued)

Table 10.3: (Continued)

Microorganism	Arthropod species (order)	Effect reference(s)
Thelytoky lost maternally inherited	(Hymenoptera) <i>Trichogramma</i> species (Hymenoptera)	Cure with antibiotics results in bisexual (arrhenotokous) populations. Stouthamer et al. 1990, Stouthamer and Werren 1993, Stouthamer and Luck 1993
<i>Wolbachia</i>	<i>Leptopilina heterotoma</i>	Tri-infected strain males crossed with uninfected or monoinfected females result in killed female eggs; monoinfected males crossed with uninfected females result in reduced progeny and more males
	(Hymenoptera)	Vavre et al. 2000

<sup>a</sup>*Wolbachia* also causes cytoplasmic incompatibility and other effects (see Chapter 4).

inherited). Cytoplasmic agents that can manipulate their host's sex ratio and promote their own spread are called **cytoplasmic sex-ratio distorters**. The spread of a cytoplasmic sex-ratio distorter often reduces the fitness of its host and can drive populations to extinction (Cordaux et al. 2011). Sex-ratio distorters are usually suspected if crosses produce a heavily female-biased sex ratio, although meiotic drive and hybrid dysgenesis agents are other possible mechanisms (Hurst 1993). Cordaux et al. (2011) evaluated the impact of symbionts on the evolution of sex-determination mechanisms in their hosts and noted that there are four types of reproductive mechanisms caused by symbionts: cytoplasmic incompatibility (as discussed in Chapter 4), male killing (increase in the proportion of females through targeted death of male progeny), feminization of genetic males (conversion of genetic males to functional females), and induction of asexual daughter development (thelytoky).

Ebbert (1991, 1993) described >50 cases in which cytoplasmically inherited organisms alter sex ratios in the Diptera, Heteroptera, Coleoptera, Lepidoptera, and Acari (mites). Such sex-ratio distorters may be widespread, but undiscovered, in other arthropods because scientists rarely assess sex ratios by making single-pair crosses. Transmission rates of these agents typically are high, although a few progeny may fail to become infected. The altered sex-ratio conditions are found in natural populations at frequencies ranging from low to high. The infections may reduce fitness of the hosts, and reduce egg hatch or larval survival in the progeny of infected females. Some examples are described in Sections 10.11.1–10.11.5.

### 10.11.1 Spiroplasma Strains

The sex-ratio condition of *Drosophila willistoni*, and at least 10 other *Drosophila* species, is due to a *Spiroplasma* strain (Ashburner 1989, Williamson et al. 1999, Anbutsu and Fukatsu 2003). *Spiroplasma* strains are maternally inherited, transovarially transmitted, and lethal to male embryos. Male eggs die at early developmental stages before gastrulation. Anbutsu and Fukatsu (2003) compared the titers of two strains of *Spiroplasma* by a quantitative PCR method; one strain killed males and one did not but was otherwise very similar. They compared the ability of the *Spiroplasma* strains to proliferate within their hosts to determine whether their titer altered the ability of the benign strain and discovered that, indeed, the benign strain did not replicate as well. Thus, the *titer* of the strain was important in resulting in male killing.

*Spiroplasma* strains can be transmitted between species by injecting hemolymph, but *Spiroplasma* strains from different species are different, and a different virus is associated with each. When *Spiroplasma* strains from different species are mixed, they clump because the viruses lyse the *Spiroplasma* of the other species.

Why would male-killing bacteria persist in insect populations? Martins et al. (2010) tested the hypothesis that death of males results in increased resources for sibling females. This suggests that infected females should be larger than uninfected females, or have higher viability, or shorter development times. Tests involving infected and uninfected *D. melanogaster* females collected from the wild indicated that infected females produced more daughters than uninfected females, and there was a decrease in development time of infected females, perhaps as a result of reduced competition with sibling males (which died).

### 10.11.2 L-Form Bacteria

The *Drosophila paulistorum* complex contains six semispecies (subgroups derived from a single species that are thought to be in the process of speciation) that do not normally interbreed. When they are crossed in the laboratory, fertile daughters and sterile sons are produced. Streptococcal L-form bacteria were isolated and cultured in artificial media that are associated with the sterility (Somerson et al. 1984). The L-forms are transferred through the egg and each semispecies appears to have a different microorganism. L-forms can be microinjected into females and can produce the expected male sterility. This suggests that an L-form normally has a benign relationship with its own host; however, if it is transferred to a closely related host, male sterility is induced.

### 10.11.3 Rickettsia

*Neochrysocharis formosa* (Hymenoptera: Eulophidae) is an important natural enemy of leafminers. It has both thelytokous and arrhenotokous strains, with thelytoky induced by *Rickettsia*. The mechanism of producing diploid females (thelytoky) was found to be due to the fact that meiotic cells during female-gamete formation undergo only a single equational division, followed by the expulsion of a single polar body. This means that meiotic recombination does not occur and diploidy is maintained by an apomictic cloning mechanism that differs from the mechanism of thelytoky-induction by *Wolbachia* (Adachi-Hagimori et al. 2008).

### 10.11.4 Wolbachia

These rickettsia-like bacteria are one of the most commonly described cytoplasmically inherited microorganisms in arthropods. *Wolbachia* are discussed in Chapter 4 with regard to their ability to induce cytoplasmic incompatibility when populations with and without the same *Wolbachia* strain mate. *Wolbachia* are gram-negative rods that cannot be cultured easily outside their hosts and they are widespread, with 17–76% of all arthropod species containing them (Werren et al. 1995, Werren 1997, Jeyaprakash and Hoy 2000). Knowledge of the evolution and physiological and phenotypic effects of *Wolbachia* on most of their hosts remains limited; only a few of the thousands of insects that contain *Wolbachia* have been studied to determine the phenotype (if any) produced by the symbiont (Rigaud 1999). One of their effects in arthropods is to alter sex ratio (Rigaud and Rousset 1996, O'Neill et al. 1997, Bourtzis and O'Neill 1998, Cook and Butcher 1999, Stouthamer et al. 1999, Stevens et al. 2001, Vavre et al. 2001).

In addition to infecting insects, *Wolbachia* infect isopods (Crustacea), including *Armadillidium album*, *Ligia oceanica*, *Armadillidium nasatum*, *Porcellionides pruinosus*, *Chaetophiloscia elongata*, and *Spaeroma rugicauda* (Rigaud and Rousset 1996). Some *Wolbachia*-infected isopods produce female-biased broods because the *Wolbachia* change genetic males (homogametic ZZ individuals) into functional “females.” These ZZ individuals are chromosomally male, but phenotypically appear and function as females. “Daughters” of infected mothers produce all-female or highly female-biased progeny, resulting in isopod lineages that are chromosomally males (ZZ) but are functional females (Rigaud and Rousset 1996). There is speculation that some *Wolbachia* genes have been transferred to the isopod nuclear genome. If the “females” are cured of their *Wolbachia* with antibiotics, their progeny are all males (ZZ).

*Wolbachia* can cause thelytoky, male killing, and female mortality (Rousset and Raymond 1991, Stouthamer and Werren 1993, Bandi et al. 1999, Majerus

et al. 1999, Vavre et al. 2001). Some *Wolbachia* improve fertility or vigor, whereas others appear to decrease these traits in their hosts. Some species appear to have *Wolbachia* only in their germ line (ovaries and testes), whereas others have *Wolbachia* in somatic tissues, as well (Dobson et al. 1999). Large numbers of *Wolbachia* have been found in ovaries and testes of populations with cytoplasmic incompatibilities. Sometimes, infection with *Wolbachia* can increase fecundity in females, thus favoring its transmission. For example, in *Drosophila mauritiana* infected with a native strain of *Wolbachia*, females produce four times as many eggs compared to uninfected strains. This is apparently due to increased activity of stem cells in the germ-line stem-cell niche (Fast et al. 2011).

*Wolbachia* may cause thelytoky in the Hymenoptera, which typically are arrhenotokous (Stouthamer 1997). *Wolbachia*-induced thelytoky (parthenogenesis in which only females are known) has been found in the Tenthredinoidea, Signiforidae, and Cynipoidea (Stouthamer 1997), as well as at least 70 species of parasitoids (Aphelinidae, Encyrtidae, Eulophidae, Pteromalidae, Torymidae, Trichogrammatidae, Cynipidae, Eucoilidae, Braconidae, Ichneumonidae, and Proctotrupoidae) (Stouthamer 1997, Cook and Butcher 1999, Russell and Stouthamer 2011). Many hymenopteran parasitoid species have both arrhenotokous and thelytokous strains.

In thelytokous populations of parasitoids, unfertilized eggs give rise to females. Several thelytokous parasitoids (*Ooencyrtus submetallicus*, *Pauridia peregrina*, *Trichogramma* sp., and *Ooencyrtus fecundus*) produce a few males, usually <5%, when reared at temperatures over 30 °C (Stouthamer 1997), suggesting incomplete transmission of *Wolbachia* or a low titer of *Wolbachia*. Theytokous populations of the braconid *Asobara japonica* produce small numbers of males, which Reumer et al. (2011) attributed to the recent invasion of these parasitoid populations, resulting in “a maladaptive side effect of incomplete coevolution between symbiont and host in this relatively young infection.”

Sometimes, the rare males produced in thelytokous populations have been shown to mate and transfer sperm to conspecific females, indicating that the male retained normal vigor and fertility. In other cases, the rare males are infertile, suggesting that the *Wolbachia* infection has existed for a long time in the population, which could have relaxed selection for essential fertility genes over evolutionary time (Russell and Stouthamer 2011). In addition to heat, several antibiotics (tetracycline hydrochloride, sulfamethoxazole, and rifampin) can induce the production of males in some thelytokous parasitoid populations infected with *Wolbachia*.

The cytogenetic changes that occur during meiosis to restore an unfertilized haploid egg to diploidy (thus permitting thelytoky) has been studied in several

arthropods, and the mechanisms vary, indicating that *Wolbachia* can cause parthenogenesis by several mechanisms. In the eggs of *Wolbachia*-infected *Trichogramma* females, meiosis progresses to the stage of a single haploid pronucleus and the diploid chromosome number is restored during the first mitotic division. Thus, during anaphase, the two identical sets of chromosomes do not separate and the result is a single nucleus containing two copies of the same set of chromosomes, resulting in a female that is completely homozygous at all loci (Stouthamer 1997). A *Wolbachia*-infected strain of the parasitoid *Leptopilina clavipes* has a meiosis in which diploidy is restored by anaphase restitution during the first somatic mitosis, similar to that in *Trichogramma* (Pannebakker et al. 2004). However, in *Diplolepis rosae* and *Muscidifurax uniraptor* parasitoids the first mitotic division is normal and diploidization occurs through a fusion of the two mitotic nuclei in the second prophase (Pannebakker et al. 2004).

How might parasitoids (or other insects) be infected with *Wolbachia*? Phylogenetic analysis of *Wolbachia* suggests that both horizontal and vertical transfer of *Wolbachia* occurs among insects (Jeyaprakash and Hoy 2000). Huigens et al. (2000) documented horizontal transmission of thelytoky-inducing *Wolbachia* from one parasitoid strain to another within a shared lepidopteran host. Offspring of uninfected *Trichogramma* females can acquire sufficient thelytoky-inducing *Wolbachia* to express the trait when they share a host egg with progeny of *Wolbachia*-infected females. The process by which the uninfected *Trichogramma* larvae acquire the *Wolbachia* remains unclear. However, this intraspecific horizontal transfer suggests that interspecific horizontal transfers from parasitoid to parasitoid could occur by sharing a common host.

*Wolbachia* infections causing thelytoky are hypothesized to be a mechanism that contributes to the process of speciation. For example, some populations of the parasitoid *Encarsia formosa* no longer have males, so that these populations essentially become clonal and over time could differentiate genetically. *Wolbachia*-induced incompatibility is thought to precede hybrid incompatibilities in the parasitoid *Nasonia* (Bordenstein et al. 2001). Under this scenario, an uninfected ancestral population gives rise to two geographically isolated daughter populations. If each population is infected with a different strain of *Wolbachia*, the populations could become reproductively isolated due to their infections. The role of *Wolbachia* in speciation is controversial and, according to some, unproven (Werren 1997, Hurst and Schilthuizen 1998, Shoemaker et al. 1999, Rokas 2000).

*Wolbachia* can influence mating behavior and kill males in populations of the butterfly *Acraea encedon* across Africa. In many populations, females produce only female progeny, whereas other populations produce both males and females in a normal 1:1 sex ratio. *Wolbachia* kills males, resulting in the production of all-female progeny in this butterfly (Jiggins et al. 1998). *Acraea encedon* females typically deposit clutches of 50–300 eggs and newly hatched larvae often cannibalize unhatched eggs, only gradually dispersing into smaller groups. Jiggins et al. (1998) speculated that the evolution of male-killing behavior by *Wolbachia* may be favored when the behavior and ecology of a species makes antagonistic interactions between siblings or sib cannibalism likely. Under field conditions, *Wolbachia* infections in *A. encedon* females may result in populations with a serious shortage of males. As a consequence, the mating behavior of *Wolbachia*-infected *A. encedon* has been altered. Normally males seek out and compete for individual females near larval food plants. However, when male-killing *Wolbachia* are present in high frequency in a population, females instead form dense aggregations in grassy areas near trees, perhaps to attract rare males as mates.

In adzuki bean borer *Ostrinia scapulalis* (Lepidoptera: Crambidae) populations infected with *Wolbachia*, sexual mosaics are produced (Sugimoto 2010). The female-specific form of the *doublesex<sup>+</sup>* gene is transcribed as well as the male-specific form, indicating that *Wolbachia* manipulates the sex of this insect by interfering either with the sex-specific splicing of *dsx<sup>+</sup>* itself or with another upstream sex-determination process.

*Wolbachia* have been proposed as “drive” mechanisms to introduce transgenes into arthropod populations (Turelli and Hoffman 1999), as is discussed further in Chapter 14. *Wolbachia* infections may reduce the ability of mosquitoes to transmit pathogens (viruses, nematodes, and protozoa); this also is being evaluated as a pest-management method and is described in Chapter 14. However, Cordaux et al. 2011) noted that resistance to feminization by *Wolbachia* and resistance to male killing by *Wolbachia* has developed in wild populations. This suggests that there could be an arms race between symbionts and their hosts.

#### 10.11.5 Cardinium

The false spider mite *Brevipalpus phoenicis* (Acari: Tenuipalpidae) consists entirely of females that have only a haploid chromosome set (Weeks et al. 2001). This unusual genetic system is due to the presence of an endosymbiotic bacterium, *Cardinium*, discovered relatively recently by Zchori-Fein et al. (2001). The *Cardinium* in *Brevipalpus* results in feminized haploid males, but how the bacterium induces feminization of genetic males is unknown.

## 10.12 Paternal Sex-Ratio Chromosomes and Cytoplasmic Incompatibility in *Nasonia*

Sex ratio in the parasitoid *Nasonia vitripennis* can be altered by at least two different mechanisms. Some natural populations of *N. vitripennis* carry a supernumerary or **B chromosome** that causes a condition called **paternal sex ratio (PSR)**. B chromosomes are found in many plant and animal species and are small non-vital chromosomes mostly consisting of heterochromatin. B chromosomes have few genes and often cause a small fitness cost to their host, making them “selfish” genetic elements. Some B chromosomes are thought to be derived from normal chromosomes and may be transmitted at higher rates than expected, exhibiting “drive” (Jones and Rees 1982).

The PSR chromosome is carried only by male *N. vitripennis* and is transmitted via sperm to fertilized eggs. After an egg is fertilized by a PSR-bearing sperm, the paternally derived chromosomes condense into a chromatin mass and are lost, leaving only the maternal chromosomes. The PSR chromosome itself persists, changing fertilized diploid (female) eggs into haploid PSR males. PSR is unusual in its ability to destroy the genome of its carrier each generation (Werren et al. 1987; Nur et al. 1988; Beukeboom and Werren 1992, 1993; Beukeboom et al. 1992; Reed and Werren 1995).

Where did the PSR chromosome come from? The PSR chromosome has sequences that are homologous with autosomal sequences of *Nasonia giraulti*, *N. longicornis*, and *Trichomalopsis dubius*, but not with *N. vitripennis* (Eickbush et al. 1992). The PSR chromosome could have been present before the divergence of the genera *Trichomalopsis* and *Nasonia* (Eickbush et al. 1992). Alternatively, PSR may have crossed the species barrier more recently (horizontal transfer) through a series of interspecific transfers between species capable of mating (Dobson and Tanouye 1998b). Experimental interspecific transfer of the PSR was successful after these species were cured of *Wolbachia*, which causes cytoplasmic incompatibility between them (Dobson and Tanouye 1998b). The transferred PSR chromosome functioned in both recipient species.

The sex ratio of *N. vitripennis* is modified by other non-Mendelian factors including “Son-killer,” a maternally transmitted bacterium that prevents development of unfertilized male eggs (Huger et al. 1985) and “Maternal Sex Ratio,” a cytoplasmically inherited agent that causes female wasps to produce nearly 100% daughters.

## 10.13 Male Killing in the Coccinellidae

Male killing is associated with a variety of microorganisms. The Coccinellidae appear particularly prone to infection by male-killing endosymbionts, with four

different groups (*Rickettsia*, *Spiroplasma*, Flavobacteria, and *Wolbachia*) identified (Majerus and Hurst 1997, Hurst et al. 1999a,b, Majerus et al. 1999, Hurst and Jiggins 2000, Sokolova et al. 2002). Coccinellids may be especially susceptible to invasion by and establishment of male-killing microbes due to their biology. Coccinellids feed on aphids and lay eggs in tight batches, which promotes sibling-egg cannibalism and significant levels of mortality of newly hatched larvae due to starvation (Majerus et al. 1999).

The evolution of male killing may have evolved because the bacteria are almost exclusively transmitted vertically from mother to eggs. As a result, bacteria in male hosts are at an evolutionary dead end, so male killing has a fitness cost of zero from the bacterial point of view (Randerson et al. 2000). Furthermore, the death of male embryos could augment the fitness of the remaining female brood by providing food to those females carrying the clonal relatives of the male-killing bacteria (Randerson et al. 2000). However, Majerus and Majerus (2010) found that resistance to a male-killer can develop, suggesting that there is an “arms race” between host and microbial symbiont.

## 10.14 Sex and the Sorted Insects

Resolving the molecular genetics of sex determination in arthropods and learning how to modify sex ratio or fertility will have both theoretical and applied applications, and could lead to improved genetic control of pests or useful genetic modifications of beneficial biological control agents.

### 10.14.1 Genetic Control

Genetic control of pest insects represents an attractive alternative to chemical control in terms of safety, specificity, and the limited negative impact it has upon the environment. The screwworm (*Cochliomyia hominivorax*) eradication campaign demonstrates what can be achieved with mass releases of males sterilized by irradiation (Box 10.1). The principle of sterile-insect releases has been applied to other pests, including the Mediterranean fruit fly (*Ceratitis capitata*), tsetse flies (*Glossina palpalis* and *G. morsitans*), mosquitoes (*Anopheles albimanus*), codling moth (*Cydia pomonella*), and ticks (LaChance 1979).

Sterile-insect release programs usually require only males, but both sexes typically must be reared. Not only is it expensive to rear large numbers of “useless” females, but, in the case of species that vector disease or annoy or bite humans or domestic animals, it is undesirable to release any females, sterile or not! As a result, genetic methods have been used to develop “genetic sexing strains,” strains that make it easy to separate males and females. For example, slight differences in size or color of pupae have been used to sort out undesirable

### Box 10.1 Eradication of the “Flesh-Eating Fly”

The genetic-control method used to eradicate the screwworm *Cochliomyia hominivorax* was called the sterile-insect release method (SIRM) or sterile-insect technique (SIT). The SIRM involves mass rearing, sterilization of males by chemicals or irradiation, and their subsequent release to mate with wild females. Because females of the screwworm mate only once, any wild female mating with a sterile male fails to contribute progeny to the next generation (Knipling 1955, 1985). By releasing an excess of sterile males (compared with the number of wild males), populations decline in a predictable manner, ultimately becoming extinct. Because absolute population densities of *C. hominivorex* were typically low during the winter in the USA, the number of sterile males that had to be released could be produced in fly factories.

The screwworm eradication program was initiated in Florida with small-scale trials on Sanibel Island in 1951. The results were promising and the project was geared up to cover the state of Florida and then the southeastern United States. The screwworm was declared eradicated from the southeastern United States in 1959, one year ahead of schedule. Eradication was achieved in a surprisingly short time due to the combined effects of a severe winter in Florida during 1957–1958, which greatly reduced the overwintering screwworm population, and a 17-month eradication program beginning in July 1958 that cost ≈US\$7 million and involved the release of almost 9 billion sterile screwworm flies over an area of ≈56,000 square miles (Meadows 1985).

Since 1959, the livestock industry of Florida and adjacent states have saved at least US\$20 million each year because the screwworm is no longer present; actual benefits are even greater in today’s dollars (Meadows 1985). Furthermore, the elimination of losses due to the deaths of livestock and the elimination of labor and control costs are only part of the benefits; loss of wildlife to screwworm attack also was eliminated.

The success of the SIRM program in the southeastern United States led the cattle growers of Texas to mount, in collaboration with the state and the U.S. Department of Agriculture, a similar but more ambitious program in the southwestern United States in the 1960s (Bushland 1985). This program required more time and effort because the area from which the screwworm was to be eradicated bordered on a front 2400 km long, stretching from the Gulf of Mexico to the Pacific Ocean. Despite this challenge, and some setbacks with quality control and reinvasion of flies from Mexico, both Texas and New Mexico were declared “screwworm free” in 1964.

The SIRM program was moved into Arizona and California in 1965, and by 1966 the entire United States could be declared free of screwworms. To reduce the likelihood that the screwworm would reinvoke the United States from Mexico, the program was expanded into Mexico in 1972, with the goal of eradicating the screwworm all the way south to the Isthmus of Tehuantepec (Pineda-Vargas 1985).

After successfully eliminating the pest in Mexico, the SIRM program was expanded to cover all of Central America (Wyss 2000). Screwworms were eliminated from Guatemala between 1988 and 1994, from Belize between 1988 and 1994, from El Salvador between 1991 and 1995, from Honduras between 1991 and 1996, from Nicaragua between 1992 and 1998, from Costa Rica between 1995 and 1999, and from Panama between 1997 and 2000. These eradication programs were carried out so that a barrier zone could be set up at the 90 km wide Isthmus of Panama, which is only easier to maintain compared to the 2400-km border that the United States and Mexico share (Snow et al. 1985). This barrier zone is being maintained by a combination of quarantines and mass releases of sterile screwworms.

Benefits of this massive, and expensive, screwworm eradication program are large (Wyss 2000). In 1996, the producer benefits in the United States, Mexico, and Central America were estimated to be US\$796 million, US\$292 million, and US\$77.9 million annually, respectively. These benefits were due to decreases in deaths of livestock, reduced veterinary services, medicines, insecticides, inspections, handling costs, and increases in meat and milk production (Richardson and Averhoff 1978). The estimated benefit to cost ratios for the eradication programs average 12.2:1 for Central America to 18:1 for the United States and Mexico (Wyss 2000).

females during mass rearing. Most genetic-sexing strains are based on maintaining marker genes (such as *white pupa* or a temperature-sensitive lethal) within chromosomal translocations. However, because translocations can undergo recombination in the region between the translocation breakpoint and the marker gene, the strains are not completely stable. As a result, if no practical means exist to remove the recombinants, an increasing number of "undesirable" females will be reared and released.

Ideally, a genetic-sexing method would produce ONLY males of high quality and vigor to compete with wild males for female mates. Because an all-male colony will be difficult to maintain (!), this character ideally would be a conditional trait, perhaps dependent upon temperature or some other environmental cue. Developing genetic-sexing systems based on transgenic methods could result in more-stable lines than those based on translocations. Likewise, developing transgenic methods to cause male sterility also could be beneficial for a sterile-insect release method (SIRM) program. Sterilizing males by irradiation makes them less fit because it causes general somatic damage. Eliminating this fitness loss could allow fewer males to be released, also resulting in a significant savings in program costs.

Recent studies on dipteran pests, such as *Lucilia cuprina* (Concha and Scott 2009), the Mediterranean fruit fly (Pane et al. 2005, Saccone et al. 2011), *Anastrepha* fruit flies (Ruiz et al. 2007, Sarno et al. 2010), and *Bactrocera dorsalis* and *B. correcta* (Permpoon et al. 2011), indicate that *doublesex<sup>+</sup>*, *transformer<sup>+</sup>* and *transformer-2<sup>+</sup>* are homologs of the *Drosophila* genes and function in a similar manner. In fact, Schetelig et al. (2012) were able to induce male-only progeny in *Anastrepha suspensa* using RNA interference (RNAi) by injecting double-stranded (ds) RNA for *transformer<sup>+</sup>* and *transformer-2<sup>+</sup>* into embryos. Nearly all the XX embryos, which should have been female, developed into fully masculinized males with no apparent female morphology and with gonads that were hypertrophied. They concluded that both *transformer<sup>+</sup>* and *transformer-2<sup>+</sup>* genes are essential to development of *Anastrepha* females and that this technology could result in male-only populations for genetic-control programs.

Sex determination in *Bombyx mori*, in which females are ZW and males are WW, has been studied (Ohbayashi et al. 2001, Suzuki et al. 2001, Nagaraja et al. 2005, Shukla et al. 2011). The knowledge obtained holds promise of developing genetic-sexing strains in pest Lepidoptera for genetic-control programs (Marec et al. 2005).

#### 10.14.2 Genetic Improvement of Parasitoids

Genetic improvement of parasitoids reared for augmentative biological control could be achieved if the proportion of females produced in the rearing program

could be increased. Experiments have been conducted to determine whether it is possible to artificially transfer thelytoky-inducing strains of *Wolbachia* from one *Trichogramma* species to another so that the new strain could be improved by increasing the proportion of female progeny (Grenier et al. 1998). *Wolbachia* isolated from an infected species, *Trichogramma pretiosum*, was transferred by microinjection into the pupae of an uninfected species, *T. dendrolimi*. The *Wolbachia* were found in the recipient species 26 generations after the transfer, but only a partial level of thelytoky was observed, perhaps because the density of symbionts was too low or because symbiont–host interactions interfered with the expression of the thelytoky phenotype.

### 10.15 Conclusion

Gempe and Bey (2010) reviewed sex-determination systems in insects and concluded:

*"One common theme from these studies is that evolved mechanisms produce activities in either males or females to control a shared gene switch that regulates sexual development. Only a few small-scale changes in existing and duplicated genes are sufficient to generate large differences in sex determination systems."*

### References Cited

- Adachi-Hagimori, T., Miura, K., and Stouthamer, R., (2008). A new cytogenetic mechanism for bacterial endosymbiont-induced parthenogenesis in Hymenoptera. *Proc. R. Soc. Lond. B* **275**: 2667–2673.
- Anbutsu, H., and Fukatsu, T., (2003). Population dynamics of male-killing and non-male-killing spiroplasmas in *Drosophila melanogaster*. *Appl. Environ. Microbiol.* **69**: 1428–1434.
- Antolin, M.F., (1999). A genetic perspective on mating systems and sex ratios of parasitoid wasps. *Res. Popul. Ecol.* **41**: 29–37.
- Ashburner, M., (1989). *Drosophila A Laboratory Handbook*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Baker, B.S., (1989). Sex in flies: the splice of life. *Nature* **340**: 521–524.
- Baker, B.S., Gorman, M., and Marin, I., (1994). Dosage compensation in *Drosophila*. *Annu. Rev. Genet.* **28**: 491–521.
- Bandi, C., Slatko, B., and O'Neill, S.L., (1999). *Wolbachia* genomes and the many faces of symbiosis. *Parasitol. Today* **15**: 428–429.
- Barton, N.H., and Charlesworth, B., (1998). Why sex and recombination? *Science* **281**: 1986–1990.
- Beeman, R.W., and Friesen, K.S., (1999). Properties and natural occurrence of maternal-effect selfish genes ('Medea' factors) in the red flour beetle, *Tribolium castaneum*. *Heredity* **82**: 529–534.
- Beeman, R.W., Friesen, K.S., and Denell, R.E., (1992). Maternal-effect selfish genes in flour beetles. *Science* **256**: 89–92.
- Belote, J.M., McKeown, M.B., Andrew, D.J., Scott, T.N., Wolfner, M.F., and Baker, B.S., (1985). *Cold Spring Harbor Symp. Quant. Biol.* **50**: 605–614.
- Beukeboom, L.W., (1995). Sex determination in Hymenoptera: a need for genetic and molecular studies. *Bioessays* **17**: 813–817.

- Beukeboom, L.W., and Werren, J.H., (1992). Population genetics of a parasitic chromosome: experimental analysis of PSR in subdivided populations. *Evolution* **46**: 1257–1268.
- Beukeboom, L.W., and Werren, J.H., (1993). Deletion analysis of the selfish B chromosome, paternal sex ratio (PSR), in the parasitic wasp *Nasonia vitripennis*. *Genetics* **133**: 637–648.
- Beukeboom, L.W., Reed, K.M., and Werren, J.H., (1992). Effects of deletions on mitotic stability of the paternal-sex-ratio (PSR) chromosome from *Nasonia*. *Chromosoma* **102**: 20–26.
- Beukeboom, L.W., Kamping, A., and van de Zande, L., (2007). Sex determination in the haplodiploid wasp *Nasonia vitripennis* (Hymenoptera: Chalcidoidea): a critical consideration of models and evidence. *Semin. Cell Dev. Biol.* **18**: 371–378.
- Beye, M., Moritz, R.F.A., Crozier, R.H., and Crozier, Y.C., (1996). Mapping the sex locus of the honeybee (*Apis mellifera*). *Naturwissenschaften* **83**: 424–426.
- Beye, M., Hunt, G.J., Page, R.E., Fondrk, M.K., Grohmann, L., and Moritz, R.F.A., (1999). Unusually high recombination rate detected in the sex locus region of the honey bee (*Apis mellifera*). *Genetics* **153**: 1701–1708.
- Bigot, Y., Jegot, G., Casteret, S., Aupine, P., and Tasel, J.N., (2010). DNA modifications and genome rearrangements during the development and sex differentiation of the bumble bee. *Bombus terrestris*. *Insect Mol. Biol.* **20**: 165–175.
- Bordenstein, S.R., O'Hara, F.P., and Werren, J.H., (2001). *Wolbachia*-induced incompatibility precedes other hybrid incompatibilities in *Nasonia*. *Nature* **409**: 707–710.
- Bourtzis, K., and O'Neill, S., (1998). *Wolbachia* infections and arthropod reproduction. *Bioscience* **48**: 287–293.
- Bownes, M., (1992). Molecular aspects of sex determination in insects. In: Crampton, J.M., and Eggleston, P., Eds., *Insect Molecular Science* Academic, London, UK, pp. 76–100.
- Bull, J.J., (1983). *Evolution of Sex Determining Mechanisms*. Benjamin/Cummings, London, UK.
- Bushland, R.C., (1985). Eradication program in the southwestern United States. In: Graham, O.H., Ed., *Symposium on Eradication of the Screwworm from the United States and Mexico*. Entomological Society of America, College Park, MD, pp. 12–15. Misc. Publ. **62**.
- Butcher, R.D.J., Whitfield, W.G.F., and Hubbard, S.F., (2000). Single-locus complementary sex determination in *Diadegma crysostictos* (Gmelin) (Hymenoptera: Ichneumonidae). *Heredity* **91**: 104–111.
- Charnov, E., (1982). *The Theory of Sex Allocation*. Princeton University Press, NJ.
- Clausen, S., and Ullerich, F.H., (1990). Sequence homology between a polytene band in the genetic sex chromosomes of *Chrysomya rufifacies* and the *daughterless* gene of *Drosophila melanogaster*. *Naturwissenschaften* **77**: 137–138.
- Cline, T.W., and Meyer, B.J., (1996). Vive la difference: males vs females in flies vs worms. *Annu. Rev. Genet.* **30**: 637–702.
- Concha, C., and Scott, M.J., (2009). Sexual development in *Lucilia cuprina* (Diptera, Calliphoridae) is controlled by the *transformer* gene. *Genetics* **182**: 785–798.
- Cordaux, R., Bouchon, D., and Greve, P., (2011). The impact of endosymbionts on the evolution of host sex-determination mechanisms. *Trends Genet.* **27**: 332–341.
- Cook, J.M., (1993). Sex determination in the Hymenoptera: a review of models and evidence. *Heredity* **71**: 421–435.
- Cook, J.M., and Butcher, R.D.J., (1999). The transmission and effects of *Wolbachia* bacteria in parasitoids. *Res. Popul. Ecol.* **41**: 15–28.
- Crozier, R.H., (1971). Heterozygosity and sex determination in haplo-diploidy. *Am. Nat.* **105**: 339–412.
- Crow, J.F., (1994). Advantages of sexual reproduction. *Dev. Genet.* **15**: 205–213.
- Dallai, R., Fanciulli, P.P., and Frati, F., (1999). Chromosome elimination and sex determination in springtails (Insecta, Collembola). *J. Exp. Zool.* **285**: 215–225.
- Dallai, R., Fanciulli, P.P., and Frati, F., (2000). Aberrant spermatogenesis and the peculiar mechanism of sex determination in symphyleonan Collembola (Insecta). *Heredity* **91**: 351–358.

- DeLoof, A., and Huybrechts, R., (1998). "Insects do not have sex hormones": a myth? *Gen. Comp. Endocrinol.* **111**: 245–260.
- Dobson, S.L., and Tanouye, M.A., (1998a). Evidence for a genomic imprinting sex determination mechanism in *Nasonia vitripennis* (Hymenoptera: Chalcidoidea). *Genetics* **149**: 233–242.
- Dobson, S.L., and Tanouye, M.A., (1998b). Interspecific movement of the paternal sex ratio chromosome. *Heredity* **81**: 261–269.
- Dobson, S.L., Bourtzis, K., Braig, H.R., Jones, B.F., Zhou, W., Rousset, F., and O'Neill, S.C., (1999). *Wolbachia* infections are distributed throughout insect somatic and germ line tissues. *Insect Biochem. Mol. Biol.* **29**: 153–160.
- Dubendorfer, A., Hediger, M., Burghardt, G., and Bopp, D., (2002). *Musca domestica*, a window on the evolution of sex-determining mechanisms in insects. *Int. J. Dev. Biol.* **46**: 75–79.
- Ebbert, M.A., (1991). The interaction phenotype in the *Drosophila willistoni*-*spiroplasma* symbiosis. *Evolution* **45**: 971–988.
- Ebbert, M.A., (1993). Endosymbiotic sex ratio distorters in insects and mites. In: Wrensch, D.L., and Ebbert, M.A., Eds., *Evolution and Diversity of Sex Ratio in Insects and Mites*, Chapman and Hall, New York, pp. 150–191.
- Eickbush, D.G., Eickbush, T.H., and Werren, J.H., (1992). Molecular characterization of repetitive DNA sequences from a B chromosome. *Chromosoma* **101**: 575–583.
- El Agoze, M., Drezen, J.M., Renault, S., and Periquet, G., (1994). Analysis of the reproductive potential of diploid males in the wasp *Diadromus pulchellus* (Hymenoptera: Ichneumonidae). *Bull. Entomol. Res.* **84**: 213–218.
- Ephrussi, A., and Lehmann, R., (1992). Induction of germ cell formation by oskar. *Nature* **358**: 387–392.
- Erickson, J.W., and Quintero, J.J., (2007). Indirect effects of ploidy suggest X chromosome dose, not the X:A ratio, signals sex in *Drosophila*. *PLoS Biol.* **5**: e332.
- Fast, E.M., Toomey, M.E., Panaram, K., Desjardins, D., Kolaczyk, E.D., and Frydman, H.M., (2011). *Wolbachia* enhance *Drosophila* stem cell proliferation and target the germline stem cell niche. *Science* **334**: 990–992.
- Fialho, R.E., and Stevens, L., (2000). Male-killing *Wolbachia* in a flour beetle. *Proc. R. Soc. Lond. B* **267**: 1469–1474.
- Ganetzky, B., (2000). Tracking down a cheating gene. *Am. Sci.* **88**: 128–135.
- Gempe, T., and Beye, M., (2010). Function and evolution of sex determination mechanisms, genes and pathways in insects. *Bioessays* **33**: 52–60.
- Gherna, R.L., Werren, J.H., Weisburg, W., Cote, R., Woese, C.R., and Mandelco, L., et al. (1991). *Arsenophonus nasoniae* gen. nov., sp. nov., the causative agent of the son-killer trait in the parasitic wasp *Nasonia vitripennis*. *Int. J. Syst. Bacteriol.* **41**: 563–565.
- Greenspan, R.J., and Ferveur, J.F., (2000). Courtship in *Drosophila*. *Annu. Rev. Genet.* **34**: 205–232.
- Grenier, S., Pintureau, B., Heddi, A., Lassabliere, F., Jager, C., Louis, C., and Khatchadourian, C., (1998). Successful horizontal transfer of *Wolbachia* symbionts between *Trichogramma* wasps. *Proc. R. Soc. Lond. B* **265**: 1441–1445.
- Hackstein, J.H.P., and Hochstenbach, R., (1995). The elusive fertility genes of *Drosophila*: the ultimate haven for selfish genetic elements. *Trends Genet.* **11**: 195–200.
- Hamilton, W.D., (1967). Extraordinary sex ratios. *Science* **156**: 477–488.
- Hamilton, W.D., Axelrod, R., and Tanese, R., (1990). Sexual reproduction as an adaptation to resist parasites (a review). *Proc. Natl. Acad. Sci. USA* **87**: 3566–3573.
- Hartl, D.L., and Brown, S.W., (1970). The origin of male haploid genetic systems and their expected sex ratio. *Theor. Pop. Biol.* **1**: 165–190.
- Hasselmann, M., Gempe, T., Schiott, M., Nunes-Silva, C.G., Otte, M., and Beye, M., (2008). Evidence for the evolutionary nascence of a novel sex determination pathway in honeybees. *Nature* **454**: 519–522.

- Heimpel, G.E., and de Boer, J.G., (2008). Sex determination in the Hymenoptera. *Annu. Rev. Entomol.* **53**: 209–230.
- Heimpel, G.E., Antolin, M.F., and Strand, M.R., (1999). Diversity of sex-determining alleles in *Bracon hebetor*. *Heredity* **82**: 282–291.
- Hennig, W., (1993). Conventional protein coding genes in the *Drosophila* Y chromosome: Is the puzzle of the fertility gene function solved? *Proc. Natl. Acad. Sci. USA* **90**: 10904–10906.
- Heinrich, J.C., and Scott, M.J., (2000). A repressible female-specific lethal genetic system for making transgenic insect strains suitable for a sterile-release program. *Proc. Natl. Acad. Sci. USA* **97**: 8229–8232.
- Hilfiker-Kleiner, D., Dubendorfer, A., Hilfiker, A., and Nothiger, R., (1993). Developmental analysis of two sex-determining genes, *M* and *F*, in the housefly, *Musca domestica*. *Genetics* **134**: 1187–1194.
- Hiroki, M., et al. (2002). Feminization of genetic males by a symbiotic bacterium in a butterfly, *Eurema hecabe* (Lepidoptera: Pieridae). *Naturwissenschaften* **89**: 167–170.
- Hochstenbach, R., Brand, R., and Hennig, W., (1994). Transcription of repetitive DNA sequences in the lampbrush loop pair *Nooses* formed by sterile alleles of fertility gene *Q* on the Y chromosome of *D. hydei*. *Mol. Gen. Genet.* **244**: 653–660.
- Huger, A.M., Skinner, S.W., and Werren, J.H., (1985). Bacterial infections associated with the son-killer trait in the parasitoid wasp *Nasonia* (=*Mormoniella*) *vitripennis* (Hymenoptera: Pteromalidae). *J. Invert. Path.* **46**: 272–280.
- Huigens, M.E., Luck, R.F., Klaassen, R.H.G., Maas, M.F.P.M., Timmermans, M.J.T.N., and Stouthamer, R., (2000). Infectious parthenogenesis. *Nature* **405** (178–179).
- Hunter, M.S., Nur, U., and Werren, J.H., (1993). Origin of males by genome loss in an autoparasitoid wasp. *Heredity* **70**: 162–171.
- Hurst, L.D., (1993). The incidences, mechanisms and evolution of cytoplasmic sex ratio distorters in animals. *Biol. Rev.* **68**: 121–193.
- Hurst, G.D.D., and Jiggins, F.M., (2000). Male-killing bacteria in insects: mechanisms, incidence and implications. *Emerg. Infect. Dis.* **6**: 329–336.
- Hurst, L.D., and Pomiankowski, A., (1991). Causes of sex ratio bias may account for unisexual sterility in hybrids: a new explanation of Haldane's rule and related phenomena. *Genetics* **128**: 841–858.
- Hurst, G.D.D., and Schilthuizen, M., (1998). Selfish genetic elements and speciation. *Heredity* **80**: 2–8.
- Hurst, G.D.D., Hammarton, T.M., Bandi, C., Majerus, T.M.O., Bertrand, D., and Majerus, M.E.N., (1997). The diversity of inherited parasites of insects: the male-killing agent of the ladybird beetle *Coleomegilla maculata* is a member of the Flavobacteria. *Genet. Res.* **70**: 1–6.
- Hurst, G.D.D., Schulenburg, J.H.G.V.D., Majerus, T.M.O., Bertrand, D., Zakarov, I.A., and Baungaard, J., et al. (1999a). Invasion of one insect species, *Adalia bipunctata*, by two different male-killing bacteria. *Insect Mol. Biol.* **8**: 133–139.
- Hurst, G.D.D., Jiggins, F.M., Schulenburg, J.H.G.V.D., Bertrand, D., West, S.A., and Goriacheva, I.I., et al. (1999b). Male-killing *Wolbachia* in two species of insect. *Proc. R. Soc. Lond. B* **266**: 735–740.
- Inoue, H., and Hiroyoshi, T., (1986). A maternal-effect sex-transformation mutant of the housefly, *Musca domestica* L. *Genetics* **112**: 469–482.
- Jaenike, J., (2001). Sex chromosome meiotic drive. *Annu. Rev. Ecol. Syst.* **32**: 25–49.
- Janzer, B., and Steinmann-Zwicky, M., (2001). Cell-autonomous and somatic signals control sex-specific gene expression in XY germ cells of *Drosophila*. *Mech. Dev.* **100**: 3–13.
- Jeyaprakash, A., and Hoy, M.A., (2000). Long PCR improves *Wolbachia* DNA amplification: *wsp* sequences found in 76% of 63 arthropod species. *Insect Mol. Biol.* **9**: 393–405.
- Jiggins, F.M., Hurst, G.D.D., and Majerus, M.E.N., (1998). Sex ratio distortion in *Acraea encedon* (Lepidoptera: Nymphalidae) is caused by a male-killing bacterium. *Heredity* **81**: 87–91.

- Jiggins, F.M., Hurst, G.D.D., and Majerus, M.E.N., (1999). How common are meiotically driving sex chromosomes in insects? *Am. Nat.* **154**: 481–483.
- Jones, R.N., and Rees, H., (1982). *B Chromosomes*. Academic Press, New York.
- Kageyama, D., and Traut, W., (2004). Opposite sex-specific effects of *Wolbachia* and interference with the sex determination of its host *Ostrinia scapulalis*. *Proc. R. Soc. Lond. B* **271**: 251–258.
- Kaiser, V.B., and Bachtrog, D., (2010). Evolution of sex chromosomes in insects. *Annu. Rev. Genet.* **44**: 91–112.
- Keightley, P.D., and Eyre-Walker, A., (2000). Deleterious mutations and the evolution of sex. *Science* **290**: 331–333.
- Kelley, R.L., and Kuroda, M.I., (2000). Noncoding RNA genes in dosage compensation and imprinting. *Cell* **103**: 9–12.
- Knipling, E.F., (1955). Possibilities of insect control or eradication through the use of sexually sterile males. *J. Econ. Entomol.* **48**: 459–462.
- Knipling, E.F., (1985). Sterile insect technique as a screwworm control measure: the concept and its development. In: Graham, O.H., Ed., *Symposium on Eradication of the Screwworm from the United States and Mexico*. Entomological Society of America, College Park, MD, pp. 4–8. Misc. Publ. 62.
- Kopp, A., Duncan, I., and Carroll, S.B., (2000). Genetic control and evolution of sexually dimorphic characters. *Nature* **408**: 553–559.
- Kuhn, S., Sievert, V., and Traut, W., (2000). The sex-determining gene *doublesex* in the fly *Megaselia scalaris*: conserved structure and sex-specific splicing. *Genome* **43**: 1011–1020.
- LaChance, L.E., (1979). Genetic strategies affecting the success and economy of the sterile insect release method. In: Hoy, M.A., and McKelvey Jr., J.J., Eds., *Genetics in Relation to Insect Management*, Rockefeller Found Press, New York, pp. 8–18.
- Lattorff, H.M.G., Moritz, R.F.A., and Fuchs, S., (2005). A single locus determines thelytokous parthenogenesis of laying honeybee workers (*Apis mellifera capensis*). *Heredity* **94**: 533–537.
- Lattorff, H.M.G., Moritz, R.F.A., Crewe, R.M., and Solignac, M., (2007). Control of reproductive dominance by the thelytoky gene in honeybees. *Biol. Lett.* **3**: 292–295.
- Lauge, G., (1985). Sex determination: genetic and epigenetic factors. In: Kerkut, G.A., and Gilbert, L.I., Eds., *Comprehensive Insect Physiology Biochemistry and Pharmacology, Volume 1, Embryogenesis and Reproduction*, Pergamon Press, Oxford, UK, pp. 298–317.
- Laurie, C.C., (1997). The weaker sex is heterogametic: 75 years of Haldane's Rule. *Genetics* **147**: 937–951.
- Laverty, C., Lucci, J., and Akhtar, A., (2010). The MSL complex: X chromosome and beyond. *Curr. Opin. Genet. Dev.* **20**: 171–178.
- Lawson, E.T., Mousseau, T.A., Klaper, R., Hunter, M.D., and Werren, J.H., (2001). Rickettsia associated with male-killing in a buprestid beetle. *Heredity* **87**: 497–505.
- Lehtonen, J., Jennions, M.D., and Kokko, H., (2012). The many costs of sex. *Trends Ecol. Evol.* **27**: 172–178.
- Leslie, J.F., (1984). A “sex-ratio” condition in *Oncopeltus fasciatus*. *Heredity* **75**: 260–264.
- Leventhal, E., (1968). The sex ratio condition in *Drosophila bifasciata*; its experimental transmission to several species of *Drosophila*. *J. Invertebr. Pathol.* **11**: 170–183.
- Lewis, W.M., (1987). The cost of sex. In: Stearns, S.C., Ed., *The Evolution of Sex and Its Consequences* Birkhauser Verlag, Basel. Switzerland, pp. 33–57.
- Li, H., and Baker, B.S., (1998). *hermaphrodite* and *doublesex* function both dependently and independently to control various aspects of sexual differentiation in *Drosophila*. *Development* **125**: 2641–2651.
- Lucchesi, J.C., (1999). On the origin of sex chromosomes. *Bioessays* **21**: 188–190.
- Lucchesi, J.C., (2009). The structure-function link of compensated chromatin in *Drosophila*. *Curr. Opin. Genet. Dev.* **19**: 550–556.
- Lyttle, T.W., (1993). Cheaters sometimes prosper: distortion of mendelian segregation by meiotic drive. *Trends Genet.* **9**: 205–210.

- Macke, E., Magalhaes, S., Do-Thi Khan, H., Luciano, A., Frantz, A., Facon, B., and Olivier, I., (2011). Sex allocation in haplodiploids is mediated by egg size: evidence in the spider mite *Tetranychus urticae* Koch. *Proc. R. Soc. London* **278**: 1054–1063.
- Majerus, M.E.N., and Hurst, G.D.D., (1997). Ladybirds as a model system for the study of male-killing symbionts. *Entomophaga* **42**: 13–20.
- Majerus, T.M., and Majerus, M.E., (2010). Intergenomic arms races: detection of a nuclear rescue gene of male-killing in a ladybird. *PLoS Path.* **6**: e1000987.
- Majerus, T.M.O., von der Schulenburg, J.H.G., Majerus, M.E.N., and Hurst, G.D.D., (1999). Molecular identification of a male-killing agent in the ladybird *Harmonia axyridis* (Pallas) (Coleoptera: Coccinellidae). *Insect Mol. Biol.* **8**: 551–555.
- Marin, I., and Baker, B.S., (1998). The evolutionary dynamics of sex determination. *Science* **281**: 1990–1994.
- Marin, I., Siegal, M.L., and Baker, B.S., (2000). The evolution of dosage-compensation mechanisms. *Bioessays* **22**: 1106–1114.
- Mank, J.E., (2009). The W, X, Y and Z of sex-chromosome dosage compensation. *Trends Genet.* **25**: 226–233.
- Marec, F., Neven, L.G., Robinson, A.S., Vreysen, M., Goldsmith, M.R., Nagaraju, J., and Franz, G., (2005). Development of genetic sexing strains in Lepidoptera: from traditional to transgenic approaches. *J. Econ. Entomol.* **98**: 248–259.
- Martins, A.B., Ventura, I.M., and Klaczko, L.B., (2010). *Spiroplasma* infection in *Drosophila melanogaster*: what is the advantage of killing males? *J. Invertebr. Pathol.* **105**: 145–150.
- Maynard Smith, J., (1978). *The Evolution of Sex*. Cambridge University Press, Cambridge.
- McAllister, B.E., and McVean, G.A.T., (2000). Neutral evolution of the sex-determining gene *transformer*. *Drosophila. Genetics* **154**: 1711–1720.
- Meadows, M.E., (1985). Eradication program in the southeastern United States. In: Graham, O.H., Ed., *Symposium on Eradication of the Screwworm from the United States and Mexico*. Entomological Society of America, College Park, MD, pp. 8–11. *Misc. Publ.* **62**.
- Meise, M., Hilfiker-Kleiner, D., Dubendorfer, A., Brunner, C., Nothiger, R., and Bopp, D., (1998). *Sex-lethal*, the master sex-determining gene in *Drosophila*, is not sex-specifically regulated in *Musca domestica*. *Development* **125**: 1487–1494.
- Merrill, C., Bayraktaroglu, L., Kusano, A., and Ganetzky, B., (1999). Truncated RanGAP encoded by the *Segregation Distorter* locus of *Drosophila*. *Science*, **293**: 1742–1745.
- Michod, R.E., and Levin, B.R., Eds., (1988). *The Evolution of Sex*, Sinauer, Sunderland, MA.
- Mittwoch, U., (1996). Sex-determining mechanisms in animals. *Trends Ecol. Evol.* **11**: 63–67.
- Muller, H.J., (1964). The relation of recombination to mutational advance. *Mutat. Res.* **1**: 2–9.
- Muller-Holtkamp, F., (1995). The *Sex lethal* gene homologue in *Chrysomya rufifacies* is highly conserved in sequence and exon-intron organization. *J. Mol. Evol.* **41**: 467–477.
- Nagaraja, G.M., Mahesh, G., Satish, V., Madhu, M., Muthulakshmi, M., and Nagaraju, J., (2005). Genetic mapping of W chromosome-specific markers in the silkworm, *Bombyx mori*. *Heredity* **95**: 148–157.
- Naito, T., and Suzuki, H., (1991). Sex determination in the sawfly, *Athalia rosae ruficornis* (Hymenoptera): occurrence of triploid males. *J. Hered.* **82**: 101–104.
- Narita, S., Kageyama, D., Nomuroa, M., and Fukatsu, T., (2007). Unexpected mechanism of symbiont-induced reversal of insect sex: feminizing *Wolbachia* continuously acts on the butterfly *Eurema hecate* during larval development. *Appl. Environ. Microbiol.* **73**: 4322–4341.
- Negri, I., Pellecchia, M., Mozzoglio, P.J., Patetta, A., and Alma, A., (2006). Feminizing *Wolbachia* in *Zyginidia pullula* (Insecta: Hemiptera), a leafhopper with an XX/XO sex-determination system. *Proc. Biol. Sci.* **273**: 2409–2416.
- Negri, I., Franchini, A., Gonella, E., Daffonchio, D., Mazzoglio, P.J., Mandrioli, M., and Alma, A., (2009). Unravelling the *Wolbachia* evolutionary role: the reprogramming of the host genomic imprinting. *Proc. Biol. Sci.* **278**: 2485–2491.

- Normark, B.B., (2004). Haplodiploidy as an outcome of coevolution between male-killing cytoplasmic elements and their hosts. *Evolution* **58**: 790–798.
- Nothiger, R., and Steinmann-Zwicky, M., (1985). A single principle for sex determination in insects. *Cold Spring Harbor Symp. Quant. Biol.* **50**: 615–621.
- Nur, U., Werren, J.H., Eickbush, D.G., Burke, W.D., and Eickbush, T.H., (1988). A “selfish” B chromosome that enhances its transmission by eliminating the paternal genome. *Science* **240**: 512–514.
- Ohbayashi, F., Suzuki, M.G., Mita, K., Okano, K., and Shimada, T., (2001). A homologue of the *Drosophila doublesex* gene is transcribed into sex-specific mRNA isoforms in the silkworm. *Bombyx mori. Comp. Biochem. Physiol. B. Biochem. Mol. Biol* **128**: 145–158.
- Oliveira, D.C.S.G., Werren, J.H., Verhulst, E.C., Giebel, J.D., Kamping, A., Beukeboom, L.W., and van de Zande, L., (2009). Identification and characterization of the *doublesex* gene of *Nasonia*. *Insect Mol. Biol.* **18**: 315–324.
- O'Neill, S.L., Hoffmann, A.A., and Werren, J.H., Eds., (1997). *Influential Passengers. Inherited Microorganisms and Arthropod Reproduction*, Oxford University Press, Oxford, UK.
- Palopoli, M., Doshi, P., and Wu, C.I., (1994). Characterization of two *Segregation distorter* revertants: evidence that the tandem duplication is necessary for *Sd* activity in *Drosophila melanogaster*. *Genetics* **136**: 209–215.
- Pane, A., De Simone, A., Saccone, G., and Polito, C., (2005). Evolutionary conservation of *Ceratitis capitata transformer* gene function. *Genetics* **171**: 615–624.
- Pannebakker, B.A.L.P., Pijnacker, B.J., Zwaan, and Beukeboom, L.W., (2004). Cytology of *Wolbachia*-induced parthenogenesis in *Leptopilina clavipes* (Hymenoptera: Figitidae). *Genome* **47**: 299–303.
- Partridge, L., and Hurst, L.D., (1998). Sex and conflict. *Science* **281**: 2003–2008.
- Pauli, D., and Mahowald, A.P., (1990). Germ-line sex determination in *Drosophila melanogaster*. *Trends Genet.* **6**: 259–264.
- Permpoon, R., Aketarawong, N., and Thanaphum, S., (2011). Isolation and characterization of *doublesex* homologues in the *Bactrocera* species: *B. dorsalis* (Hendel) and *B. correcta* (Bezzi) and their putative promoter regulatory regions. *Genetica* **139**: 113–127.
- Pineda-Vargas, N., (1985). Screwworm eradication in Mexico: activities of the Mexico-American screwworm eradication commission, 1977–84. In: Graham, O.H., Ed., *Symposium on Eradication of the Screwworm from the United States and Mexico*. Entomological Society of America, College Park, MD, pp. 22–27. *Misc. Publ.* **62**.
- Presgraves, D.C., Severance, E., and Wilkinson, G.S., (1997). Sex chromosome meiotic drive in stalk-eyed flies. *Genetics* **147**: 1169–1180.
- Rabeling, C., Gonzales, O., Schultz, T.R., Bacci, M., Garcia, M.V.B., and Verhaagh, M., et al. (2011). Cryptic sexual populations account for genetic diversity and ecological success in a widely distributed, asexual fungus-growing ant. *Proc. Natl. Acad. Sci. USA* **108**: 12366–12371.
- Randerson, J.P., Smith, N.G.C., and Hurst, L.D., (2000). The evolutionary dynamics of male-killers and their hosts. *Heredity* **84**: 151–160.
- Rao, S.R.V., and Padmaja, M., (1992). Mammalian-type dosage compensation mechanism in an insect—*Gryllotalpa fossor* (Scudder)—Orthoptera. *J. Biosci.* **17**: 253–273.
- Reed, K.M., and Werren, J.H., (1995). Induction of paternal genome loss by the paternal-sex-ratio chromosome and cytoplasmic incompatibility bacteria (*Wolbachia*): a comparative study of early embryonic events. *Mol. Reprod. Dev.* **40**: 408–418.
- Retnakaran, A., and Percy, J., (1985). Fertilization and special modes of reproduction. In: Kerkut, G.A., and Gilbert, L.I., Eds., *Comprehensive Insect Physiology Biochemistry and Pharmacology, Volume 1, Embryogenesis and Reproduction*, Pergamon Press, Oxford, UK, pp. 213–293.
- Reumer, B.M., van Alphen, J.J.M., and Kraaijeveld, K., (2011). Occasional males in parthenogenetic populations of *Asobara japonica* (Hymenoptera: Braconidae): low *Wolbachia* titer or incomplete coadaptation. *Heredity* **108**: 341–346.
- Rice, W.R., (1994). Degeneration of a nonrecombining chromosome. *Science* **263**: 230–232.

- Rice, W.R., and Chippindale, A.K., (2001). Sexual recombination and the power of natural selection. *Science* **294**: 555–559.
- Richardson, R.H., and Averhoff, W.W., (1978). A cattleman's view of beef production with and without screwworms. In: Richardson, R.H., Ed., *The Screwworm Problem. Evolution of Resistance to Biological Control*. University of Texas Press, Austin, TX, pp. 3–9.
- Rigaud, T., (1999). Further *Wolbachia* endosymbiont diversity: a tree hiding in the forest? *Trends Ecol. Evol.* **14**: 212–213.
- Rigaud, T., and Rousset, F., (1996). What generates the diversity of *Wolbachia*-arthropod interactions? *Biodivers. Conserv.* **5**: 999–1013.
- Robinson, A.S., and Franz, G., (2000). The application of transgenic insect technology in the sterile insect technique. In: Handler, A.M., and James, A.A., Eds., *Insect Transgenesis. Methods and Applications*, CRC Press, Boca Raton, FL, pp. 307–318.
- Rokas, A., (2000). *Wolbachia* as a speciation agent. *Trends Ecol. Evol.* **15**: 44–45.
- Rousset, F., and Raymond, M., (1991). Cytoplasmic incompatibility in insects: why sterilize females? *Trends Ecol. Evol.* **6**: 54–57.
- Ruiz, M.F., Milano, A., Calvemini, M., Elrin-Lopez, J.M., Perondini, A.L.P., and Selivon, D., et al. (2007). The gene *transformer* of *Anastrepha* fruit flies (Diptera, Tephritidae) and its evolution in insects. *PLoS ONE* **11**: e1239.
- Russell, J.E., and Stouthamer, R., (2011). The genetics and evolution of obligate reproductive parasitism in *Trichogramma pretiosum* infected with parthenogenesis-inducing *Wolbachia*. *Heredity* **106**: 58–67.
- Saccone, G., Salvemini, M., and Polito, L.C., (2011). The *transformer* gene of *Ceratitis capitata*: a paradigm for a conserved epigenetic master regulator of sex determination in insects. *Genetica* **139**: 99–111.
- Salz, H.K., (2011). Sex determination in insects: a binary decision based on alternative splicing. *Curr. Opin. Genet. Dev.* **21**: 395–400.
- Sandrock, C., and Vorburger, C., (2011). Single-locus recessive inheritance of asexual reproduction in a parasitoid wasp. *Curr. Biol.* **21**: 433–437.
- Sarno, F., Ruiz, M.F., Elrin-Lopez, J.M., Perondini, A.L.P., Selivon, D., and Sanchez, L., (2010). The gene *transformer-2* of *Anastrepha* fruit flies (Diptera, Tephritidae) and its evolution in insects. *BMC Evol. Biol.* **10**: 140.
- Sawamura, K., (1996). Maternal effect as a cause of exceptions for Haldane's rule. *Genetics* **143**: 609–611.
- Schetelig, M.F., Milano, A., Saccone, G., and Handler, A.M., (2012). Male only progeny in *Anastrepha suspense* by RNAi-induced sex reversion of chromosomal females. *Insect Biochem. Mol. Biol.* **42**: 51–57.
- Schurko, A.M., Neiman, M., and Logsdon, J.M., (2008). Signs of sex: what we know and how we know it. *Trends Ecol. Evol.* **24**: 2008–2217.
- Shirk, P.D., O'Brochta, D.A., Roberts, P.E., and Handler, A.M., (1988). Sex-specific selection using chimeric genes. In: Hedin, P.A., Menn, J.J., and Hollingsworth, R.M., Eds., *Biotechnology for Crop Protection*, American Chemical Society, Washington, DC, pp. 135–146. Am. Chem. Soc. Symp. Series No. 379.
- Shoemaker, D.D., Katju, V., and Jaenike, J., (1999). *Wolbachia* and the evolution of reproductive isolation between *Drosophila recens* and *Drosophila subquadrata*. *Evolution* **53**: 1157–1164.
- Shukla, J.N., Jadhav, S., and Nagaraju, J., (2011). Novel female-specific splice form of *dsx* in the silkworm, *Bombyx mori*. *Genetica* **139**: 23–31.
- Siegal, M.L., and Baker, B.S., (2005). Functional conservation and divergence of *intersex*, a gene required for female differentiation in *Drosophila melanogaster*. *Dev. Genes Evol.* **215**: 1–12.
- Siever, V., Kuhn, S., Paululat, A., and Traut, W., (2000). Sequence conservation and expression of the Sex-lethal homologue in the fly *Megaselia scalaris*. *Genome* **43**: 382–390.
- Slee, R., and Bownes, M., (1990). Sex determination in *Drosophila melanogaster*. *Quart. Rev. Biol.* **65**: 175–204.

- Smith, E.R., Pannuti, A., Gu., W., Steurnagel, A., Cook, R.G., Allis, C.D., and Lucchesi, J.C., (2000). The *Drosophila* MSL complex acetylates histone H4 at lysine 16, a chromatin modification linked to dosage compensation. *Mol. Cell. Biol.* **20**: 312–318.
- Snow, J.W., Whitten, C.J., Salinas, A., Ferrer, J., and Sudlow, W.H., (1985). The screwworm (*Cochliomyia hominivorax*) (Diptera: Calliphoridae) in Central America and proposed plans for its eradication south to the Darien Gap in Panama. *J. Med. Entomol.* **22**: 353–360.
- Sokolova, M.I., Zinkevich, N.S., and Zakharov, A., (2002). Bacteria in the ovarioles of females from maleless families of ladybird beetles *Adalia bipunctata* L. (Coleoptera: Coccinellidae) naturally infected with *Rickettsia*, *Wolbachia*, and *Spiroplasma*. *J. Invertebr. Pathol.* **79**: 72–79.
- Somerson, N.L., Ehrman, L., Kocka, J.P., and Gottlieb, F.J., (1984). Streptococcal L-forms isolated from *Drosophila paulistorum* semispecies cause sterility in male progeny. *Proc. Natl. Acad. Sci. USA* **81**: 282–285.
- Stevens, L., (1993). Cytoplasmically inherited parasites and reproductive success in *Tribolium* flour beetles. *Anim. Behav.* **46**: 305–310.
- Stevens, L., Giordano, R., and Fialho, R.F., (2001). Male-killing, nematode infections, bacteriophage infection, and virulence of cytoplasmic bacterial in the genus *Wolbachia*. *Annu. Rev. Ecol. Syst.* **32**: 519–545.
- Stouthamer, R., (1997). *Wolbachia*-induced parthenogenesis. In: O'Neill, S.L., Hoffmann, A.A., and Werren, J.H., Eds., *Influential Passengers. Inherited Microorganisms and Arthropod Reproduction*, Oxford University Press, Oxford, UK, pp. 102–124.
- Stouthamer, R., and Luck, R.F., (1993). Influence of microbe-associated parthenogenesis on the fecundity of *Trichogramma deion* and *T. pretiosum*. *Entomol. Exp. Appl.* **67**: 183–192.
- Stouthamer, R., and Werren, J.H., (1993). Microbes associated with parthenogenesis in wasps of the genus *Trichogramma*. *J. Invertebr. Pathol.* **61**: 6–9.
- Stouthamer, R., Luck, R.F., and Hamilton, W.D., (1990). Antibiotics cause parthenogenetic *Trichogramma* (Hymenoptera/Trichogrammatidae) to revert to sex. *Proc. Natl. Acad. Sci. USA* **87**: 2424–2427.
- Stouthamer, R., Luck, R.F., and Werren, J.H., (1992). Genetics of sex determination and the improvement of biological control using parasitoids. *Environ. Entomol.* **21**: 427–435.
- Stouthamer, R., Breeuwer, J.A.J., and Hurst, G.D.D., (1999). *Wolbachia pipiensis*: microbial manipulator of arthropod reproduction. *Annu. Rev. Microbiol.* **53**: 71–102.
- Sugimoto, T.N., Fujii, T., Kayukawa, T., Sakamoto, H., and Ishidawa, Y., (2010). Expression of a doublesex homologue is altered in sexual mosaics of *Ostrinia scapulalis* moths infected with *Wolbachia*. *Insect Biochem. Mol. Biol.* **40**: 847–854.
- Suzuki, M.G., Ohbayashi, F., Mita, K., and Shimada, T., (2001). The mechanism of sex-specific splicing at the doublesex gene is different between *Drosophila melanogaster* and *Bombyx mori*. *Insect Biochem. Mol. Biol.* **31**: 1201–1211.
- Thomson, M.S., and Beeman, R.W., (1999). Assisted suicide of a selfish gene. *Heredity* **90**: 191–194.
- Thornhill, R., and Alcock, J., (1983). *The Evolution of Insect Mating Systems*. Harvard University Press, Cambridge, MA.
- Traut, W., (1994). Sex determination in the fly *Megaselia scalaris*, a model system for primary steps of sex-chromosome evolution. *Genetics* **126**: 1097–1104.
- Traut, W., and Willhoeft, U., (1990). A jumping sex determining factor in the fly *Megaselia scalaris*. *Chromosoma* **99**: 407–412.
- Turelli, M., and Hoffmann, A.A., (1999). Microbe-induced cytoplasmic incompatibility as a mechanism for introducing transgenes into arthropod populations. *Insect Mol. Biol.* **8**: 243–255.
- Vavre, F.F., Flury, J., Varaldi, P., Fouillet, and Bouletreau, M., (2000). Evidence for female mortality in *Wolbachia*-mediated cytoplasmic incompatibility in haplodiploid insects: epidemiologic and evolutionary consequences. *Evolution* **54**: 191–200.

- Vavre, F., Dedeine, F., Quillon, M., Fouillet, P., Fleury, F., and Bouletreau, M., (2001). Within-species diversity of *Wolbachia*-induced cytoplasmic incompatibility. *Evolution* **55**: 1710–1714.
- Veneti, Z., et al. (2005). A functional dosage compensation complex required for male killing in *Drosophila*. *Science* **307**: 1461–1463.
- Verhulst, E.C., van de Zande, L., and Beukeboom, L.W., (2010a). Insect sex determination: it all evolves around *transformer*. *Curr. Opin. Genet. Dev.* **20**: 376–383.
- Verhulst, E.C., Beukeboom, L.W., and van de Zande, L., (2010b). Maternal control of haplodiploidy sex determination in the wasp *Nasonia*. *Science* **328**: 620–623.
- Vincent, S., Perkins, L.A., and Perrimon, N., (2001). Doublesex surprises. *Cell* **106**: 399–402.
- von der Schulenburg, J.H.G., Habig, M., Sloggett, J.J., Webberley, K.M., Bertrand, D., and Hurst, G.D.D., et al. (2001). Incidence of male-killing *Rickettsia* spp. ( $\alpha$ -Proteobacteria) in the ten-spot ladybird beetle *Adalia decempunctata* L. (Coleoptera: Coccinellidae). *Appl. Environ. Microbiol.* **67**: 270–277.
- Wang, W., Kidd, B.J., Carroll, S.B., and Yoder, J.H., (2011). Sexually dimorphic regulation of the Wingless morphogen controls sex-specific segment number in *Drosophila*. *Proc. Natl. Acad. Sci. USA* **108**: 11139–11144.
- Weeks, A.R., Marec, F., and Breeuwer, J.A.J., (2001). A mite species that consists entirely of haploid females. *Science* **292**: 2479–2482.
- Werren, J.H., (1997). Biology of *Wolbachia*. *Annu. Rev. Entomol.* **42**: 587–609.
- Werren, J.H., and Beukeboom, L.W., (1998). Sex determination, sex ratios, and genetic conflict. *Annu. Rev. Ecol. Syst.* **29**: 233–261.
- Werren, J.H., Nur, U., and Eickbush, D., (1987). An extrachromosomal factor causing loss of paternal chromosomes. *Nature* **327**: 75–76.
- Werren, J.H., Windsor, D., and Guo, L., (1995). Distribution of *Wolbachia* among Neotropical arthropods. *Proc. R. Soc. Lond. B* **262**: 197–204.
- Werren, J.H., Hurst, G.D.D., Zhang, W., Breeuwer, J.A.J., Stouthamer, R., and Majerus, M.E.N., (1994). Rickettsial related male-killing in the ladybird beetle (*Adalia bipunctata*). *J. Bacteriol.* **176**: 388–394.
- West, S.A., Herre, E.A., and Sheldon, B.C., (2000). The benefits of allocating sex. *Science* **290**: 288–290.
- White, M.J.D., (1973). *Animal Cytology and Evolution*. Cambridge Univ. Press.
- Whiting, P.W., (1943). Multiple alleles in complementary sex determination of *Habrobracon*. *Genetics* **28**: 365–382.
- Wilkins, A.S., (1995). Moving up the hierarchy: a hypothesis on the evolution of a genetic sex determination pathway. *Bioessays* **17**: 71–77.
- Wilkinson, G.S., and Reillo, P.R., (1994). Female choice response to artificial selection on an exaggerated male trait in a stalk-eyed fly. *Proc. R. Soc. Lond. B* **255**: 1–6.
- Wilkinson, G.S., Presgraves, D.C., and Grymes, L., (1998). Male eye span in stalk-eyed flies indicates genetic quality by meiotic drive suppression. *Nature* **391**: 276–279.
- Williamson, D.L., Sakaguchi, B., Hackett, K.J., Whitcomb, R.F., Tully, J.G., and Carle, P., et al. (1999). *Spiroplasma poulonii* sp. nov., a new species associated with male-lethality in *Drosophila willistoni*, a Neotropical species of fruit fly. *Int. J. Syst. Bacteriol.* **49**: 611–618.
- Wood, R.J., and Newton, M.E., (1991). Sex-ratio distortion caused by meiotic drive in mosquitoes. *Am. Nat.* **137**: 379–391.
- Wrensch, D.L., and Ebbert, M.A., Eds., (1993). *Evolution and Diversity of Sex Ratio in Insects and Mites*, Chapman and Hall, New York.
- Wyss, J.H., (2000). Screw-worm eradication in the Americas—overview. In: Tan, K.H., Ed., *Area-Wide Control of Fruit Flies and Other Insect Pests*. Penerbit Universiti Sains Malaysia, Pulau Pinang, pp. 79–86.
- Xie, T., and Spradling, A.C., (2000). A niche maintaining germ line stem cells in the *Drosophila* ovary. *Science* **290**: 2328–2330.

- Zchori-Fein, E., Roush, R.T., and Hunter, M.S., (1992). Male production induced by antibiotic treatment in *Encarsia formosa* (Hymenoptera: Aphelinidae), an asexual species. *Experientia* **48**: 102–105.
- Zchori-Fein, E., Gottlieb, Y., Kelly, S.E., Brown, J.K., Wilson, M.N., Karr, T.L., and Hunter, M.S., (2001). A newly discovered bacterium associated with parthenogenesis and a change in host selection behavior in parasitoid wasps. *Proc. Natl. Acad. Sci. USA* **98**: 12555–12560.

# Molecular Genetics of Insect Behavior

## Chapter Outline

- 11.1 Overview 466
- 11.2 Introduction 467
- 11.3 The Insect Nervous System 469
- 11.4 Traditional Genetic Analyses of Behavior 471
  - 11.4.1 Crossing Experiments 471
    - 11.4.1.1 Susceptibility to American Foulbrood in *Apis mellifera* 472
    - 11.4.1.2 House-Entering Behavior in *Aedes aegypti* 473
    - 11.4.1.3 Foraging in *Drosophila* 474
    - 11.4.1.4 Other Behaviors Influenced by One or a Few Genes 475
  - 11.4.2 Selection Experiments 476
    - 11.4.2.1 Migratory Behavior in *Oncopeltus fasciatus* 476
    - 11.4.2.2 Analysis of Selection Experiments 477
  - 11.4.3 Some Polygenically Determined Behaviors 478
- 11.5 Molecular-Genetic Analyses of Insect Behavior 479
  - 11.5.1 The Photoperiodic Clock 480
    - 11.5.1.1 The period<sup>+</sup> Locus of *Drosophila* 482
    - 11.5.1.2 Song-Cycle Behavior in Transgenic *Drosophila* 484
    - 11.5.1.3 Other Effects of per 485
  - 11.5.2 Learning in *Drosophila* 486
  - 11.5.3 Functional Genomics of Odor Behavior in *Drosophila* 490
  - 11.5.4 Behavior of *Apis mellifera* 491
  - 11.5.5 Pheromones in Insects 492
  - 11.5.6 Neurobiochemistry of *Drosophila* 493
    - 11.5.6.1 Electrical Signaling 493
    - 11.5.6.2 Neurotransmitters 495
    - 11.5.6.3 Ion Channels 495
  - 11.5.7 Divergent Functions of *Est-6* and *Est-5* in Two *Drosophila* Species: A Cautionary Tale of Homologs 496
  - 11.5.8 Courtship Behavior in *Drosophila* 497
  - 11.5.9 Speciation Genes in *Drosophila* and Other Insects 499
  - 11.5.10 Personality in Insects: *Tribolium confusum*, *Apis mellifera*, *Acyrtosiphon pisum*, and *Pyrrhocoris apterus* 501

11.6 Symbionts and Insect Behavior	502
11.7 Human Neurodegenerative Diseases and Addictions in <i>Drosophila</i>	503
11.8 High-Throughput Ethomics	506
11.9 Systems Genetics of Complex Traits in <i>Drosophila</i>	506
11.10 Social Behavior in Bees and Ants	507
11.11 Conclusions	508
References Cited	509

## 11.1 Overview

The study of insect behavior involves the analysis of any and all activities performed by an insect in relation to its surrounding environment. Behavior genetics is the study of the underlying hereditary basis of the behavior. For many years, Mendelian genetic analyses were conducted on a few traits determined by one or a few genes, or quantitative-genetic methods were used for traits determined by “many” genes. The sequencing of whole genomes and the use of molecular-genetic methods are revolutionizing the genetic analysis of insect behavior. However, one caveat should be recognized, based on a recent discovery about the discrepancy between behavior in the laboratory and in the field by the well-studied *Drosophila melanogaster*: the behavior you evaluate in the laboratory should be tested under conditions approaching the natural environment as much as possible or erroneous conclusions could be reached.

Circadian behaviors, mating behavior, and learning in *Drosophila* have been dissected with the tools of molecular genetics and inter- and intraspecific comparisons can be made of the DNA sequences associated with these behaviors. The circadian clock of *Drosophila* involves several genes, including *period<sup>+</sup>* (*per<sup>+</sup>*). Mutants of *period<sup>+</sup>* influence activity patterns and other circadian rhythms, as well as altering song cycles in courting males. The *per<sup>+</sup>* locus has been cloned and sequenced in *D. melanogaster* and *D. simulans*. After the *per<sup>+</sup>* gene of *D. simulans* was inserted by *P*-element-mediated transformation into a strain of *D. melanogaster* that is arrhythmic, transgenic *D. melanogaster* males produced song cycles like those of *D. simulans*. As few as four amino acids may account for the differences in song rhythm determined by the *per<sup>+</sup>* locus. The *timeless<sup>+</sup>*, *doubletime<sup>+</sup>*, *cycle<sup>+</sup>*, *cryptochrome<sup>+</sup>*, and *Clock<sup>+</sup>* genes are involved in the circadian clock. The clock involves transcription of the *per<sup>+</sup>* and *tim<sup>+</sup>* genes, followed by production of the PER and TIM proteins and subsequent negative feedback on self-transcription. Degradation of proteins then releases the negative feedback, allowing a new round of transcription, resulting in oscillations of RNA and protein.

*Drosophila* learning mutants, such as *dunce* and *couch potato*, are providing insights into the fundamental processes involved in short-term, intermediate, long-term, and anesthetic-resistant learning in insects and other organisms. Analysis of sleep behaviors in *D. melanogaster* may offer clues to sleep in humans (Shaw and Franken 2003). The discovery of “personalities” of diverse insects changes our interpretations of insect behavior. Analyses of behaviors that are determined by many genes are being revolutionized by the use of molecular-genetic methods and whole-genome sequencing projects. It now is possible to map the number and location of genes or single nucleotide polymorphisms (SNPs) affecting complex traits by correlating their inheritance with a variety of DNA markers and by conducting microarray analyses.

## 11.2 Introduction

Insect behavior covers a very wide range of activities, including locomotion, grooming, feeding, communication, reproduction, dispersal, flight, learning, migration, host or prey selection, diapause, and various responses to environmental hazards such as temperature, humidity, parasites, and toxins (Dingle 1978, Beck 1980, Dingle and Hegmann 1982, Alcock 1984, Tauber et al. 1986, Gatehouse 1989, 1997, Sokolowski 2001, Bazzett 2008, Ikeno et al. 2011). Understanding the behavior of pest and beneficial insects could allow improved pest-management programs (Foster and Harris 1997, Renou and Guerrero 2000, Bendena 2010).

One definition of **behavior** is any action that an individual carries out in response to a stimulus or its environment, especially an action that can be observed and described. However, insects also behave spontaneously, in the absence of any obvious stimulus. Thus, behavior includes studies to understand how an insect takes in information from its environment, processes that information, and acts. Processing information in the central nervous system may involve integrating information over time, including stimuli such as hormones coming from within the insect. Thus, the connection between **stimulus** and **response** can be delayed and indirect.

The genetic analysis of behavior rightfully has been perceived to be more complex than the analysis of morphological or anatomical traits (Baker et al. 2001, Sokolowski 2001, Vanin et al. 2012). One of the complications in genetic analyses of behavior is the difficulty in defining the behavior in a clear manner. Often “a behavior” may consist of multiple components, which can lead to confusion regarding the number of genes involved. Distinguishing between

behavior and physiology can be particularly difficult. The same behavior can be examined from at least four different viewpoints: 1) the immediate cause (or control), 2) its development during the individual's life span, 3) the function of the behavior, and 4) how the behavior evolved (Wyatt 1997).

Behavior genetics began to develop as a field of study in the 1960s, but was limited to demonstrating that a behavioral trait was heritable, determining whether its **mode of inheritance** was dominant or recessive, sex-linked or autosomal, and resolving whether the variation was due to single or multiple genes (Ehrman and Parsons 1973). Genetic analyses of insect behavior require careful control of environmental conditions, because even subtle differences in test conditions can influence the results of assays (Vanin et al. 2012). Objective measures of insect behavior are difficult, and considerable efforts have been devoted to devising specific and appropriate assays. The possible influence of learning always must be considered and, to complicate matters further, learning rates no doubt vary among different populations of the same species so that both heredity and environment must be considered. Furthermore, recent studies indicate that individuals within a population may vary in their behavior and have "personalities"; for example, some may be more adventurous than others. Genetic analyses of insect behavior involve, in many cases, analyses of the physiological or morphological changes that are associated with the change in behavior. Sometimes, however, behavior is changed in an insect because a morphological trait has been altered through mutation.

The genetic basis of insect behavior has been analyzed most extensively using *Drosophila melanogaster* and a few other species, such as honey bees, grasshoppers, *Nasonia* parasitoids, and crickets (Benzer 1973, Matthews and Matthews 1978, Ehrman and Parsons 1981, Hall 1984, Kalmring and Elsner 1985, Huettel 1986, Huber et al. 1989, Menzel 1999, Beukeboom and van den Assem 2001).

Molecular-genetic techniques provide powerful methods to analyze insect behavior, including olfaction, learning, circadian rhythms, and mating behavior. Having the complete sequence of the genome of *D. melanogaster* simplifies the isolation of specific genes that are involved in the behavior. *P*-element-mediated transformation makes it possible to insert genes from one species of *Drosophila* into the genome of another, and their effect(s) on behavior can be determined. Transgenic *D. melanogaster* carrying markers such as green fluorescent protein (GFP) allow scientists to determine when and where specific genes are active.

Molecular-genetic analyses of learning and memory in *Drosophila* provide a means to study one of the most challenging frontiers in neurobiology

(Waddell and Quinn 2001). Molecular-genetic methods may allow us to localize and identify some of the individual genes among the “many” involved in determining the interesting and complex behaviors exhibited by insects (Doerge 2002). Perhaps the most significant advance in the study of behavior has been the sequencing of whole genomes of insects other than *Drosophila*. This has allowed novel and detailed studies of behavior in a variety of insects and promises to provide exceptional new insights.

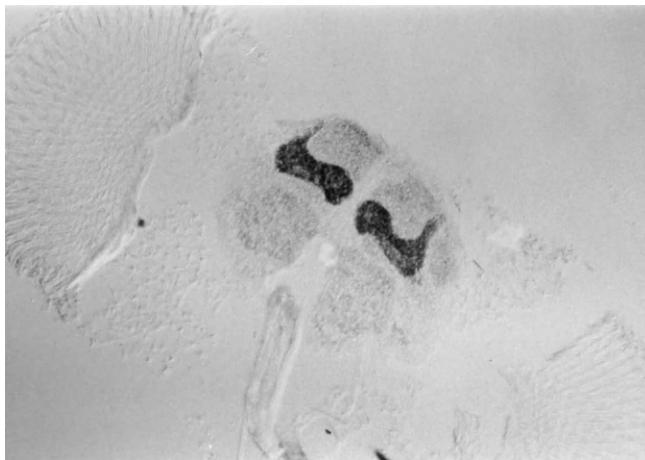
Analyses of insect behavior use techniques from several disciplines including anatomy, biochemistry, ecology, **ethology** (study of animal behavior in the natural environment), genetics, psychology, physiology, and statistics (Matthews and Matthews 1978, Hay 1985, Bell 1990, Holman et al. 1990, Via 1990, Barton Browne 1993, Heisenberg 1997, Doerge 2002). These disciplines are required because an insect perceives the environment through its sensory systems. The external sensory stimuli are transduced into electrical information, which is then processed and decoded, leading to a behavioral response. Behavior can be divided into several sequential steps: **stimulus recognition, signal transduction, integration, and response** or motor output.

### 11.3 The Insect Nervous System

The insect brain contains  $\approx 10^5$ – $10^6$  neurons. It consists of three main divisions, the **protocerebrum**, **deutocerebrum**, and **tritocerebrum**. In each division, different **neuropil** regions are located; a neuropil is a dense network of interwoven axons and dendrites of neurons and neuroglial cells in the central nervous system and parts of the peripheral nervous system.

In the protocerebrum, centers are present that are associated with vision and other sensory receptors (the mushroom bodies and central complex). The superior protocerebrum, with the pars intercerebralis, contains sets of neurosecretory cells that supply neurohemal organs in the corpora cardiaca and corpora allata, which are located in the head or prothorax in insects. The optic lobes flanking the protocerebrum contain the most well organized neuropiles in the brain.

Mushroom bodies in the brain are associated with olfactory pathways, including olfactory learning (Figure 11.1). Among the insects, mushroom bodies differ greatly in size and shape, with the number of cells ranging from 2500 in *Drosophila* to 50,000 in the cricket *Acheta*, 170,000 in the honey bee, and 200,000 in the cockroach *Periplaneta* (Heisenberg 1998, Strausfeld et al. 1998). The antennal centers are found in the deutocerebrum. In the tritocerebrum, neurosecretory neurons and neurons associated with the control of feeding and



**Figure 11.1** Photograph of the brain structures known as mushroom bodies in *D. melanogaster*. The dark areas show the mushroom bodies stained with an antibody to the *dunce*-encoded enzyme. *dunce*<sup>+</sup> encodes an enzyme called cyclic AMP phosphodiesterase. This enzyme destroys cyclic AMP, which is important in learning and memory. (Photograph provided by R. L. Davis.)

foregut activity are found (Homberg et al. 1989). The brain is connected to the subesophageal ganglion via connectives and to the thoracic and abdominal ganglia, or ventral nerve cord (Strausfeld 1976).

Behavior and development are coordinated in the insect by both nerves and neuropeptides. Both neurosecretory cells and neurons use **neuropeptides** as messengers. Many different types of neuropeptides have been identified (Scharrer 1987, Masler et al. 1993, Raina and Menn 1993, Nassel 2002, Bendena 2010). Neuropeptides range in size from three amino-acid residues (thyrotropin-releasing hormone) to >50 (insulin). They are generated from larger precursor proteins, ranging from 90 to 250 amino acids in length. Several neuropeptide genes have been cloned (Nassel 1993, 2002; Sato et al. 1993; Tillman et al. 1999; Bendena 2010).

Neuropeptides are released as cotransmitters and modulate fast transmission at neuromuscular junctions. A given neuropeptide may occur at several sites, including central nervous system circuits, peripheral synapses, and at the peripheral targets (muscles and glands). Neuropeptides regulate behavior by coordinating temporal and spatial activity of many neuronal circuits. Each of the circuits controlling behavior uses sets of sensory neurons, interneurons, and motor neurons. Thus, multiple neural networks share neural elements. Molecular-genetic analysis is providing rapid progress in understanding neuropeptide receptors and second-messenger pathways. Research on neuropeptides and their receptors indicates that they have roles during embryonic

development and as cytokines in the immune systems of insects (Nassel 1993). Neuropeptides are being studied as an alternative to chemical pesticides. However, neuropeptides offer a challenge because the insect cuticle contains a lipid component that inhibits penetration of peptides and the gut, hemolymph, and membranes of tissues contain peptidases that rapidly degrade peptides so topical application or ingestion are not likely to be effective in delivering these proteins (Bendena 2010).

## 11.4 Traditional Genetic Analyses of Behavior

Sometimes, as is demonstrated in examples below, mutations in a single gene or a few major genes alter a behavior and the mode of inheritance can be assessed by traditional methods. Traditional behavior-genetic analysis uses two main experimental approaches: **crossing** and **selection**. A third, limited to *D. melanogaster*, involves analysis of **fate maps** in genetic mosaics to locate the anatomical site of abnormalities that affect behavior (Hotta and Benzer 1972).

Many genes may influence an insect's behavior, although a specific behavior sometimes can be altered by the mutation of a single gene (Plomin 1990). When many genes are involved, analyses of behavior traditionally have required the use of **quantitative genetic** methods and sophisticated statistical analysis.

### 11.4.1 Crossing Experiments

A crossing experiment involves mating individuals that differ in a particular kind of behavior and then examining the behavior of their F<sub>1</sub> and backcross progeny. (A backcross is a cross of F<sub>1</sub> individuals to a parental line, usually the homozygous recessive one.) Ideally, the environment is controlled so that all individuals experience the same conditions. It is easiest to interpret the results of the experiment *if* the individuals that are crossed differ *only* with regard to a single behavioral attribute.

The phenotype of the F<sub>1</sub> and backcross progeny indicates whether a single gene or multiple genes determine the behavior, and whether there is dominance, sex linkage, or maternal influences. If many genes influence the trait, it is difficult to determine the number of loci, their relationship to each other, or their location on specific chromosomes because most insect species lack sufficient genetic markers. New molecular and statistical methods using quantitative trait loci (QTLs) may provide greater power to study and locate multiple and interacting loci (Doerge 2002).

One aspect of honey bee behavior provides an example of a trait that appears to be determined by a few genes. Other behaviors, including house-entering

behavior in the mosquito *Aedes aegypti*, foraging in *D. melanogaster*, and host choice in the parasitoid *Nasonia* also appear to be determined by one or a few genes.

#### 11.4.1.1 Susceptibility to American Foulbrood in *Apis mellifera*

Susceptibility to foulbrood disease, caused by *Bacillus larvae*, originally was analyzed by crossing two inbred *A. mellifera* strains with differing levels of resistance (Table 11.1). The differences in resistance were attributed to differences in “hygienic behavior” in worker (sterile female) bees and represent one of the first demonstrations that behavior is genetically determined (Rothenbuhler 1964). Resistant workers (=hygienic) consistently remove dead larvae and pupae from the brood nest at a high rate, thus slowing the spread of the bacteria through the colony by reducing contamination. Crosses between “hygienic” queens and susceptible “nonhygienic” haploid males yield  $F_1$  worker progeny that are nonhygienic, indicating that the “genes” conferring resistance are recessive.

Progeny produced by backcrosses to the homozygous-recessive hygienic strain yielded approximately 25% hygienic workers, which is consistent with the hypothesis that hygienic behavior is determined by two recessive loci (Table 11.1). Under a two-locus model, hygienic worker queens are homozygous for two genes, *uu* and *rr*. The hygienic workers (*uu, rr*) both uncap the cells (*uu*) containing dead brood **and** remove them (*rr*). However, *uu, r<sup>+</sup>r* individuals will uncap the cells, but not remove dead brood. The *u<sup>+</sup>u, rr* individuals do not uncap brood, but will remove them if the cells are uncapped for them. Individuals that are *u<sup>+</sup>u, r<sup>+</sup>r* are unhygienic, and will neither uncap nor remove

**Table 11.1: Crossing Experiment to Explain Differences in Nest-Cleaning Behavior Among Inbred Lines,  $F_1$  and Backcross Progeny of the Hapl-Diploid Honey Bee *Apis mellifera* Supports the Two-Locus, Two-Allele Model.**

Parental lines	Hygienic (diploid) <i>uu, rr</i> queen	X	Unhygienic <i>u<sup>+</sup>, r<sup>+</sup></i> (haploid) male		
		↓			
$F_1$ progeny	All diploid sterile workers				
	<i>u<sup>+</sup>u, r<sup>+</sup>r</i> (Unhygienic workers)				
Ratio of backcross progeny (workers)					
Cross of <i>u<sup>+</sup>u, r<sup>+</sup>r</i> queen X hygienic <i>u, r</i> male	→	1: <i>uu, rr</i> Hygienic	1: <i>uu, r<sup>+</sup>r</i> Uncaps, doesn't remove	1: <i>u<sup>+</sup>u, rr</i> Removes, doesn't uncaps	1: <i>u<sup>+</sup>u, r<sup>+</sup>r</i> Unhygienic

Data from Rothenbuhler (1964).

brood. Hygienic behavior appears to be a general response to remove pathogens and parasites from the nest (Spivak and Gilliam 1993).

Rothenbuhler's research on hygienic behavior became a classic in textbooks on behavioral genetics because it was one of the first examples that demonstrated that behavior is inherited (Rothenbuhler 1964). More recently, Moritz (1988) proposed a three-locus model for hygienic behavior in bees. Research on hygienic behavior continues because such behavior is important in managing bees. It is clear that the expression of hygienic behavior depends on colony strength and composition of worker types within the colony (Spivak and Gilliam 1993, Arathi et al. 2000). Electro-antennogram analyses of the olfactory and behavioral responses of hygienic and nonhygienic bees to diseased brood indicated that hygienic bees have a higher sensitivity to low concentrations of the odor of diseased bee pupae (Masterman et al. 2001). Such differences are due to a lower stimulus threshold and are not a direct result of age or experience of the bee. Thus, nonhygienic bees may be unable to detect diseased brood. Understanding hygienic behavior in *A. mellifera* has resulted in practical recommendations to beekeepers for selecting colonies resistant to chalkbrood (a fungal disease) and the pest bee mite Varroa. So far, no negative effects have been found associated with hygienic colonies and such colonies produce as much honey as nonhygienic ones (Moritz 1994, Spivak and Gilliam 1998).

#### 11.4.1.2 House-Entering Behavior in *Aedes aegypti*

House-entering behavior by the mosquito *Ae. aegypti* from East Africa was analyzed by crossing different populations with different behaviors (Trpis and Hausermann 1978). One population of *Ae. aegypti* commonly entered houses (domesticated or D), whereas others rarely did so (either peridomestic, P, or feral, F). House-entering behavior is important in determining whether a population transmits yellow fever to humans.

Three populations of *Ae. aegypti* collected either inside houses (D), near a village (P), or from tree holes in a forest (F) were bred in insectaries and crossed to produce hybrid (DP, PD, DF, FD, PF, and FP) populations (Trpis and Hausermann 1978). The original and hybrid populations were then marked with different colored fluorescent powders and released near houses. Marked mosquitoes were captured inside houses and in the village area. Of the mosquitoes entering houses, 45% were from the D population, 13.9 % from hybrids between the domestic and peridomestic population (DP and PD), 9.8% from the P population, and 5.7% were hybrids (DF and FD). Only 1.5 and 0.6% of the PF and FP hybrids were collected in the house, and the feral population entered the house with a frequency of only 0.6%. The recapture rates in the village area

were in the reverse order. The data indicate a few genes with additive effects determine the behavior.

"Domesticity" in *Ae. aegypti* is a complex phenomenon that includes a variety of behaviors, including a preference for ovipositing in man-made containers, the ability of larvae to develop in drinking water stored in clay pots with a low nutritional content, and preferences for feeding on man (rather than birds) inside houses, as well as resting and mating indoors. It is likely that *Ae. aegypti* speciated long before man began to build houses, but *Ae. aegypti* has adapted rapidly to human habitats, and the domestic form of *Ae. aegypti* is the only one known that is entirely dependent on humans ([Trpis and Hausermann 1978](#)).

#### 11.4.1.3 Foraging in *Drosophila*

*Drosophila melanogaster* larvae feed on yeast growing on decaying fruit. Naturally occurring populations contain individuals that vary in the distance the larvae travel while foraging for food, a difference attributed to a single *foraging* gene ([Osborne et al. 1997](#), [Sokolowski et al. 1997](#)). Natural populations comprise approximately 70% "rovers" (who forage long distances) and 30% "sitters" (short-distance foragers), with the rover phenotype dominant to sitter, indicating a single-gene mode of inheritance ([de Belle and Sokolowski 1987](#), [Sokolowski 2001](#)).

Sitter larvae grow at a normal rate and are of normal size. Both sitters and rovers are maintained in the field by natural selection; density-dependent selection can shift allele frequencies so that rovers are selected for in crowded larval environments and sitters in less-crowded ones. The *foraging* gene encodes a cyclic guanosine monophosphate (cGMP)-dependent protein kinase, and rovers have higher kinase activity levels than sitters. Subtle differences in this kinase lead to naturally occurring variation in behavior ([Shaver et al. 1998](#)). Sleep deprivation and starvation studies with *D. melanogaster* individuals with natural variants of the *foraging* gene suggest one of the reasons *foraging* gene polymorphisms persist in wild populations is due to fitness tradeoffs ([Donlea et al. 2012](#)). The *foraging* gene influences sleep, learning and memory, as well as feeding behavior. Flies with the "rover" genotype have a better short-term memory than "sitters," but sitter flies have better long-term memory. Rover flies do not lose their short-term memory if well fed, but do so if starved overnight. Sitter flies have reduced short-term memory after sleep deprivation, but are able to learn after 12 hours of starvation. Thus, the *foraging* gene alleles appear to provide benefits in some environments, but not others. [Donlea et al. \(2012\)](#) conclude that "understanding how these tradeoffs confer resilience or vulnerability to specific environmental challenges may provide additional clues

as to why an evolutionary alternative to sleep has not emerged," although it is not clear how the ability to survive sleep loss can confer a fitness advantage. Another gene, *Chaser*, affects larval foraging by increasing foraging path length (Pereira et al. 1995).

#### 11.4.1.4 Other Behaviors Influenced by One or a Few Genes

Crossing experiments indicate that one or a few genes influence a specific behavior in the flour moth *Epehestia kuhniella* (silk-mat spinning by larvae before pupation), the mosquito *Aedes atropalpus* (egg maturation without an exogenous source of protein such as blood), and the parasitoid wasp *Habrobracon juglandis* (flightlessness) (Ehrman and Parsons 1981). In *Bombyx mori*, females with the *piled egg* gene deposit eggs in a peculiar manner; *B. mori* larvae with the *Non-preference* gene are unable to discriminate mulberry leaves from others (Tazima et al. 1975), and Huettel and Bush (1972) found that when two monophagous tephritid flies (*Procecidochares*) were crossed, the host-preference behavior segregated in a manner consistent with control by a single locus. Desjardins et al. (2010) found that crosses in the laboratory between the parasitoids *Nasonia vitripennis* and *N. giraulti* resulted in a change in host-preference behavior. This behavior was found to be dominant and was localized to 16 Mb of sequence on chromosome 4.

A variety of behavioral mutants determined by single-major genes have been identified in *D. melanogaster* (Grossfield 1975, Hall 1985, Pavlidis et al. 1994), including a group of sex-linked, incompletely dominant mutants (*Shaker*, *Hyperkinetic*, and *eag*) that are expressed when the flies are anesthetized with ether. The sex-linked temperature-sensitive recessive mutant *para<sup>ts</sup>* causes *D. melanogaster* to become immobile above 29°C. Mutants of the *couch potato* (*cpo*) locus cause flies to be hypoactive and exhibit abnormal geotaxis (response to gravity), phototaxis (response to light), and flight behavior. This gene is unusually complex, spanning >100kb and encoding three different messages (Bellen et al. 1992).

Many "single-gene" mutants that affect the morphology of *D. melanogaster* affect behavior because they are unable to perform the reaction to a stimulus due to altered effector structures. Other mutants exhibit altered behavior because perception of cues is impaired. For example, flies with *white* eyes may exhibit abnormal courtship behaviors (Grossfield 1975).

Pheromone communication in the European corn borer is genetically determined (Klun and Maini 1979, Klun and Huettel 1988, Löfstedt et al. 1989, Löfstedt 1990). Females of the E- and Z-strains of *Ostrinia nubilalis* produce different enantiomeric ratios of sex pheromone. Hybrids between these two strains

produce an intermediate pheromone blend. Analysis of the F<sub>2</sub> and backcross progeny indicates pheromone type is controlled by two alleles at a single autosomal locus, although one or more modifier genes control the precise ratio of the isomers in heterozygous females. Males of the two *O. nubilalis* strains are attracted to the appropriate pheromone blends in the field, and hybrid males respond preferentially to the pheromone produced by heterozygous females rather than to the pheromones produced by the two parental female types. A single-sex-linked gene with two alleles determines the response of males to the pheromone. The olfactory sensillae of the two types of males are different, which is controlled by an autosomal locus with two alleles. Hybrid males give intermediate results when tested for their electrophysiological responses, with E- and Z-cells yielding approximately equal spike amplitudes. The genes determining variation in pheromone production and organization of male olfactory sensillae are not closely linked and are probably on different chromosomes (Löfstedt 1990). [Yasukochi et al. \(2011\)](#) found sex-linked pheromone receptor genes of *Ostrinia nubilalis* were present in tandem array on the Z chromosome. In addition, an autosomal locus responsible for male response to the sex pheromone contains at least four odorant-receptor genes, suggesting that additional copies of odorant-receptor genes can increase the potential for male moths to acquire altered specificity for pheromone components and facilitate differentiation of sex pheromones ([Yasukochi et al. 2011](#)).

Sexual isolation between the moths *Heliothis subflexa* and *H. virescens* was determined to be due to a single quantitative trait locus (QTL) that consists of at least four odorant receptor genes ([Gould et al. 2010](#)).

#### 11.4.2 Selection Experiments

Selection experiments provide another traditional method to determine the degree to which a given behavior is determined genetically. In a selection experiment, individuals with a specific behavioral attribute are allowed to reproduce and this process is repeated over succeeding generations. Eventually, the behavior of the selected population is altered *if* genetic variation for the attribute is present in the initial colony *and* the selection procedures were appropriate. The response of the population to selection can be analyzed to estimate the heritability of the trait.

##### 11.4.2.1 Migratory Behavior in *Oncopeltus fasciatus*

Migratory behavior of *Oncopeltus fasciatus* is under genetic control ([Palmer and Dingle 1989](#)). Strains of *O. fasciatus* were selected for wing length and propensity to fly. Bidirectional selection on wing length (selection for increased and

decreased wing length) was performed for 13 generations, and the flight behavior of individuals monitored. Individuals also were selected for flight time, and those whose flight times totaled 30 minutes were considered "fliers," whereas those with a shorter flight time were labeled "nonfliers." Response to selection on wing length was rapid, and flight tests of the long- and short-winged insects indicated there was a positive correlation between wing length and flight duration. Selection after two generations for flight or nonflight likewise resulted in divergent responses, indicating a large genetic component to flight behavior.

#### 11.4.2.2 Analysis of Selection Experiments

To estimate the degree of genetic influence on a specific behavior, two measures are used, the selection differential and the estimate of heritability. The **response to selection (R)** is the difference in mean phenotypic value between the offspring of the selected parents and the mean phenotypic value of the entire parental generation before selection (Falconer 1989).

$$R = h^2 S$$

R is the improvement or response to selection,  $h^2$  is the heritability of the characteristic under selection in the population, and S is the selection differential. The **selection differential (S)** is the average superiority of the selected parents expressed as a deviation from the population mean (Falconer 1989). The selection differential measures the difference between the average value of a quantitative character in the whole population and the average value of those selected to be parents of the next generation. It is measured in standard deviation units.

**Heritability in the broad sense** is the degree to which a trait is genetically determined. Because both genes and environment influence behavioral traits, heritability is expressed as the ratio of the total genetic variance to the phenotypic variance ( $V_G/V_P$ ). Heritability *in the narrow sense* is the degree to which a trait is transmitted from parents to offspring, and is expressed as the ratio of the additive genetic variance to the total phenotypic variance ( $V_A/V_P$ ) (Falconer 1989).

Heritability could be estimated to be "zero" if the specific population being selected had no variability for the behavioral attribute under study because it was inbred. Heritability could be estimated to be "one" if the trait was completely determined by genes, and the environment had little effect on the phenotype, although this would be an unusual outcome.

Heritability estimates provide no information about the actual mode of inheritance of a quantitative trait because they represent the cumulative effect of all

loci affecting the trait. The number of loci involved generally can be determined only with elaborate and specially designed experiments. Several assumptions are made when estimating heritability: 1) all loci affecting the trait act *independently* of one another and 2) the loci are *unlinked* (located on different chromosomes). Another assumption 3) is that environment affects all genotypes in a similar manner. These three assumptions are not always justified. Thus, heritability estimates are difficult to interpret, although they are useful for predicting response to selection under specific environmental conditions.

Heritability estimates usually are made by regression-correlation analyses of close relatives (parent-offspring, full sibs, half sibs), experiments involving response to selection, or analysis-of-variance components. Traits with high heritability respond readily to selection with an appropriate selection method. The magnitude of the response to the selection, that is the differences in mean values between parent and progeny generations, provides an estimate of heritability in the narrow sense ( $h_n^2$ ). This estimate is valid only for the population being examined, under the test conditions used, for the behavior observed, and for the method of measurement used.

Heritability of most insect behaviors is relatively high, probably because many arthropod behaviors are highly stereotyped (Ehrman and Parsons 1981). For example, the heritability of locomotor activity of *D. melanogaster* has been estimated to be 0.51, and heritability of mating speed of male *D. melanogaster* has been estimated to be 0.33. Heritability for honey production from honey bees ranged from 0.23 to 0.75, depending upon the experimental conditions and colonies tested (Rinderer and Collins 1986). Italian honey bees are less able to remove the parasitic mite Varroa than Africanized bees, and the heritability of this ability was estimated to be 0.71 (Moretto et al. 1993). Heritability of the length of the prereproductive period in *Helicoverpa armigera*, which is when migratory flight occurs in this noctuid moth, ranged from 0.54 to 0.16 (Colvin and Gatehouse 1993). Heritability of host selection behavior by *Asobara tabida*, a parasitoid of *Drosophila subobscura*, ranged from 0.03 to 1.0 depending upon the test method used (Mollema 1991), illustrating that test conditions are crucial to behavior analysis.

#### **11.4.3 Some Polygenically Determined Behaviors**

Behavior often is a continuous variable, controlled by multiple genes with small additive effects (Plomin 1990, Heisenberg 1997). The task of teasing apart the respective roles of genes and environment requires statistical analysis (Doerge 2002). *Drosophila* behaviors determined by multiple genes include locomotor activity, chemotaxis, duration of copulation, geotaxis, host-plant preference, mating speed, phototaxis, preening, and the level of sexual isolation within and

between species. Multiple genes influence host-plant adaptation and host preference in insects, and learning also may affect host preference (Papaj and Prokopy 1989, Via 1990). Host-plant choice usually is a hierarchy of several components. For example, attraction to a site from a distance and oviposition-site preference (egg laying at the site) are genetically distinct in *Drosophila tripunctata* (Jaenike 1986).

The genetic basis of host-plant specialization in the fruit flies *Drosophila sechellia* and *D. simulans* is determined by a minimum of three or four loci that affect egg production, survival, and host preference (R'Kha et al. 1991). *Drosophila sechellia* breeds in a single plant, *Morinda citrifolia*, which is toxic to other *Drosophila* species. Its sympatric relative, *D. simulans*, breeds on a variety of plants. The two species can be crossed, and the F<sub>1</sub> hybrid embryos produced by *D. simulans* females are susceptible to *Morinda* fruit because susceptibility is maternally inherited and fully dominant. Females of *D. sechellia* are stimulated by *Morinda* to produce eggs, but this plant inhibits oviposition in *D. simulans*. In hybrid progeny, the inhibition observed in *D. simulans* is dominant. F<sub>1</sub> hybrids and backcross progeny exhibit intermediate, approximately additive, behavior. These differences result in isolation of the two species in nature, although their ranges overlap geographically. Thus, their ecological niches are determined by tolerance to toxic products in the ripe *Morinda* fruit, with *D. sechellia* exhibiting a strong preference for *Morinda*, an ability to detect fragrant volatiles from *Morinda* over a long distance, and a stimulation of egg production by *Morinda*. By contrast, *Morinda* inhibits egg production in *D. simulans*.

Other specific behavioral attributes inherited in a complex manner include number of attempts to mate by males (*Musca domestica*), high and low ability to learn to extend the proboscis to a stimulus applied to the forelegs (*Phormia regina*), call rhythm of males and female response to calling songs (hybrid crickets), ability to avoid pesticides (*Anopheles albimanus*), high and low collection of alfalfa pollen and stinging behavior by bees (*Apis mellifera*) (Ehrman and Parsons 1981, Hall 1985, Rinderer 1986). Gould (1986) found that the propensity for cannibalism by larvae of *Heliothis virescens* is polygenically determined. Most of these behaviors were analyzed by selection experiments.

## 11.5 Molecular-Genetic Analyses of Insect Behavior

Molecular-genetic analyses are providing significant advances in our knowledge of behavior (Plomin 1990, Sokolowski 2001, Doerge 2002). Molecular-genetic methods are unlikely to replace traditional methods of behavior analysis, but the ability to identify, clone, and sequence specific genes makes it easier to understand several behaviors, including the periodicity of biological rhythms, mating behavior, locomotion, and learning. It is possible to clone a gene from

one *Drosophila* species, insert it into a *P*-element vector, and introduce the exogenous gene into mutant strains of *D. melanogaster* to confirm that the putative gene does, in fact, code for the behavior of interest. Cloned genes from *Drosophila* can, in some cases, be used as probes to identify genes from other arthropods, which then can be sequenced and compared. The availability of the complete genome of *D. melanogaster*, and of other insects, allows analyses of behavior that could not be conducted previously.

### 11.5.1 The Photoperiodic Clock

The potential that molecular genetics offers was first exemplified by analyses conducted using the *period<sup>+</sup>*, and other clock genes, of *D. melanogaster* (Table 11.2). Most insects exhibit particular behaviors at a specific time of the day, which are due to the action of a **circadian clock** that allows the insect to

**Table 11.2: Some Genes Involved in the Circadian Clock of *Drosophila melanogaster*.**

Gene (abbreviation)	Mutant phenotype(s)	Function(s)
<i>period<sup>+</sup></i> ( <i>per<sup>+</sup></i> )	Short-period, long-period and arrhythmic flies Affects locomotion, eclosion, courtship rhythms	Negative transcription element; <i>per<sup>+</sup></i> mRNA levels rise late in the day Activated by the heterodimer of the CLOCK and CYCLE proteins PER proteins feed back negatively on their own transcription PER contains a protein dimerization domain called PAS
<i>timeless<sup>+</sup></i> ( <i>tim<sup>+</sup></i> )	Short-period, long-period, and arrhythmic flies Affects locomotion, eclosion, sleep	Negative element; <i>tim<sup>+</sup></i> mRNA levels rise late in the day TIM protein destabilized by light TIM proteins feed back negatively on their own transcription, interacts with PER
<i>Clock<sup>+</sup></i> ( <i>Clk<sup>+</sup></i> )	Arrhythmic Affects locomotion, eclosion, rhythm	CLK, in combination with CYC, activates transcription of <i>per<sup>+</sup></i> and <i>tim<sup>+</sup></i> CLK negatively regulates itself Mutants blind for “lights-on” response
<i>cycle<sup>+</sup></i> ( <i>cyc<sup>+</sup></i> )	Arrhythmic Affects locomotion and eclosion	CYC, in combination with CLK, activates transcriptions of <i>per<sup>+</sup></i> and <i>tim<sup>+</sup></i> CYC negatively regulates itself Mutants respond poorly to light-dark cycles
<i>cryptochrome<sup>+</sup></i> ( <i>cry<sup>+</sup></i> )	Photoreceptor Resets rhythms	Sequence homologous to a photolyase; binds TIM in a light-dependent manner Altered light response in mutants
<i>doubletime<sup>+</sup></i> ( <i>d<sup>+</sup></i> )	Lengthens cycle in constant darkness Affects locomotion, sleep	Casein kinase I involved in phosphorylating PER, rendering it unstable in absence of TIM

(Adapted from Emery et al. 2000, Sancar 2000, Harmer et al. 2001, Merrow and Roenneberg 2001, Sokolowski 2001, Williams and Sehgal 2001.)

measure time (Kyriacou 1990, Takahashi 1992). Such circadian rhythms have several characteristics:

1. The clocks that regulate such behavior usually are “free running” in constant environments and are not simple responses to changes in light or temperature.
2. Although the rhythms are free running, an initial environmental signal is required to start the clock. Among the cues that “set” the clock are alternating light and dark cycles, high and low temperature cycles, or short pulses of light.
3. The circadian rhythm is relatively insensitive to changes in temperature (temperature compensated).
4. Altering the cues that entrain the clock can reset the clock.

*Drosophila melanogaster* reared in constant darkness exhibit circadian locomotor activity rhythms as adults. However, the rhythms of the individual flies in these populations are not synchronized with one another (Sehgal et al. 1992). Rhythms can be synchronized if dark-reared flies are exposed to light treatments as first-instar larvae. Light treatments occurring before hatching of the first-instar larvae fail to synchronize adult locomotor activity rhythms, indicating the clock functions continuously from the time larvae hatch until they reach adulthood. The rhythm can be advanced, delayed, or unchanged, depending on the phase of the cycle at which the cue is given.

The circadian rhythm has an approximate periodicity of 24 hours. Molecular-genetic analyses of *Drosophila* clock mutants provided fundamental understanding of the mechanisms of the circadian clock. Rapid advances have been made in understanding the molecular aspects of circadian clocks in a variety of organisms (Hogenesch and Ueda 2011). Circadian rhythms are found in all organisms and probably evolved early. Common genetic elements are present in *Drosophila*, *Neurospora*, mammals, and cyanobacteria (Lakin-Thomas 2000, Loudon et al. 2000, Allada et al. 2001, Harmer et al. 2001, Merrow and Roenneberg 2001, Williams and Sehgal 2001). In *Drosophila*, the genes *period*<sup>+</sup>, *timeless*<sup>+</sup>, *Clock*<sup>+</sup>, *cycle*<sup>+</sup>, *doubletime*<sup>+</sup>, and *cryptochrome*<sup>+</sup> are involved in the circadian clock (Lakin-Thomas 2000, Emery et al. 2000, Table 11.2).

Numerous reviews have compared the molecular, genetic, and neurological components of biological rhythms, reflecting the excitement of the scientific community in understanding the molecular basis of this complex behavior (Hall 1995, 1998a; Iwasaki and Thomas 1997; Dunlap 1998, 1999; Young 1998; Ishida et al. 1999; Giebultowicz 2000; Lakin-Thomas 2000; Wager-Smith and Kay 2000; Allada et al. 2001; Harmer et al. 2001; Williams and Sehgal 2001). The numerous reviews

are nearly overwhelming, and Hall (1998b) questioned how it is possible “to review an over-reviewed subject--one whose reviews have even been reviewed.”

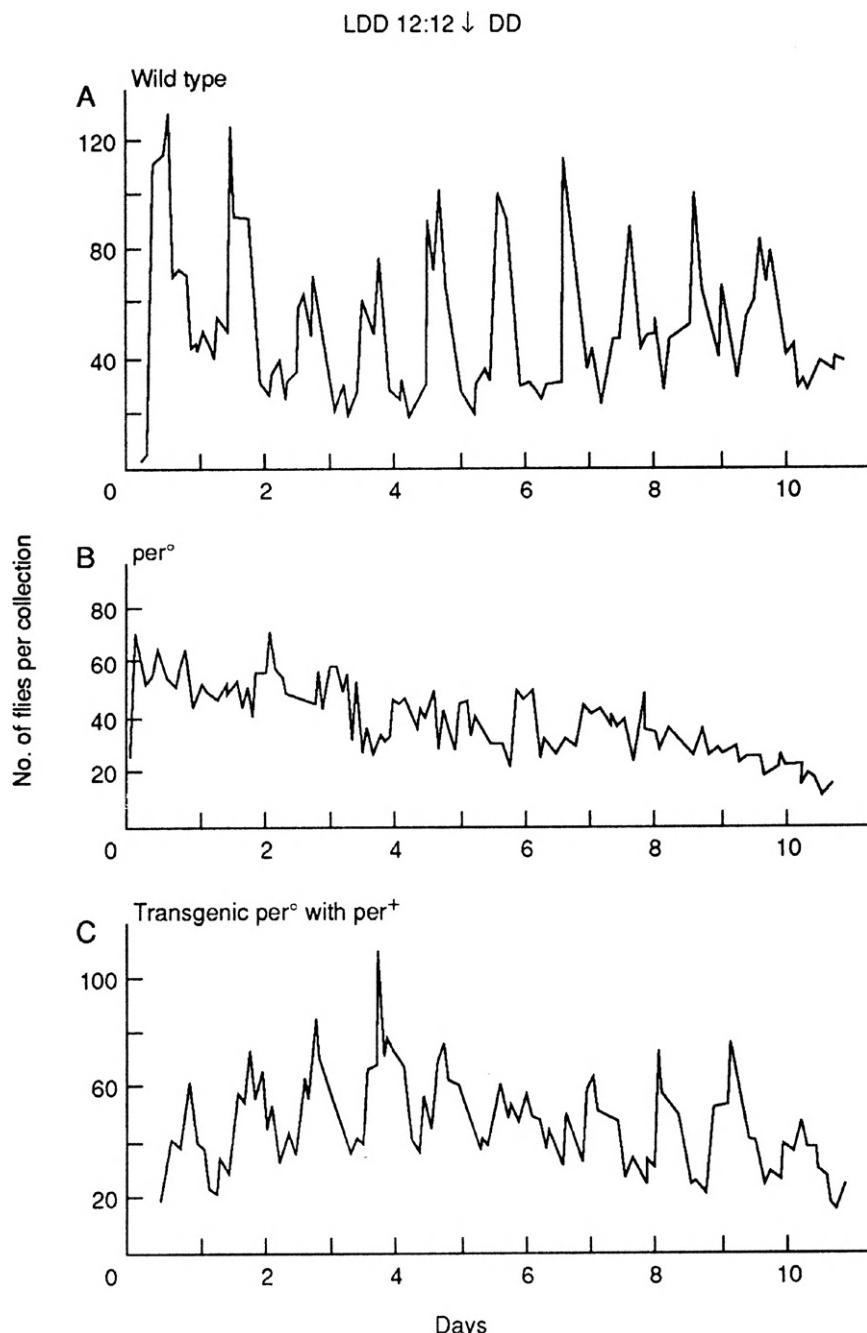
#### 11.5.1.1 *The period<sup>+</sup> Locus of Drosophila*

The *Drosophila period<sup>+</sup>* locus (*per<sup>+</sup>*) is on the X chromosome and mutations of it influence eclosion, locomotor activity, and the length of the interpulse interval of the courtship song (Table 11.2). Eclosion of wild-type flies (emergence of adults from the pupal case) typically occurs around dawn, when the presence of dew and high relative humidity increases their survival rate (Figure 11.2A). Locomotor activity then decreases during midday and is followed by increased activity again in the evening. Three classes of mutant alleles exist; they shorten (*per<sup>s</sup>* mutants have 19-hour eclosion rhythms), lengthen (*per<sup>l</sup>* mutants have 29-hour eclosion rhythms), or abolish circadian eclosion and locomotor activity rhythms (*per<sup>0</sup>* mutants). Flies with the *per<sup>0</sup>* mutation eclose arrhythmically, but periodicity in eclosion can be restored by *P* element-mediated transformation of arrhythmic flies using the wild-type *per<sup>+</sup>* allele (Bargiello et al. 1984, Figures 11.2B, 11.2C).

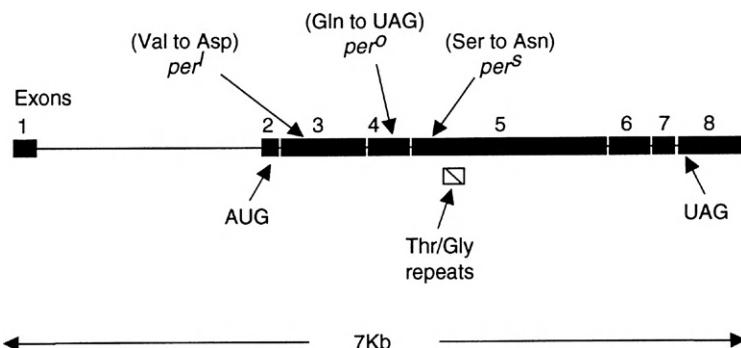
The *per<sup>+</sup>* gene is  $\approx$ 7 kb long, encodes a 4.5-kb transcript with eight exons, the first of which is noncoding (Figure 11.3). One of the most striking features of the protein is a series of threonine-glycine (Thr-Gly) repeats in the middle of the gene (Costa et al. 1992, Guantieri et al. 1999). The region encoding the Thr-Gly repeat is polymorphic in length within and between *Drosophila* species and plays a role in the thermal stability of the circadian phenotype. For example, 17, 20, or 23 repeats are found in *D. melanogaster* populations, and a clinal pattern occurs along a north-south axis in Europe and North Africa, with the shorter sequences in southern Europe (Costa et al. 1992). Costa et al. (1992) suggested that the length polymorphism cline is maintained by natural selection under different temperature conditions.

A large number of tissues express the *per<sup>+</sup>* product, including embryonic, pupal, and adult nervous systems, as well as the esophagus, gut, and ovaries. Liu et al. (1992), demonstrated that the *per<sup>+</sup>* gene product (PER protein) is predominantly found in cell nuclei in adult *Drosophila*, and Hardin et al. (1992) showed that *per<sup>+</sup>* mRNA levels undergo daily fluctuations, which constitutes a **feedback loop** in which the PER protein affects the oscillations of its own mRNA. The fluctuation in *per<sup>+</sup>* mRNA is due to fluctuations in gene transcription because the *per<sup>+</sup>* mRNA has a relatively short half-life (Zerr et al. 1990), which is consistent with the hypothesis that PER acts as a transcription factor (Table 11.2).

The *per<sup>+</sup>* genes from *D. simulans*, *D. virilis*, *D. pseudoobscura*, and *D. yakuba* have been cloned and sequenced. Parts of the gene are conserved and parts are highly diverged, which suggests that conserved regions may encode basic



**Figure 11.2** Profiles of eclosing (emergence of adults from pupal cases) for populations of A) *D. melanogaster* wild-type females, B)  $per^{\circ}$  males and females, C) and transgenic  $per^{\circ}$  individuals that have received a wild-type  $per^+$  gene by *P*-element-mediated transformation. LDD 12:12 indicates there is a 12-hour light-dark cycle. (Modified from Bargiello et al. 1984.)



**Figure 11.3** The exon/intron structure of the *D. melanogaster* *period* gene. The gene is  $\approx 7$  kb long, with seven exons. The location of the *per<sup>L</sup>*, *per<sup>O</sup>*, and *per<sup>S</sup>* mutations are indicated, as is the region which codes for the variable number (17, 20, or 23) of Thry/Gly repeats.

functions common to all (clock-type functions), whereas species-specific differences, such as love songs, locomotor activity, and eclosion profiles, may be encoded within the variable regions (Kyriacou 1990).

*Clock<sup>+</sup>*, *timeless<sup>+</sup>*, *cycle<sup>+</sup>*, and *doubletime<sup>+</sup>* are components of the circadian clock (Kyriacou 1993, Table 11.2). In addition, autosomal mutations induce flies to eclose early in a light–dark cycle; flies with mutations of *phase-angle<sup>+</sup>* emerge in the pre-dawn part of the cycle instead of just after dawn, while flies with mutations of *gate<sup>+</sup>* fail to eclose during this narrow time window.

The *cryptochrome<sup>+</sup>* (*cry<sup>+</sup>*) gene is an important clock gene because it encodes a critical circadian photoreceptor in *Drosophila* (Egan et al. 1999, Emery et al. 2000). The gene product CRY belongs to a family of blue light-sensitive proteins that includes photolyases and plant blue-light photoreceptors. Flies overexpressing CRY are hypersensitive to light. The CRY protein is probably the only dedicated circadian photoreceptor in *Drosophila* (Emery et al. 2000).

#### 11.5.1.2 Song-Cycle Behavior in Transgenic Drosophila

The courtship song is produced when *D. melanogaster* males vibrate their wings. The song consists of two components: 1) courtship hums, and 2) a series of pulses with interpulse intervals that can fluctuate between 15 and 85 milliseconds (ms) (Kyriacou and Hall 1989). The variation in interpulse intervals ranges from a period of 56 ms in *D. melanogaster* and 35–40 ms in *D. simulans*. The males of *D. melanogaster* with the *per<sup>S</sup>* mutation sing with 40-ms periods, *per<sup>L</sup>* males sing with 76-ms periods, and *per<sup>O</sup>* males are arrhythmic.

The genetic basis of species-specific song instructions was confirmed by the transfer of the *per<sup>+</sup>* gene cloned from *D. simulans* into *D. melanogaster* via

*P*-element-mediated transformation (Wheeler et al. 1991). The *D. simulans* *per*<sup>+</sup> gene restored a rhythm in *D. melanogaster* and transgenic *D. melanogaster* males produced song cycles characteristic of *D. simulans* males. Wheeler et al. (1991) concluded that substitutions in four or fewer amino acids in the *per*<sup>+</sup> locus are responsible for the species-specific courtship behavior.

#### 11.5.1.3 Other Effects of *per*

*per* alleles affect locomotion, cellular rhythms, and development time. Flies with *per*<sup>s</sup> develop faster than wild-type flies, and *per*<sup>l</sup> flies develop more slowly than the wild type (Kyriacou 1990, 1993). It has long been thought that circadian oscillations provided the clock for photoperiodically induced diapause in insects (Takeda and Skopik 1997). Diapause is a genetically determined state of arrested development that is induced before the onset of detrimental conditions. Hibernal diapause, which allows insects to survive overwinter, can be induced when insects develop during a period of cool temperatures under a short daylength, which means they must be able to measure light and dark cycles. However, *per*<sup>+</sup> appears to have no influence on the photoperiodic clock in *D. melanogaster* (Saunders 1990). Females of a wild-type strain of *D. melanogaster* (Canton-S) and strains with *per* mutations were able to discriminate between diapause-inducing short days and noninductive daylengths. Adult females of *D. melanogaster* exhibit an ovarian diapause when reared and held under short days and low temperature (12°C). Females exposed to long days at the same temperature reproduce. The critical daylength (the photoperiod at which 50% of the individuals enter diapause) for Canton-S females at 12°C is ≈14 hours of light per 24 hours. Photoperiodic response curves for the *per*<sup>s</sup>, *per*<sup>l</sup>, and Canton-S strains were almost identical, although *per*<sup>0</sup> flies showed shortened critical daylengths. However, *per*<sup>0</sup> females are able to discriminate between a long and a short day.

Ikeno et al. (2011), however, found that the circadian clock genes *period*<sup>+</sup> and *cycle*<sup>+</sup> regulate reproductive diapause in males of the bean bug *Riptortus pedestris* (Ikeno et al. 2011). RNA interference (RNAi) analyses showed that disruption of *per*<sup>+</sup> and *cycle*<sup>+</sup> disrupted the rhythm of cuticle deposition. RNAi of *per*<sup>+</sup> also induced development of male reproductive organs even under diapause-inducing short-day conditions. RNAi of *cycle*<sup>+</sup> suppressed development of the reproductive organs even under diapause-preventing long-day conditions. Larval diapause in *Chymomyza costata* also is affected by mutations affecting the photoperiodic clock (Kostal and Shimada 2001).

Many behaviors, including learning, involve temporally patterned events. The interval between presentation of the conditioned stimulus and reinforcement is

important in associative learning. The conditioned stimulus must be presented before the unconditioned stimulus, and the unconditioned stimulus must follow the conditioned within a relatively brief interval. It was thought that the *per<sup>+</sup>* gene could be involved in learning, based on the observation that males with the *per<sup>-</sup>* allele in one experiment did not exhibit normal experience-dependent courtship behavior. However, males with the wild type *per<sup>+</sup>* or the *per<sup>s</sup>* and *per<sup>d</sup>* alleles could be conditioned normally (Gailey et al. 1991).

Although the above-described analyses of *Drosophila* circadian behavior have informed us about the genetics of circadian rhythms, Vanin et al. (2012) discovered that the behavior of flies under natural conditions is different from that observed in artificial laboratory conditions with constant temperatures and light-dark cycles without dawn- or dusk-lighting regimes. Vanin et al. (2012) recommended that studies of circadian rhythms be conducted in the future using more-natural temperature and light regimes in the laboratory. Under artificial laboratory conditions, where lights were either on or off, flies anticipated lights turning on by increasing their locomotor activity about two hours before that event. However, under field conditions, the flies did not behave in this manner. In fact, their sleep-wake cycle was quite different when temperature, sunlight, moonlight, and humidity varied. Instead of sleeping during the middle of the day, as they did in the laboratory, they became active. Instead of anticipating dawn, they reacted to changing light during twilight, but did not respond to dawn. In addition, moonlight had no effect on behavior under field conditions, although it does under artificial laboratory conditions.

Sandrelli et al. (2008) note that the mammalian genes *Clock* and *Period* share a common evolutionary origin with that of insects, although the mammalian-clock components consist of multiple gene copies, increasing complexity and redundancy. They also reviewed what has been learned about circadian clock genes in insects (Diptera, Lepidoptera, and Orthoptera) other than *Drosophila*. Sandrelli et al. (2008) note that peripheral clocks are common in insects and can be entrained by light in extraretinal photoreceptors, but these have not yet been studied.

### 11.5.2 Learning in *Drosophila*

It is difficult to produce a single definition of learning (Meller and Davis 1996, Tully 1996, Waddell and Quinn 2001). Learning can be defined as a change in behavior with experience, but this definition would not exclude responses such as growth and maturation, or other processes that are triggered by events such as mating or feeding. Another definition is a reversible change in behavior with

experience, but this excludes phenomena in which the modification caused by some experience is fixed and resistant to further change. Another definition is that learning is a more-or-less permanent change in behavior that occurs as a result of practice, but this definition is ambiguous (Papaj and Prokopy 1989).

Papaj and Prokopy (1989) suggested the following properties are characteristic of learning in insects: 1) An individual's behavior changes in a repeatable way as a consequence of experience. 2) Behavior changes gradually with continued experience, often following a "learning curve" to an asymptote. 3) The change in behavior accompanying experience declines in the absence of continued experience of the same type or as a consequence of a novel experience or trauma.

Insect populations and species vary in their ability to learn (Hoedjes et al. 2011). Genetic variability within strains has been used to analyze learning in *Drosophila*, *Phormia* flies, and the honey bee (McGuire and Hirsch 1977, McGuire 1984, McGuire and Tully 1987, Tully 1996, Menzel 1999). *Drosophila melanogaster* can be sensitized and habituated, learn associations with positive or negative reinforcement, and be classically conditioned (Davis and Dauwalder 1991). *Drosophila melanogaster* can learn to run away from specific odors that they previously experienced with an electric shock and hungry flies can learn to run toward odors previously associated with a sugar reward. Flies can learn visual, tactile, spatial, and proprioceptive cues (Waddell and Quinn 2001). Analyses of memory mutants in *Drosophila*, including *dunce*, *rutabaga*, *amnesiac*, *radish*, *zucchini*, *cabbage*, *tetanic*, *turnip*, *linotte* and *latheo*, indicate that memory consists of distinct phases: short-term, intermediate, long-term, and anesthesia-resistant memory (Table 11.3, Davis 1996, Sokolowski 2001).

Genetic analyses of learning in *D. melanogaster* began in the mid-1970s in Seymour Benzer's laboratory when *D. melanogaster* was trained to avoid an odor associated with a shock (Benzer 1973, Vosshall 2007). The learned avoidance lasted only a few hours, but the odor-avoidance test was used to screen mutagenized flies for strains that had normal olfaction and aversion to shock, but an abnormally low ability to associate odors with shocks. The mutant flies obtained were poor learners, but each had different phenotypes (Table 11.3). Flies with the mutant gene *amnesiac* had a nearly normal learning ability, but forgot rapidly. Flies with mutated *dunce* genes had a shortened memory for several different conditioned behaviors (Davis and Dauwalder 1991) due to a defective gene for cAMP-specific phosphodiesterase, an enzyme that regulates levels of cyclic AMP (cAMP). The *dunce* flies have elevated cAMP levels (Zhong and Wu 1991). Cyclic AMP is part of a "second messenger" signaling pathway in nerve

**Table 11.3: Some Single Genes Involved in Learning and Memory of *Drosophila melanogaster*.**

Gene	Mutant phenotype(s)	Function(s)
<i>dunce</i> <sup>+</sup>	Short-term memory defective	cAMP-specific phosphodiesterase involved in olfactory learning and memory Affects locomotor rhythms, ethanol tolerance, learning
<i>rutabaga</i> <sup>+</sup>	Short-term memory defective	Adenylate cyclase decreases expression of cAMP, affects courtship, learning, ethanol tolerance, grooming
<i>amnesiac</i> <sup>+</sup>	Middle-term memory defective	Neuropeptide, stimulates cAMP synthesis Affects ethanol tolerance, sleep regulation
<i>radish</i> <sup>+</sup>	Anesthesia-resistant long-term memory	Affects only one type of long-term memory
<i>cabbage</i> <sup>+</sup>	Long- and short-term memory (?)	Involved in olfactory learning
<i>turnip</i> <sup>+</sup>	Long- and short-term memory	Involved in the protein kinase C pathway Affects olfactory discrimination, larval, visual and reward learning
<i>latheo</i> <sup>+</sup>	Acquisition of initial memory defective?	Involved in short-term memory; affects DNA replication and synaptic plasticity?
<i>linotte</i> <sup>+</sup>	Retarded learning Mutants have structural brain defects (mushroom bodies and central complex)	Encodes receptor tyrosine kinase
<i>Volado</i> <sup>+</sup>	Short-term memory Expressed in mushroom bodies	Cell surface receptor altered, involved in synaptic remodeling underlying learning and memory; two variants of $\alpha$ -integrin coded for
<i>leonardo</i> <sup>+</sup>	Short-term and olfactory learning	Affects protein 14-3-3-, which is involved in intracellular signaling that activates and represses protein kinase C activity, activates tyrosine hydroxylase and tryptophan hydroxylase (enzymes involved in catecholamine and serotonin synthesis)

(Adapted from Dubnau and Tully 1998, Sokolowski 2001, Waddell and Quinn 2001, FlyBase.)

cells that help form associative memories. The *dunce* flies have impaired synaptic transmission because the excess of cAMP leads to hyperpolarization of the synaptic terminals, resulting in a chronically lowered availability of neurotransmitter (Delgado et al. 1992).

The *dunce*<sup>+</sup> gene is one of the largest and most complex identified in *Drosophila*, extending over 140 kb. It produces, by the use of multiple transcription start sites and alternative splicing of exons and differential processing of 3' sequences, at least eight to ten RNAs ranging from 4.2 to 9.5 kb. One unusually large intron, 79 kb in length, contains at least two genes (*Sgs-4* and *Pig-1*) within it (Chen et al. 1987, Qiu et al. 1991). This "genes-within-an-intron" arrangement is uncommon. One of the contained genes, *Sgs-4*<sup>+</sup>, is expressed in larval salivary glands and provides the glue used by larvae to attach themselves to the

surface for pupation. The second gene, *pre-intermolt<sup>+</sup>*, also is expressed in larval salivary glands. Genes homologous to *dunce<sup>+</sup>* have been identified in mice, rats, and humans, and the mammalian counterpart of *dunce<sup>+</sup>* functions in regulating mood (Tully 1991a).

*dunce<sup>+</sup>* is expressed in the mushroom bodies in the brain of *D. melanogaster* (Figure 11.1). This was discovered because the mushroom bodies can be stained with an antibody to the *dunce<sup>+</sup>*-encoded protein (Figure 11.1). The activity of the *dunce<sup>+</sup>* gene was identified by the **enhancer trap** method (O'Kane and Gehring 1987), a technique that involves placing a **reporter gene** (such as  $\beta$ -galactosidase which turns the fly's brain tissues blue when the substrate is added) into the *P* element under the control of a weak constitutive promoter. When this *P* element is brought in proximity to a tissue-specific enhancer after the *P* inserts into a chromosome,  $\beta$ -galactosidase expression will be regulated by the "native" enhancer in a tissue- and stage-specific pattern. Ideally,  $\beta$ -galactosidase will be expressed in a manner similar to the native gene. To determine which genes are expressed in the mushroom bodies, fly brains were screened and some 50 learning mutants were identified, including several alleles of *rutabaga<sup>+</sup>*. Mutations of *rutabaga<sup>+</sup>* cause decreased expression of cAMP and the rutabaga protein was identified as an adenylate cyclase (Han et al. 1992, Table 11.3).

Mushroom bodies are important for olfactory learning and memory. In *D. melanogaster* these structures are paired and consist of  $\approx$ 2500 neurons (Davis 1993, Heisenberg 1998, Figure 11.1). Mushroom bodies receive olfactory information from the antennal lobes. Mushroom bodies house part of the short-term memory for odors, are required for courtship conditioning memory, and are necessary for context generalization in visual learning, as well as regulating the transition from walking to rest (Zars 2000).

Learning requires other brain centers, including the antennal lobes, the central complex, and the lateral protocerebrum in insects (Davis 1993, Hansson and Anton 2000). During metamorphosis, the nervous system of holometabolous insects such as *Drosophila* changes significantly. A controversy has existed as to whether flies retain learned behavior after metamorphosis from larvae to adults. There is no evidence that larval conditioning induces a change in adult olfactory responses (Barron and Corbet 1999). This is not surprising, because larval sense organs undergo histolysis during the pupal stage and adult sense organs are formed *de novo* from imaginal discs. The mushroom bodies of the fly brain are extensively rewired during metamorphosis.

*Drosophila* carrying a mutant version of the *turnip<sup>+</sup>* gene have difficulty in olfactory discrimination, conditioning of leg position, larval, visual and reward learning (Table 11.3, Choi et al. 1991). Additional mutated genes, including

*radish*, *amnesiac*, *cabbage*, *latheo* and *linotte*, are involved in abnormal learning or memory of *D. melanogaster* (Table 11.3). For example, flies with the X-linked *radish* mutation initially learn in olfactory tests, but their subsequent memory decays rapidly at both early and late times after learning. The *radish* flies show normal locomotor activity and sensitivity to odor cues and electric-shock reinforcements used in the learning tests. Anesthesia-resistant memory, or consolidated memory, is strongly reduced in *D. melanogaster* with the *radish* phenotype (Folkers et al. 1993).

The *rutabaga<sup>+</sup>* gene codes for an adenyl cyclase and is expressed in *Drosophila* mushroom bodies. This gene is involved in olfactory short-term memory (Zars et al. 2000). Likewise, *Volado<sup>+</sup>* is expressed in mushroom body cells of *Drosophila* and mediates short-term memory in olfactory learning (Grotewiel et al. 1998).

How does long-term memory develop and persist over years? Majumdar et al. (2012) found that the *Drosophila* Orb2 protein forms amyloid-like oligomers upon neuronal stimulation and these are enriched in the synaptic membrane, which may be critical for the persistence of long-term memory. Mutants of the Orb2 protein had impaired long-term memories, suggesting that self-sustaining amyloid-like conversion of the neuronal cytoplasmic polyadenylation element-binding proteins (CPED) is involved in long-term memory lasting >12 hours.

The enlightenment obtained from the study of *Drosophila* learning mutants is providing an understanding of learning in higher organisms (Tully 1991a,b, 1996, Dubnau and Tully 1998, Sokolowski 2001, Waddell and Quinn 2001, Majumdar et al. 2012).

### 11.5.3 Functional Genomics of Odor Behavior in *Drosophila*

The ability to respond to odors is essential for survival and reproduction, allowing insects to select mates, find and choose food, and locate appropriate oviposition sites. Insects detect odors with antennae and maxillary palps, upon which sensory hairs (sensilla) are present. Each sensillum houses the dendrites of a few olfactory-receptor neurons. Insect antennae can contain >100,000 sensilla but *D. melanogaster* antennas contain ≈400 sensilla. Sexual dimorphism in antennal structure is common in insects, and immature insects typically have fewer sensilla than adults, perhaps because insect eggs typically are deposited in or near the appropriate food for the larvae.

Studies of the genetic basis of odor behavior in insects first used *D. melanogaster* as a model system (Field et al. 2000, Vosshall 2000, Anholt et al. 2001). Efforts also are being made to evaluate olfaction, learning, and memory in the honey bee (Maleszka 2000, Robertson and Wanner 2006). Robertson and

Wanner (2006) identified 170 odorant-receptor genes in the honey bee, a dramatic increase compared to the 62 and 79 genes found in *D. melanogaster* and *Anopheles gambiae* genomes.

Odors are received by olfactory receptors located on the antennae and the maxillary palps, which send their axons to the antennal lobes in insect brains. Each third antennal segment in *D. melanogaster* contains  $\approx$ 1300 olfactory receptor cells and each maxillary palp carries 120 chemosensory neurons (Anholt et al. 2001). These neurons project to 43 glomeruli in the antennal lobe of the brain. From there, processed olfactory information is relayed to the mushroom body and the lateral horn of the protocerebrum. Insect odorant receptors are seven-transmembrane-domain proteins, but are not related to G protein-coupled receptors (Carey and Carlson 2011).

It is thought that there are fewer than 100 types of odor receptors in insects (Vosshall et al. 1999), perhaps as few as 50 or 60 (Vosshall et al. 2000), and these are different from those found in mammals (Carey and Carlson 2011). Mammals have  $>10,000$  different receptor types. Insect receptors consist of large multi-gene families (Clyne et al. 2000, Vosshall et al. 1999). Once an odor or pheromone has activated the olfactory receptors, it needs to be deactivated. Several enzymes have been found that appear to degrade odor stimulants, including esterases, oxidases, and glutathione transferases (Field et al. 2000).

*Drosophila melanogaster* is able to recognize and discriminate between a large number of odors (Vosshall 2001, Rutzler and Zwiebel 2005). Because there are as few as 50 or 60 types of receptors, each olfactory sensory neuron responds to several odorants, but responds maximally to one (Dryer 2000). Although the average olfactory-receptor gene is expressed in 20 olfactory neurons, some receptor genes are expressed in only two to three neurons. Seven olfactory receptor genes are expressed solely in the maxillary palp (Vosshall et al. 2000).

The sequencing of the pea aphid, *Acyrtosiphon pisum*, genome allowed Smadja et al. (2009) to identify 79 putative odorant and gustatory receptor genes.

#### 11.5.4 Behavior of *Apis mellifera*

Mushroom bodies in the Hymenoptera are much larger than those in *Drosophila*, which may reflect the importance of mushroom bodies for social behavior, learning, and memory in the honey bee (Rinderer 1986, Rybak and Menzel 1993, Meller and Davis 1996).

Social Hymenoptera (ants, bees, and wasps) have complex behaviors, including caring for their brood. Social bee species such as *Apis mellifera* feed, protect

and nurse larvae; store food; and respond to adverse environmental factors. Adult workers search for nectar and pollen at unpredictable sites, they learn the celestial and terrestrial cues that guide their foraging trips over long distances and allow them to find their nest sites once again. Workers learn how to respond to the changing position of the sun, to a pattern of polarized light during the day, and to landmarks. Associative learning is an essential component to foraging behavior and dance communication. Hive mates attending a dance performance learn the odor the dancing bee carries and seek out that same odor when they forage for food.

The complexity of bee behavior makes it an ideal organism to better understand learning, especially odor responses (Hammer and Menzel 1995, Ray and Ferneyhough 1999, Menzel 1999), as well as to analyze social behavior (Bloch and Grozinger 2011). Associative olfactory learning in honey bees has several features similar to higher forms of learning in vertebrates (Grunbaum and Muller 1998).

#### **11.5.5 Pheromones in Insects**

Many insects use chemical cues as signals to find mates and molecular-genetic methods are now used to study various aspects of pheromone-response behavior. For example, genes are being identified that code for proteins involved in the synthesis of **pheromones** (substances released by the body that cause a predictable reaction by another individual of the same species), the perception of **semiochemicals** (chemicals that influence insect interactions), and the processing of the signals (Krieger and Breer 1999, Mombaerts 1999, Tillman et al. 1999, Field et al. 2000).

Pheromone biosynthesis appears to use one or a few enzymes that convert the products of normal primary metabolism into compounds that act as pheromones (Tillman et al. 1999). For example, pheromones arise from isoprenoid biosynthesis, or by the transformation of amino acids or fatty acids. Several genes encoding the enzymes involved in transforming metabolites into pheromones have been cloned and sequenced (Field et al. 2000). Three hormonal messengers regulate production of pheromones by insects: juvenile hormone III, ecdysteroids, and a neuropeptide called pheromone-biosynthesis-activating neuropeptide (PBAN).

The antennae contain olfactory organs (*sensillae*) that mediate pheromone perception. Some receptor neurons on the antennae appear to respond to one particular chemical (specialist neurons), but others appear to respond to a number of compounds (generalist neurons). Pheromones often are perceived in combination with other chemicals, including plant volatiles.

The detection of pheromones and other chemicals by insects involves proteins (**odorant binding proteins [OBPs]**) that carry the compounds from the surface of the antennal sensilla through the sensillum lymph to the receptors and the olfactory neurons (Prestwich 1996, Krieger and Breer 1999). The odorant-binding proteins (which includes pheromone-binding proteins) are small, soluble proteins that are concentrated in the sensillum lymph (Christophides et al. 2000). Analysis indicates that the binding proteins of unrelated species have low levels of amino-acid sequence similarity. It appears that there has been gene duplication and divergence of odorant-binding protein genes, with moth proteins belonging to one branch and the proteins of other insects more distantly related (Christophides et al. 2000).

#### 11.5.6 Neurobiochemistry of *Drosophila*

Molecular neurobiology is concerned about how the nervous system controls behavior at the molecular level (Glover and Hames 1989). What are the biochemical substrates of behavior? A molecular-genetic approach using *Drosophila* is providing interesting answers for both insects and mammals, although *D. melanogaster* has  $\approx$ 250,000 neurons and humans have  $\approx$ 10<sup>12</sup> neurons in the brain. For example, a potassium (K)-channel gene family was cloned first from *Drosophila* and subsequently from humans and mice using probes from *Drosophila*. The *Shaker*<sup>+</sup>, *Shal*<sup>+</sup>, *Shab*<sup>+</sup>, and *Shaw*<sup>+</sup> subfamilies of the K-channel gene family have been found in the Chordata, Arthropoda, and Mollusca, suggesting that the ancestral K-channel gene had already given rise to these subfamilies by the time of the Cambrian radiation (Salkoff et al. 1992).

Many enzymes and receptors are involved in neurobiology, including receptors for neurotransmitters and hormones, ion-channel proteins and associated signal-transduction components, brain-specific protein kinases, enzymes for transmitter synthesis, neuropeptide-processing enzymes, neuron-specific growth factors and their receptors, inhibitors of neuronal growth, glial-specific growth factors and their receptors, proteins associated with memory, neuronal cytoskeleton and axonal-transport proteins, and others not listed here or yet to be identified. A major endeavor in molecular neurobiology involves establishing the primary structure of all the categories of proteins involved in nerve-signal reception and transmission (Barnard 1989).

##### 11.5.6.1 Electrical Signaling

The nervous system receives information about its internal and external environment, processes this information, and produces an appropriate response. The signaling of nerve cells depends on the electrical status of their outer

membranes. Nerve cells maintain a potential difference across the membrane with the inside of the cell negative relative to the outside of the cell. The resting nerve cell also maintains concentration gradients of sodium (Na), calcium (Ca), and potassium (K) ions. Na and Ca ions are at a relatively high concentration outside the cell, whereas K-ion levels are relatively high inside the cell. Signaling involves a change in the resting-membrane potential brought about by charge transfers carried by ionic fluxes through gated pores formed by transmembrane proteins called **channels**.

Ion-channel proteins catalyze the transmembrane flow of ionic charge by forming narrow, hydrophilic pores through which ions can diffuse (Miller 1991). **Ion channels** must open or close rapidly in response to biological signals (=gating). Furthermore, the open pore is generally selective and will determine which ions will permeate and which will not (**ionic selectivity**). Thus, a specific channel will permit K but not Na to pass.

Stimuli from the environment are perceived by specialized nerve cells (sensory cells). Each type of sensory cell responds to a particular stimulus such as light, sound, touch, heat, or chemicals such as pheromones. These sensory cells transform and amplify the energy provided by a stimulus into an electrical signal (=sensory transduction). **Sensory transduction** is probably due to an alteration in the ionic permeability of the sensory-cell membrane, which causes a depolarization of the membrane of the sensory cell from its resting level. The amplitude and duration of this departure generally increases logarithmically with the intensity of the stimulus. This signal is local and is not transmitted along the nerve cell; however, it acts as a stimulus to the axon and if the depolarization increases over a threshold level, the signal will trigger a change in **action potential** in the axon. Action potentials are all-or-nothing electrical impulses that propagate without distortion or attenuation along the entire length of an axon.

The generation and propagation of an action potential alters ionic conditions within the cell. When axonal membranes are depolarized, Na channels open and allow Na ions to flow down their gradient into the cell, producing the depolarizing phase of an action potential. Within milliseconds after the Na channels are opened, they are inactivated, but at about the same time the membrane depolarization activates K channels, and the reciprocal K flow repolarizes the cell and restores the membrane-resting potential. During the course of an action potential, the Na currents in one region of the axon membrane cause the depolarization and firing of an action potential in an adjacent region of the membrane so that the action potential is propagated along the full length of the axon.

The electrical signal is transmitted between cells at special sites called **synapses**, which occur between two nerve cells as well as between nerve cells and effectors such as muscle cells. The signal is relayed by a chemical neurotransmitter that is packaged in membrane-bound vesicles. When an action potential reaches the pre-synaptic terminal, the depolarization activates Ca channels in the presynaptic membrane and the subsequent influx of Ca ions leads to the release of neurotransmitter. The neurotransmitter diffuses to the post-synaptic cell and interacts with specific receptors on that cell surface. Receptors are activated in response to binding of the specific neurotransmitter molecules. Generally the size and duration of a synaptic potential reflects the amount of transmitter released by the pre-synaptic terminal. By depolarizing the post-synaptic cell above the threshold, the synaptic potential triggers the generation of an action potential, which continues the signaling one step further along the neural pathway ([Ganetzky and Wu 1989](#)).

#### 11.5.6.2 Neurotransmitters

Acetylcholine (ACh) is the major neurotransmitter in the central nervous system of *Drosophila* and other insects. Choline acetyltransferase (ChAT) is the biosynthetic enzyme, and acetylcholinesterase (AChE) is the degradative enzyme. AChE terminates synaptic transmission by rapidly hydrolyzing acetylcholine ([Fournier et al. 1989, Ganetzky and Wu 1989](#)).

The acetylcholinesterase gene (Ace) from *Drosophila* is 34 kb in size and is split into 10 exons, with the splicing sites of the two last exons precisely conserved among *Drosophila* and vertebrate cholinesterases ([Fournier et al. 1989](#)). The deduced mature Ace transcript is 4.2 kb long. A gene for an acetylcholine receptor subunit has been identified, and the amino-acid sequence of this AChR shares similarity with vertebrate sequences.

#### 11.5.6.3 Ion Channels

Two types of ion channels, permeable to sodium (Na) or potassium (K) ions, are responsible for membrane electrical phenomena. Multigene families encode the Na and K channels. Mutated genes that affect Na channels include no-action-potential, temperature-sensitive, or *nap<sup>ts</sup>*, and paralytic, or para ([Salkoff et al. 1987, Kernan et al. 1991](#)). The mutation *para<sup>ts</sup>* is a temperature-sensitive mutation that causes instantaneous paralysis of adults at 29°C and of larvae at 37°C ([Loughney et al. 1989](#)). Mutations of several different genes (*Shaker*, *Shal*, *Shab*, and *Shaw*) alter K currents ([Covarrubias et al. 1991](#)). One of the best-studied mutations is *Shaker* and *D. melanogaster* carrying the *Shaker* allele exhibit aberrant behavior, shaking their legs when anesthetized with ether ([Papazian et al. 1987](#)).

Flies in which the *Shaker* gene is deleted still have K currents and [Butler et al. \(1989\)](#) isolated three additional family members, *Shab*<sup>+</sup>, *Shaw*<sup>+</sup>, and *Shal*<sup>+</sup>. These four genes define four K-channel subfamilies in *Drosophila* and homologous genes isolated from vertebrates all appear to fall into one of these four subclasses.

Other K-channel mutants, including *eag* ([Warmke et al. 1991](#)) and a calcium-activated K-channel gene (*slo*) ([Atkinson et al. 1991](#)), have been isolated. Another neurotransmitter,  $\gamma$ -aminobutyric acid (GABA), is a major inhibitory agent in the insect nervous system. The synthesis of GABA is controlled by the enzyme glutamic acid decarboxylase (GAD) ([Jackson et al. 1990](#)).

### **11.5.7 Divergent Functions of Est-6 and Est-5 in Two *Drosophila* Species: A Cautionary Tale of Homologs**

Evolutionary changes in gene regulation can be important in macroevolutionary change and species divergence. One case study involves an analysis of the *Esterase-6* gene in *Drosophila melanogaster* and its homolog (*Esterase-5*) in *D. pseudoobscura* ([Brady and Richmond 1990](#)). Both influence behavior in *D. melanogaster* but have a very different function in the two species, indicating that sequence homology may not be equivalent to behavioral homology.

The *Esterase-6* (*Est-6*) gene influences male-mating speed and rate of remating by *D. melanogaster* females. Fast- and slow-variants of the Esterase-6 protein are produced in natural populations of *D. melanogaster*. More Esterase-6 protein is produced in adult males than in females. The enzyme is highly concentrated in the anterior ejaculatory duct of males and is transferred to females during the first 2–3 minutes of the 20-minute copulation. Enzyme activity in females can be detected up to 2 hours after mating and influences the timing of remating by females. Males transfer a substance in the seminal fluid that is converted in the females' reproductive tract by the Esterase-6 protein into a pheromone that serves as an **antiaphrodisiac**. The antiaphrodisiac reduces the sexual attractiveness and receptivity of females, reducing the likelihood she will remate. Because the sperm from the most recent male takes precedence in fertilizing a female's eggs, this behavior encourages monogamy in *D. melanogaster* females ([Richmond et al. 1986](#)).

The *Esterase-6* gene also influences the rate of mating of males in *D. melanogaster*. Males with the slow variant of the protein require 10.2 minutes to achieve copulation with females, whereas males with the faster-moving protein require only 5.7 minutes. Once the *Est-6* gene was cloned, it was used as a probe to identify homologous genes in related species ([Brady and Richmond 1990](#)), and

*Est-5* was isolated from *D. pseudoobscura*. *Est-5* in *D. pseudoobscura* is expressed in the eyes and hemolymph. Despite these different patterns of expression, *Est-6* and *Est-5* have similar protein products, transcripts, and DNA sequences.

When *Est-5* from *D. pseudoobscura* was cloned into a *P* element and introduced into *D. melanogaster*, its activity and pattern of expression in *D. melanogaster* matched those of *D. pseudoobscura*, implying that regulatory elements had been conserved since the divergence of the two species 20–46 million years ago. [Brady and Richmond \(1990\)](#) speculated that the enzyme in the common ancestor of these two species had a more extensive expression pattern. After their divergence, regulatory mutations may have occurred that enhanced *Est-5* expression in the eyes of *D. pseudoobscura*, whereas mutations in *Est-6* led to increased expression in the male ejaculatory duct of *D. melanogaster*. Thus, the use of DNA sequence similarity to identify behavioral (and other) genes can lead to surprises.

#### 11.5.8 Courtship Behavior in *Drosophila*

Mating behavior of *D. melanogaster* is stereotypical, with a fixed sequence of actions that are under genetic control. Courtship involves visual stimuli, acoustic signals, and pheromones ([Hall 1994](#), [Yamamoto et al. 1997](#), [Goodwin 1999](#), [Savarit et al. 1999](#), [Greenspan and Ferveur 2000](#), [Ganter et al. 2011](#)). Male courtship behavior involves six elements in the following order: orienting→following→wing vibration→licking→attempting to copulate→copulation.

Sexual differentiation in *Drosophila*, described in Chapter 10, is controlled by a short cascade of regulatory genes. The expression of these regulatory genes determines all aspects of maleness and femaleness in the soma and the central nervous system. The genes also influence courtship behavior. Sexual behavior is irreversibly programmed during a critical period as a result of the activity, or inactivity, of the control gene *tra*<sup>+</sup>. Male behavior is replaced by female behavior when *tra*<sup>+</sup> is expressed around the time of puparium formation ([Arthur et al. 1998](#)).

Other genes indirectly affect courtship behavior in *Drosophila*, including genes that involve general behavior, visual behavior, olfaction, learning/memory genes, regulating periodicity of behavior, courtship song mutants, and female receptivity ([Hall 1994](#)). [Immonen and Ritchie \(2012\)](#) analyzed how gene expression changes in response to courtship song in *D. melanogaster*, using microarrays and quantitative polymerase chain reaction (PCR), and identified differentially expressed genes, some of which were up-regulated and some of which were down-regulated. Interestingly, some immune-response genes were

up-regulated and some down-regulated. At present, the function of immunity genes in mating behavior is not known. One of the genes identified was *glucose dehydrogenase*, which facilitates sperm storage in mated females, suggesting that transcriptional changes associated with mating may begin during courtship in advance of egg fertilization.

The *fruitless<sup>+</sup>* gene is involved in both sex determination and courtship behavior and is active in the central nervous system (Hall 1994, Ryner et al. 1996, Goodwin 1999, Baker et al. 2001). Males with a mutation in *fruitless* may court both females and males without copulating. Male flies expressing this mutated gene are unable to bend their abdomens in the presence of females they are courting because they lack a male-specific Muscle of Lawrence (MOL). Some mutant alleles of *fruitless* cause males to be homosexual (they court only males), while others cause males to be bisexual (they court both males and females) (Yamamoto et al. 1997).

The *fruitless<sup>+</sup>* gene is the first gene in the sex-determination hierarchy functioning specifically in the central nervous system, with mutants of this gene affecting nearly all aspects of male sexual behavior (Ryner et al. 1996, Villegas et al. 1997, Goodwin et al. 2000). It is at least 140 kb long and produces a complex array of transcripts by using four promoters and alternative splicing; the male-specific transcripts are only expressed in a small fraction of the central nervous system (Goodwin et al. 2000). Ito et al. (2012) showed that *fruitless<sup>+</sup>* encodes a set of transcription factors that promote male sexual behavior by forming a complex with a transcriptional cofactor, which recruits two chromatin regulators. This combination masculinizes individual sexually dimorphic neurons.

Another gene, *dissatisfaction<sup>+</sup>*, is necessary for some aspects of sex-specific courtship behavior and neural differentiation in *D. melanogaster* of both sexes. Mutant males are bisexual but, unlike *fruitless* males, attempt to copulate. Males with the mutant *dissatisfaction* phenotype take longer to copulate with females and females with the mutant *dissatisfaction* phenotype are unreceptive to male advances during courtship and do not lay mature eggs (Goodwin 1999). Mating behavior of normal females involves the following sequence: stop moving→offer the courting male a chance to lick the female's genitalia→allow males to attempt copulation. Nonreceptive females leave the courting male, and if the male pursues her, she may kick him. Nonreceptive virgin females persistently repel male approaches by lifting their abdomens up to block any physical contact with males. Nonreceptive fertilized females lower their abdomen, extrude their ovipositors and eggs to repel males. Thus, female receptivity varies with age, diet, hormonal condition, and mating experience. A mutation of

*spinster<sup>+</sup>* affects female sexual receptivity throughout life, and females with the mutated *spinster* phenotype continuously leave, kick, or fend off courting males (Hall 1994, Suzuki et al. 1997).

Both *D. melanogaster* and *D. simulans* females produce contact pheromones, which consist of cuticular hydrocarbons that elicit wing displays by males (Ferveur 1997). These chemical signals have a low volatility, act at a very short distance (a few millimeters) and are perceived by contact rather than smell. Flies from a given strain, sex, and age produce a reproducible pattern of cuticular hydrocarbons, the biochemical pathway of which is under genetic control. The most important hydrocarbons involved are 7-tricosene and 7-pentacosene. One mutation, *Ngbo*, influences the ratios of 7-tricosene and 7-pentacosene in *D. simulans*. Another, *kete*, reduces the amount of 7-tricosene and all other linear hydrocarbons but does not affect the ratio (Ferveur and Jallon 1993). Flies homozygous for both *kete* and *Ngbo* have reduced viability and fertility, perhaps because they have very little 7-tricosene.

Experiments were conducted to eliminate all known cuticular hydrocarbons in *D. melanogaster* using genetic methods to determine how mating behavior would be modified (Savarit et al. 1999). The results were surprising; contrary to the expectation that *D. melanogaster* females lacking cuticular pheromones would elicit no courtship by males, such females remained attractive. Additional analysis indicated that undetermined pheromone(s), probably also cuticular hydrocarbons, were present on both control and transgenic flies. Savarit et al. (1999) suggested that the newly discovered pheromones represent ancestral attractive substances in *D. melanogaster* and its sibling species.

A male-biased gene family, *takeout<sup>+</sup>*, affects male courtship behavior in *D. melanogaster*, as well as other aspects of its biology (Vanaphan et al. 2012). This gene family is conserved across more than 350 million years of insect evolution, and may have evolutionarily conserved sex-specific roles in male-mating behavior among all insects.

#### 11.5.9 Speciation Genes in *Drosophila* and Other Insects

Changes in sexual behavior can result in reproductive isolation between populations, leading to speciation. Studies of sexual behaviors in *Drosophila* species have led to different conclusions about the number of genes involved in speciation by this mechanism (Doi et al. 2001, Ting et al. 2001, Arbuthnott 2009).

Analyses of the genetics of speciation usually involve crossing pairs of related species that do not normally mate, but will do so under laboratory conditions

when given no choice. The progeny of such “interspecific” crosses then are examined to determine what phenotypes are related to their reproductive isolation. Reproductive isolation can be due to sterility of the hybrids (**postmating isolation**) or differences in mate preference (contributing to **premating isolation**). Study of postmating isolation mechanisms indicate that a number of genes (loci) are involved.

Premating isolation is thought to be a common cause of speciation in insects caused by a divergence in male sexual signals and female preferences. As a result, **assortative mating** occurs, with individuals preferring to mate with individuals who resemble themselves. [Ting et al. \(2001\)](#) studied the sexual isolation of two populations of *D. melanogaster* (M and Z forms). Z females strongly prefer Z males over M males; this preference is due to at least four loci on chromosome III that influence male behavior and at least three loci that influence female behavior, suggesting that premating isolation has a multigenic basis.

By contrast, [Doi et al. \(2001\)](#) used *D. ananassae* and its sibling species *D. pallidosa* to analyze sexual isolation. These species are almost completely isolated, but *ananassae* females no longer discriminate strongly against *pallidosa* males if the males are prevented from singing their songs (by removing their wings) or if females are prevented from hearing them (by removing their ears). This suggests that divergence in male song patterns and associated female preferences underlies this sexual isolation. The genetic basis of the preference of *ananassae* females for *ananassae* males appears to be a single dominant gene.

The divergence of acoustic signals alone appears to explain the isolation between the *ananassae* and *pallidosa* species, but the basis of mate choice in the M and Z forms of *D. melanogaster* appears to involve different signals, which probably are determined by multiple genes. The histories of these populations could explain the different isolation mechanisms. M and Z forms of *D. melanogaster* appear to have diverged in the same geographic area (**sympatric speciation**), but the *ananassae* and *pallidosa* species may have evolved while isolated geographically (**parapatric species**). [Butlin and Ritchie \(2001\)](#) concluded that analyses of additional populations and species are required to resolve how many behavioral genes are involved in speciation.

[Arbuthnott \(2009\)](#) reviewed the studies on reproductive isolation in insects, and concluded that the analysis method was important in the conclusions reached; if the behavior analyzed actually consisted of several behaviors (i.e. “mating behavior” consists of male signaling, female reception, female rejection, and others), then the conclusion was likely to be that multiple genes underlie reproductive isolation. If, however, specific behaviors were studied,

many examples (25 of 36, or 69%) were found in which one or a few genes were involved. However, in cases in which reproductive isolation has evolved rapidly by changes in behavior (mate signaling), postzygotic isolating mechanisms may not have developed.

#### **11.5.10 Personality in Insects: *Tribolium confusum*, *Apis mellifera*, *Acyrthosiphon pisum*, and *Pyrrhocoris apterus***

Individuals of many animal species have personality, i.e., they have consistent differences in behavior across time, situations, and/or contexts (Wolf et al. 2007, 2008; Nakayama et al. 2012). Personality in insect species is recognized as an issue relevant to ecology and evolution and can complicate molecular analyses of behavior if this is not recognized. The evolution of animal personalities is not well understood, but Wolf et al. (2007) argue that personalities can be given an adaptive explanation based on the trade-off between current and future reproduction that can result in polymorphic populations in which some individuals emphasize future fitness and others are less likely to take risks. Wolf et al. (2008) further argue that "the benefits of responsiveness are frequency-dependent; that is, being responsive is advantageous when rare but disadvantageous when common ... Second, positive-feedback mechanisms reduce the costs of responsiveness; that is, responsiveness is less costly for individuals that have been responsive before." Thus, personality variation should affect population ecology and dynamics.

Insect species documented to have personalities include *Tribolium confusum*, *Apis mellifera*, *Acyrthosiphon pisum*, and *Pyrrhocoris apterus*. For example, individuals of *Tribolium confusum*, exhibit consistent differences in walking activity and death-feigning behavior (a possible mechanism to escape predation). Selection for higher or lower frequencies and longer or shorter durations of death feigning resulted in two genetically distinct strains, and Nakayama et al. (2012) showed that lower activity levels were the result of lower brain dopamine levels. Administration of caffeine to low-activity strains resulted in decreased durations of death feigning and increased levels of activity.

Honey bee workers exhibit differences in scouting behaviors, with some more adventurous than others (Liang et al. 2012). Adventurous scouts for food and nest sites had extensive differences in brain gene expression (catecholamine, glutamate, and  $\gamma$ -aminobutyric acid signaling). Octopamine and glutamate treatments increased scouting activity, but dopamine-antagonist treatment decreased it. Interestingly, "Our results demonstrate intriguing parallels between honey bees and humans in novelty-seeking behavior. Although

the molecular mechanisms that produce this behavioral variation are similar, it is unknown whether both species inherited them from a common ancestor or evolved them independently" (Liang et al. 2012).

The pea aphid *Acyrthosiphon pisum* has differential escape responses to predator attack (dropping or not dropping off the plant) (Schuett et al. 2011). Genetically identical aphid clones expressed different phenotypes (dropper, nondropper, and inconsistent). Individuals within a clone dropped or did not drop and the clones varied in their responses. Some clones were consistent over repeated trials, others contained both consistent and inconsistent individuals, one clone failed to produce droppers. Individuals were repeatable in their escape response in six trials over five days of adult life. When individuals were reared under different conditions, they were consistent in their tendency to drop and in the repeatability of their behavior. This study showed clonal individuals expressed personality variation.

The short- and long-winged individuals of the firebug *Pyrrhocoris apterus* (Heteroptera: Pyrrhocoridae) exhibit consistent behaviors over time and across contexts, indicating that these bugs have personalities. Females of the two wing morphs have different personalities, with winged females being "braver and more exploratory." These data support the model of Wolf et al. (2007) that, "individuals choose different strategies to find the balance between present and future reproduction. In the case of firebug, it is known that there are differences in the behavior of brachypterous and macropterous individuals, e.g. in the higher walking activity and lowered mating propensity of macropterous individuals."

## 11.6 Symbionts and Insect Behavior

Discovery of the effects of symbionts on insect behavior required the use of molecular tools to identify and quantify their role in the biology of insects. Markov et al. (2009) found that mating preference of *D. melanogaster* depends, in part, on whether both males and females are infected with *Wolbachia*. Assortative mating depended on genotype, infection status, and a combination of genotype and infection status. Apparently mating choice can, "involve testing the partner for degree of genetic or biochemical similarity with self, based on chemoreception with possible immune system components" and, in this case, *Wolbachia* was a significant component of that testing.

Another example in which symbionts affect the behavior of their host involves the nucleopolyhedrosis virus of *Lymantria dispar* (Hoover et al. 2011). The virus causes infected gypsy moth larvae to climb to the top of their host trees to die

(Hoover et al. 2011). The body contents of the infected larvae become liquefied, and the body bursts, releasing millions of infective virus particles. By contrast, healthy gypsy moth larvae hide in crevices or even climb down the tree to the soil to avoid bird predation during the day. The cause of change in behavior was identified as due to a gene in the baculovirus (ecdysteroid uridine 5'-diphosphate-glucosyltransferase) that encodes an enzyme that inactivates the molting hormone of gypsy moth larvae. Hoover et al. (2011) inoculated gypsy moth larvae with genetically modified virus containing or lacking the virus gene. As expected, deletion of the gene from the virus eliminated the climbing behavior of the larvae and rescue of the gene (adding the gene or gene product) restored the climbing behavior. This behavior was termed an “extended phenotype” because the gene in one organism (the parasite) had a phenotypic effect on another (the host).

Sharon et al. (2010) found that gut bacteria played a role in mating preference of *D. melanogaster* reared on different media. A population of *D. melanogaster* was divided and part was reared on a molasses medium and the other on a starch medium. When the populations were mixed, flies reared on molasses preferred to mate with other molasses flies and vice versa. The mating preference occurred after only one generation and was maintained for 37 generations. Antibiotic treatment eliminated the preference, indicating that microbes were responsible. The authors suggest the mating preferences were caused by the bacteria “changing the levels of cuticular hydrocarbon sex pheromones.”

As noted in Chapter 4, the facultative symbiont *Hamiltonella defensa* affects the defensive behavior of the pea aphid (Dion et al. 2011). Aphids containing *H. defensa* are protected against parasitoids and, as a result, the infected aphids spend less time being aggressive against the parasitoids and exhibit fewer escape behaviors. This change in behavior benefited both the aphid and the symbiont because the aphid was able to feed and survive at a greater rate than uninfected aphids.

## 11.7 Human Neurodegenerative Diseases and Addictions in *Drosophila*

*Drosophila* is perhaps unique among eukaryotes in the variety and level of sophistication that can be applied to understand its neurobiology and behavior. As a result, *Drosophila* is being studied to gain knowledge about various neurodegenerative diseases in humans (Mutsuddi and Nambu 1998, Andretic et al. 1999, Feany 2000, Fortini and Bonini 2000).

Modeling diseases in simple invertebrate systems is attractive because genetics can define cellular cascades mediating disease states such as the death of neurons in Parkinson's disease, the second most common neurodegenerative disorder in humans (Feany and Bender 2000). Transgenic *Drosophila* containing a mutant form of the human  $\alpha$ -synuclein gene exhibit the essential features of the Parkinson's disease in humans, making it possible to study the function of  $\alpha$ -synuclein and determine the underlying pathogenic mechanisms in a genetically tractable animal.

The *spongecake* mutant of *Drosophila* shows degenerative changes similar to that seen in humans with Creutzfeldt-Jakob disease, while the *eggroll* mutant produces changes similar to those seen in humans with Tay-Sachs disease (Min and Benzer 1997). The *beta-amyloid protein precursor-like* (*Appl*) gene of *Drosophila* encodes a homolog of the human  $\beta$ -amyloid precursor protein which gives rise to  $\beta$ -amyloid, a major component of the plaques found in patients suffering from Alzheimer's disease (Luo et al. 1992). Another protein associated with Alzheimer's disease, presenilin, has been found in *Drosophila* and studies suggest it also may be involved in the development of the pathology (Fortini and Bonini 2000). A *Drosophila* homolog was identified for the human gene for copper/zinc superoxide dismutase; mutants of this gene are implicated in Lou Gehrig's disease (McCabe 1995, Phillips et al. 1995).

A recessive mutant (*bubblegum*) in *D. melanogaster* exhibits adult neurodegeneration similar to that seen in the human disease adrenoleukodystrophy (ALD), otherwise known as the disease cured in the movie *Lorenzo's Oil* (Min and Benzer 1999). In ALD, high levels of very long chain fatty acids are produced that can be lowered by dietary treatment with a mixture of unsaturated fatty acids; feeding the ALD flies one of the components, glyceryl trioleate oil, blocked the accumulation of excess very long chain fatty acids and eliminated the pathology. Thus, *bubblegum* flies provide a model system for studying mechanisms of disease and screening drugs for treatment.

*Drosophila* may serve as a model organism to study the genetics of alcohol abuse and drug addiction in humans (Bellen 1998, Moore et al. 1998, Andretic et al. 1999, Wolf 1999, Bainton et al. 2000, Singh and Heberlein 2000). Alcohol addiction and many types of drug addictions appear to share common mechanisms (Bellen 1998, Moore et al. 1998). For example, the "dopamine hypothesis" suggests that addictive drugs may activate certain areas of the human brain leading to an increase in dopamine neurotransmitter release (Bainton et al. 2000). Elevation of dopamine probably provides a sense of wellbeing, pleasure, or elation resulting in a positive reinforcement. Dopamine is not the only

neurotransmitter acting in alcohol abuse; glutamate, serotonin, and GABA also may be involved. Furthermore, four of the five circadian genes (*period*<sup>+</sup>, *clock*<sup>+</sup>, *cycle*<sup>+</sup>, and *doubletime*<sup>+</sup>) in *D. melanogaster* influence the fly's responsiveness to cocaine and suggest a biochemical regulator of cocaine sensitization (Andretic et al. 1999).

Resistance to ethanol in *D. melanogaster* appears to be determined by multiple genetic components. Singh and Heberlein (2000) analyzed 23 mutant fly strains with different responses to ethanol and the effects of acute ethanol exposure on *Drosophila* locomotor behaviors are "remarkably similar to those described for mammals." Thus, study of *Drosophila* "may pave the way for an in-depth study of the genes involved in acute and chronic effects of ethanol" (Bellen 1998). Bainton et al. (2000) showed that, as in mammals, dopaminergic pathways in *Drosophila* play a role in modulating specific behavioral responses to cocaine, nicotine, or ethanol.

*Drosophila* flies can sleep, and they have become a model for understanding sleep in insects and other animals (Hendricks et al. 2000; Greenspan et al. 2001; Harbison et al. 2009a,b; Donlea et al. 2011, 2012; Soshnev et al. 2011). Flies that are "resting" choose a preferred location, become immobile for periods of up to 157 minutes at a particular time in the circadian day, and are relatively unresponsive to sensory stimuli. When rest is prevented, the flies tend to rest despite stimulation and exhibit a "rest rebound." In fact, flies subjected to long-term sleep deprivation may die. Drugs that affect sleep in mammals alter "rest" in flies, suggesting conserved neural mechanisms.

"During sleep, an animal cannot forage for food, take care of its young, procreate or avoid the dangers of predation, indicating ... sleep must serve an important function" (Greenspan et al. 2001), although there is no agreement yet as to its function(s) (Harbison et al. 2009b). Hypotheses proposed to explain the evolutionary maintenance of sleep include conservation of energy by reduced expenditure of nutrients, restoration of brain glycogen, and maintenance of homeostasis of synapses (Harbison et al. 2009b). Sleep is important in learning and memory (Donlea et al. 2011). Sleep disorders in humans are common, but the genes underlying these disorders are difficult to study (Kolker and Turek 1999). Analysis of *Drosophila* behavior at the molecular level offers promise of elucidating this evolutionarily important aspect of survival, and mutagenesis studies suggest that many genes (perhaps as many as 1000) affect sleep (Harbison et al. 2009a,b). Harbison et al. (2009a) analyzed variation in sleep in 40 highly inbred lines of *D. melanogaster* and found many variable genes with only a few having large effects. The data suggest that, like mammals, regulation

of sleep in *Drosophila* is complex and controlled by multiple brain circuits involving sleep duration, waking activity, and number of sleep bouts. [Soshnev et al. \(2011\)](#) found that a conserved long noncoding RNA called *yar* affects length of sleep, disruption of sleep, and that it may regulate sleep by affecting stabilization or translational control of mRNAs.

## 11.8 High-Throughput Ethomics

The study of insect behavior can be very labor-intensive. A common method of analysis involves videotaping an insect's behavior, then analyzing the tape frame by frame. A camera-based system was developed to automatically quantify individual and group behaviors of *D. melanogaster* and to provide automated analyses of the data obtained ([Branson et al. 2009](#)). The system allows analysis of individuals or of groups of flies, allowing high-throughput screening. Another automated-monitoring system for analysis of *Drosophila* behavior was developed by [Dankert et al. \(2009\)](#). It also provides software that allows analysis of both single flies and groups of flies. [Reiser \(2009\)](#) hailed the development of these two systems that enable high-throughput screens that will allow scientists to resolve cellular and molecular underpinnings of behavior. The systems involve video recording of behavior, detection of flies in each frame, assignment of the trajectory, and classification of the behavior into an "ethogram." An ethogram is a catalog of the action patterns in an animal's behavioral repertoire. Both systems are freely distributed, and may be useful with insects other than *D. melanogaster*.

## 11.9 Systems Genetics of Complex Traits in *Drosophila*

The ability to sequence genomes relatively inexpensively has allowed analysis of complex behaviors in *D. melanogaster* ([Ayroles et al. 2009](#), [Edwards et al. 2009](#), [Mackay 2009](#)). What is system genetics? It is the use of *P*-element mutagenesis to identify genes affecting complex behaviors, artificial selection of natural populations to create extreme phenotypes, high-resolution mapping to identify candidate genes corresponding to quantitative trait loci (QTLs), and whole-genome transcriptional profiling to identify networks of interacting genes affecting complex traits ([Mackay 2009](#)). [Mackay \(2009\)](#) concluded that large numbers of loci affect behaviors in natural populations of *D. melanogaster*, including olfactory-avoidance behavior in response to a single odorant (97 mutations). In studies involving *P*-element insertions into the genome, 17% of the genome was involved in the effects of ethanol on *D. melanogaster*, and 34% of insertions affected locomotor behavior. Whole-genome transcript analyses of lines indicated that 530 genes were affected by a single mutation in *smell-impaired* loci.

Mackay concluded, "If a substantial fraction of the genome can affect any single trait, it follows that most genes must be pleiotropic and affect multiple traits."

[Edwards et al. \(2009\)](#) screened 170 *P*-element insertions for quantitative differences in aggressive behavior from their isogenic control line of *D. melanogaster* and identified 59 mutations in 57 genes that affect aggressive behavior, none of which previously were found to affect aggression. Among the 59 mutations, 32 resulted in increased aggression, and 27 lines were less aggressive than the control. [Edwards et al. \(2009\)](#) found, "Many of the genes affect the development and function of the nervous system ... Others affect basic cellular and metabolic processes. ..." The genes had pleiotropic effects on brain morphology.

[Dierick and Greenspan \(2006\)](#) analyzed aggressive behavior in *D. melanogaster* by selecting populations in a two-male arena assay. After 10 generations of selection, the aggressive lines became more aggressive and after 21 generations, the fighting index increased >30-fold. Microarray analysis indicated at least 42 genes were affected, but most of the expression changes were small, with only four genes showing an expression difference greater than two-fold. Six genes were then analyzed by quantitative PCR and five of the six had expression profiles that matched the microarray results. The differences in results obtained by [Edwards et al. \(2009\)](#) and [Dierick and Greenspan \(2006\)](#) are likely due to differences in experimental design.

## 11.10 Social Behavior in Bees and Ants

Honey bees and ants are eusocial insects and a great deal of controversy surrounds the evolution of sociality (e.g., [Nowak et al. 2010](#), [Bloch and Grozinger 2011](#)). Rapid advances in understanding the social life of *Apis mellifera* are being made now that the genome has been sequenced. For example, [Liang et al. \(2012\)](#) documented that scouting behavior for food and nest sites varies among workers and identified the genes involved in this behavior using whole-genome microarray analysis and quantitative reverse-transcriptase polymerase PCR. [Liang et al. \(2012\)](#) concluded, "Our results demonstrate intriguing parallels between honey bees and humans in novelty-seeking behavior. Although the molecular mechanisms that produce this behavioral variation are similar, it is unknown whether both species inherited them from a common ancestor or evolved them independently." [Whitfield et al. \(2006\)](#) analyzed the process by which workers mature from working in the hive to foraging using gene-expression microarrays of the bee brain.

[Jarosch et al. \(2011\)](#) found how Cape honey bees, *Apis mellifera capensis*, have modified their biology so that workers can reproduce by thelytoky,

producing diploid female progeny. The behavior appears to be caused by alternative splicing of a gene homologous to the *gemini* transcription factor of *Drosophila*. It appears that this switch can allow rapid worker ovary activation in Cape honey bees, turning the altruistic worker genome into a parasite.

Krieger and Ross (2002) identified a single major gene (*Gp-9*, encoding a pheromone-binding protein) that may affect fire ant (*Solenopsis invicta*) workers' ability to recognize queens and regulate their numbers. Some fire ant colonies have multiple egg-laying queens (polygyne form), whereas others have single queens (monogyne social form). Apparently, colony-queen number is associated with an allele of the *Gp-9* gene.

Lucas and Sokolowski (2009) studied the ant *Pheidole pallidula* and showed that the differences in major and minor workers are due to the ant *foraging* gene, which encodes a cGMP-dependent protein kinase. Majors, which are larger and defend the nest, have higher protein levels in five cells in the anterior face of the ant brain, whereas minors do not. Minors are involved in foraging and manipulating the level of the protein increases defense and reduces foraging behavior.

The descriptions mentioned above are a small sample of the wealth of information being obtained on the genetics of behavior in social insects. The complete genome sequences of bees and ants will allow even more detailed analysis of behaviors in the future.

## 11.11 Conclusions

Great advances have been made in understanding the behavior of insects using the many new molecular tools, especially those based on whole-genome sequencing. The ability to sequence the genomes of multiple inbred lines of *D. melanogaster*, for example, has provided exceptional opportunities to dissect the genetic bases of behavior. Statistical methods have advanced, and methods to evaluate groups of insects by recording their behavior and analyzing the data with computer programs provide new opportunities, as well, to obtain high-throughput data (Buchen 2009, Walsh 2009).

However, the report by Vanin et al. (2012) indicating that circadian behavior of *D. melanogaster* is different under natural conditions (populations contained in cages outdoors) than in the laboratory indicates great care must be taken to develop appropriate experimental methods in order to obtain a realistic understanding of an insect's behavior. Another example of the importance of experimental methods is the design of "choice tests." Martel and Boivin (2011) discuss

the appropriate design of choice tests in the laboratory and define the experimental methods that will distinguish "apparent choice" from "true choice", in which the true-choice behavior has to meet three criteria: exploitation of the resource is nonrandom, the chooser makes the same choice even in the absence of a differential response by the resource, and all resources are responded to, even in the absence of choice. Another issue to be resolved is how often gene function is conserved in behavior (Reaume and Sokolowski 2011). Clearly, the appropriate design and analysis of insect behavior remains a challenge.

## References Cited

- Allada, R., Emery, P., Takahashi, J.S., and Rosbash, M., (2001). Stopping time: the genetics of fly and mouse circadian clocks. *Annu. Rev. Neurosci.* **24**: 1091–1119.
- Alcock, J., (1984). *Animal Behavior: An Evolutionary Approach*, Sinauer, Sunderland, MA.
- Andretic, R., Chaney, S., and Hirsch, J., (1999). Requirement of circadian genes for cocaine sensitization in *Drosophila*. *Science* **285**: 1066–1068.
- Anholt, R.R.H., Fanara, J.J., Fedorowicz, G.M., Ganguly, I., Kulkarni, N.H., and Mackay, T.F.C., et al. (2001). Functional genomics of odor-guided behavior in *Drosophila melanogaster*. *Chem. Senses* **26**: 215–221.
- Arathi, H.S., Burns, I., and Spivak, M., (2000). Ethology of hygienic behaviour in the honey bee *Apis mellifera* L. (Hymenoptera: Apidae): behavioural repertoire of hygienic bees. *Ethology* **106**: 365–379.
- Arbuthnott, D., (2009). The genetic architecture of insect courtship behavior and premating isolation. *Heredity* **103**: 15–22.
- Arthur Jr., B.I., Jallon, J.M., Caflisch, B., Choffat, Y., and Nothiger, R., (1998). Sexual behaviour in *Drosophila* is irreversibly programmed during a critical period. *Curr. Biol.* **8**: 1187–1190.
- Atkinson, N.S., Robertson, G.A., and Ganetzky, B., (1991). A component of calcium-activated potassium channels encoded by the *Drosophila slo* locus. *Science* **253**: 551–555.
- Ayroles, J.F., Carbone, M.A., Stone, E.A., Jordan, K.W., Lyman, R.F., and Magwire, M.M., et al. (2009). Systems genetics of complex traits in *Drosophila melanogaster*. *Nat. Genet.* **41**: 299–307.
- Baker, B.S., Taylor, B.J., and Hall, J.C., (2001). Are complex behaviors specified by dedicated regulatory genes? Reasoning from *Drosophila*. *Cell* **105**: 13–24.
- Bainton, R.J., Tsai, L.T.Y., Singh, C.M., Moore, M.S., Neckameyer, W.S., and Heberlein, U., (2000). Dopamine modulates acute responses to cocaine, nicotine and ethanol in *Drosophila*. *Curr. Biol.* **10**: 187–194.
- Bargiello, T.A., Jackson, F.R., and Young, M.W., (1984). Restoration of circadian behavioural rhythms by gene transfer in *Drosophila*. *Nature* **312**: 752–754.
- Barnard, E.A., (1989). Molecular neurobiology--an introduction. In: Glover, D.M., and Hames, B.D., Eds., *Molecular Neurobiology*, IRL Press, Oxford, UK, pp. 1–7.
- Barron, A.B., and Corbet, S.A., (1999). Preimaginal conditioning in *Drosophila* revisited. *Anim. Behav.* **58**: 621–628.
- Barton Browne, L., (1993). Physiologically induced changes in resource-oriented behavior. *Annu. Rev. Entomol.* **38**: 1–25.
- Bazzett, T.J., (2008). *An Introduction to Behavior Genetics*. Sinauer, Sunderland, MA.
- Beck, S.D., (1980). *Insect Photoperiodism*, 2nd ed., Academic, New York.
- Bell, W.J., (1990). Searching behavior patterns in insects. *Annu. Rev. Entomol.* **35**: 447–467.
- Bellen, H.J., (1998). The fruit fly: a model organism to study the genetics of alcohol abuse and addiction? *Cell* **93**: 909–912.

- Bellen, H.J., Kooyer, S., D. D'Evelyn, and Pearlman, J., (1992). The *Drosophila* Couch potato protein is expressed in nuclei of peripheral neuronal precursors and shows homology to RNA-binding proteins. *Genes Dev.* **6**: 2125–2136.
- Bendena, W.G., (2010). Neuropeptide physiology in insects. In: Geary, T.G., and Maule, A.G., Eds., *Neuropeptide Systems as Targets for Parasite and Pest Control*, Landes Bioscience/Springer Science, Austin, TX, pp. 166–191.
- Benzer, S., (1973). Genetic dissection of behavior. *Sci. Am.* **229**: 24–27.
- Beukeboom, L.W., and van den Assem, J., (2001). Courtship and mating behaviour of interspecific *Nasonia* hybrids (Hymenoptera, Pteromalidae): a grandfather effect. *Behav. Genet.* **31**: 167–177.
- Bloch, G., and Grozinger, C.M., (2011). Social molecular pathways and the evolution of bee societies. *Philos. Trans. R. Soc. B* **366**: 2155–2170.
- Brady, J.P., and Richmond, R.C., (1990). Molecular analysis of evolutionary changes in the expression of *Drosophila* esterases. *Proc. Natl. Acad. Sci. USA* **87**: 8217–8221.
- Branson, K., Robie, A.A., Bender, J., Perona, P., and Dickinson, M.H., (2009). High-throughput ethomics in large groups of *Drosophila*. *Nat. Methods* **6**: 451–457.
- Buchen, L., (2009). Flies on film. *Nature* **462**: 562–564.
- Butler, A., Wei, A., Baker, K., and Salkoff, L., (1989). A family of putative potassium channel genes in *Drosophila*. *Science* **243**: 943–947.
- Butlin, R., and Ritchie, M.G., (2001). Searching for speciation genes. *Nature* **412**: 31–32.
- Carey, A.F., and Carlson, J.R., (2011). Insect olfaction from model systems to disease control. *Proc. Natl. Acad. Sci. USA* **108**: 12987–12995.
- Chen, C., Malone, T., Beckendorf, S.K., and Davis, R.L., (1987). At least two genes reside within a large intron of the *dunce* gene of *Drosophila*. *Nature* **329**: 721–724.
- Choi, K.W., Smith, R.F., Buratowski, R.M., and Quinn, W.G., (1991). Deficient protein kinase C activity in *turnip*, a *Drosophila* learning mutant. *J. Biol. Chem.* **266**: 15999–16006.
- Christophides, G.K., Mintzas, A.C., and Komitopoulou, K., (2000). Organization, evolution and expression of a multigene family encoding putative members of the odorant binding protein family in the medfly *Ceratitis capitata*. *Insect Mol. Biol.* **9**: 185–195.
- Clyne, P.J., Warr, C.G., and Carlson, J.R., (2000). Candidate taste receptors in *Drosophila*. *Science* **287**: 1830–1834.
- Colvin, J., and Gatehouse, A.G., (1993). Migration and genetic regulation of the pre-reproductive period in the cotton-bollworm moth, *Helicoverpa armigera*. *Heredity* **70**: 407–412.
- Costa, R.A.A., Peixoto, G., Barbujani, and Kyriacou, C.P., (1992). A latitudinal cline in a *Drosophila* clock gene. *Proc. R. Soc. Lond. B* **250**: 43–49.
- Covarrubias, M., Wei, A., and Salkoff, L., (1991). *Shaker*, *Shal*, *Shab*, and *Shaw* express independent K<sup>+</sup> current systems. *Neuron* **7**: 763–773.
- Dankert, H., Wang, L., Hoopfer, E.D., Anderson, D.J., and Perona, P., (2009). Automated monitoring and analysis of social behavior in *Drosophila*. *Nat. Methods* **6**: 297–303.
- Davis, R.L., (1993). Mushroom bodies and *Drosophila* learning. *Neuron* **11**: 1–14.
- Davis, R.L., (1996). Physiology and biochemistry of *Drosophila* learning mutants. *Physiol. Rev.* **76**: 299–317.
- Davis, R.L., and Dauwalder, B., (1991). The *Drosophila* *dunce* locus: learning and memory genes in the fly. *Trends Genet.* **7**: 224–229.
- de Belle, J.S., and Sokolowski, M.B., (1987). Heredity of rover/sitter: alternative foraging strategies of *Drosophila melanogaster* larvae. *Heredity* **59**: 73–83.
- Delgado, R., Latorre, R., and Labarca, P., (1992). K<sup>+</sup>-channel blockers restore synaptic plasticity in the neuromuscular junction of *dunce*, a *Drosophila* learning and memory mutant. *Proc. R. Soc. Lond. B* **250**: 181–185.
- Desjardins, C.A., Perfectti, F., Bartos, J.D., Enders, L.S., and Werren, J.H., (2010). The genetic basis of interspecies host preference differences in the model parasitoid *Nasonia*. *Heredity* **104**: 270–277.

- Dierick, H.A., and Greenspan, R.J., (2006). Molecular analysis of flies selected for aggressive behavior. *Nat. Genet.* **38**: 1023–1031.
- Dingle, H., Ed., (1989). *Evolution of Insect Migration and Diapause*, Springer-Verlag, New York.
- Dingle, H., and Hegmann, J.P., Eds., (1982). *Evolution and Genetics of Life Histories*, Springer-Verlag, New York.
- Dion, E., Polin, S.E., Simon, J.C., and Outreman, Y., (2011). Symbiont infection affects aphid defensive behaviours. *Biol. Lett.* **7**: 743–746.
- Doerge, R.W., (2002). Mapping and analysis of quantitative trait loci in experimental populations. *Nat. Rev. Genet.* **3**: 43–52.
- Donlea, J.M., Thimgan, M.S., Suzuki, Y., Gottschalk, L., and Shaw, P.J., (2011). Inducing sleep by remote control facilitates memory consolidation in *Drosophila*. *Science* **332**: 1571–1576.
- Donlea, J., Leahy, A., Thimgan, M.S., Suzuki, Y., Hughson, B.N., and Sokolowski, M.B., et al. (2012). Foraging alters resilience/vulnerability to sleep disruption and starvation in *Drosophila*. *Proc. Natl. Acad. Sci. USA* **109**: 1618–2613.
- Doi, M., Matsuda, M., Tomaru, M., Matsabayasi, H., and Oguma, Y., (2001). A locus for female discrimination behavior causing sexual isolation in *Drosophila*. *Proc. Natl. Acad. Sci. USA* **98**: 6714–6719.
- Dryer, L., (2000). Evolution of odorant receptors. *Bioessays* **22**: 803–810.
- Dubnau, J., and Tully, T., (1998). Gene discovery in *Drosophila*: new insights for learning and memory. *Annu. Rev. Neurosci.* **21**: 407–444.
- Dunlap, J.C., (1998). Common threads in eukaryotic circadian systems. *Curr. Opin. Genet. Dev.* **8**: 400–406.
- Dunlap, J.C., (1999). Molecular bases for circadian clocks. *Cell* **96**: 271–290.
- Egan, E.S., Franklin, T.M., Hilderbrand-Chae, M.J., McNeil, G.P., Roberts, M.A., and Schroeder, A.J., et al. (1999). An extraretinally expressed insect cryptochrome with similarity to the blue light photoreceptors of mammals and plants. *J. Neurosci.* **19**: 366–3673.
- Edwards, A.C., Zwarts, L., Yamamoto, A., Callaerts, P., and Mackay, T.F.C., (2009). Mutations in many genes affect aggressive behavior in *Drosophila melanogaster*. *BMC Biol.* **7**: 29. doi: [10.1186/1741-7007-7-29](https://doi.org/10.1186/1741-7007-7-29).
- Ehrman, L., and Parsons, P.A., (1973). *The Genetics of Behavior*, Sinauer, Sunderland, MA.
- Ehrman, L., and Parsons, P.A., (1981). *Behavior Genetics and Evolution*, McGraw-Hill, New York.
- Emery, P., Stanewsky, R., Helfric-Forster, C., Emery-Le, M., Hall, J.C., and Rosbash, M., (2000). *Drosophila* CRY is a deep brain circadian photoreceptor. *Neuron* **26**: 493–504.
- Falconer, D.S., (1989). *Introduction to Quantitative Genetics*, 3rd Edition, Longman, New York.
- Feany, M.B., (2000). Studying human neurodegenerative diseases in flies and worms. *J. Neuropathol. Exp. Neurol.* **59**: 847–856.
- Feany, M.B., and Bender, W.W., (2000). A *Drosophila* model of Parkinson's disease. *Nature* **404**: 394–398.
- Ferveur, J.F., (1997). The pheromonal role of cuticular hydrocarbons in *Drosophila melanogaster*. *Bioessays* **19**: 353–358.
- Ferveur, J.F., and Jallon, J.M., (1993). Genetic control of pheromones in *Drosophila simulans*. II. kete a locus on the X chromosome. *Genetics* **133**: 561–567.
- Field, L.M., Pickett, J.A., and Wadhams, L.J., (2000). Molecular studies in insect olfaction. *Insect Mol. Biol.* **9**: 545–551.
- Folkers, E., Drain, P., and Quinn, W.G., (1993). *radish*, a *Drosophila* mutant deficient in consolidated memory. *Proc. Natl. Acad. Sci. USA* **90**: 8123–8127.
- Fortini, M.E., and Bonini, N.M., (2000). Modeling human neurodegenerative diseases in *Drosophila*. On a wing and a prayer. *Trends Genet.* **16**: 161–167.
- Foster, S.P., and Harris, M.O., (1997). Behavioral manipulation methods for insect pest-management. *Annu. Rev. Entomol.* **42**: 123–146.

- Fournier, D., Karch, F., Bride, J.M., Hall, L.M.C., Berge, J.B., and Spierer, P., (1989). *Drosophila melanogaster* acetylcholinesterase gene structure, evolution and mutations. *J. Mol. Biol.* **210**: 15–22.
- Gailey, D.A., Villella, A., and Tully, T., (1991). Reassessment of the effect of biological rhythm mutations on learning in *Drosophila melanogaster*. *J. Comp. Physiol. A* **169**: 685–697.
- Ganetzky, B., and Wu, C.F., (1989). Molecular approaches to neurophysiology in *Drosophila*. In: Glover, D.M., and Hames, B.D., Eds., *Molecular Neurobiology*, IRL Press, Oxford, UK, pp. 9–61.
- Ganter, G.K., Panaitiu, A.E., Desilets, J.B., Davis-Heim, J.A., Fisher, E.A., and Tan, L.C.H., et al. (2011). *Drosophila* male courtship behavior is modulated by ecdysteroids. *J. Insect Physiol.* **57**: 1179–1184.
- Gatehouse, A.G., (1989). Genes, environment and insect flight. In: Goldsworthy, G.J., and Wheeler, C., Eds., *Insect Flight*, CRC Press, Boca Raton, FL, pp. 116–138.
- Gatehouse, A.G., (1997). Behavior and ecological genetics of wind-borne migration by insects. *Annu. Rev. Entomol.* **42**: 475–502.
- Giebultowicz, J.M., (2000). Molecular mechanism and cellular distribution of insect circadian clocks. *Annu. Rev. Entomol.* **45**: 769–793.
- Glover, D.M., and Hames, B.D., Eds., (1989). *Molecular Neurobiology*, IRL Press, Oxford, UK.
- Goodwin, S.F., (1999). Molecular neurogenetics of sexual differentiation and behaviour. *Curr. Opin. Neurobiol.* **9**: 759–765.
- Goodwin, S.F., Taylor, B.J., Villella, A., Foss, M., Ryner, L.C., Baker, B.S., and Hall, J.C., (2000). Aberrant splicing and altered spatial expression patterns in *fruitless* mutants of *Drosophila melanogaster*. *Genetics* **154**: 725–745.
- Gould, F., (1986). Genetic constraints on the evolution of cannibalism in *Heliothis virescens*. In: Huettel, M.D., Ed., *Evolutionary Genetics of Invertebrate Behavior, Progress and Prospects*, Plenum, New York, pp. 55–62.
- Gould, F., Estock, M., Hillier, N.K., Powell, B., Groot, A.T., and Ward, C.M., et al. (2010). Sexual isolation of male moths explained by a single pheromone response QTL containing four receptor genes. *Proc. Natl. Acad. Sci. USA* **107**: 8660–8665.
- Greenspan, R.J., and Ferveur, J.F., (2000). Courtship in *Drosophila*. *Annu. Rev. Genet.* **34**: 205–232.
- Greenspan, R.J., Tononi, G., Cirelli, C., and Shaw, P.J., (2001). Sleep and the fruit fly. *Trends Neurosci.* **24**: 142–145.
- Grossfield, J., (1975). Behavioral mutants of *Drosophila*. In: King, R.C., Ed., *Handbook of Genetics*. Vol. 3. *Invertebrates of Genetic Interest*, Plenum, New York, pp. 679–702.
- Grotewiel, M.S., Beck, C.D.O., Wu, K.H., Zhu, X.R., and Davis, R.L., (1998). Integrin-mediated short-term memory in *Drosophila*. *Nature* **391**: 455–460.
- Grunbaum, L., and Muller, U., (1998). Induction of a specific olfactory memory leads to a long-lasting activation of protein kinase C in the antennal lobe of the honeybee. *J. Neurosci.* **18**: 4384–4392.
- Guantieri, V., Pepe, A., Zordan, M., Kyriacou, C.P., Costa, R., and Tamburro, A.M., (1999). Different period gene repeats take ‘turns’ at fine-tuning the circadian clock. *Proc. R. Soc. Lond. B* **266**: 2283–2288.
- Hall, J.C., (1984). Complex brain and behavioral functions disrupted by mutations in *Drosophila*. *Dev. Genet.* **4**: 355–532.
- Hall, J.C., (1985). Genetic analysis of behavior in insects. In: Kerkut, G.A., and Gilbert, L.I., Eds., *Comprehensive Insect Physiology Biochemistry and Pharmacology*, Volume 9, Pergamon, New York, pp. 287–373.
- Hall, J.C., (1994). The mating of a fly. *Science* **264**: 1702–1714.
- Hall, J.C., (1995). Tripping along the trail to the molecular mechanisms of biological clocks. *Trends Neurosci.* **5**: 230–240.
- Hall, J.C., (1998a). Genetics of biological rhythms in *Drosophila*. *Adv. Genet.* **38**: 135–184.

- Hall, J.C., (1998b). Molecular neurogenetics of biological rhythms. *J. Neurogenet.* **12**: 115–181.
- Hammer, M., and Menzel, R., (1995). Learning and memory in the honeybee. *J. Neurosci.* **15**: 1617–1630.
- Han, P.L., Levin, L.R., Reed, R.R., and Davis, R.L., (1992). Preferential expression of the *Drosophila rutabaga* gene in mushroom bodies, neural centers for learning in insects. *Neuron* **9**: 619–627.
- Hansson, B.S., and Anton, S., (2000). Function and morphology of the antennal lobe: new developments. *Annu. Rev. Entomol.* **45**: 203–231.
- Harbison, S.T., Carbone, M.A., Ayroles, J.F., Stone, E.A., Lyman, R.F., and Mackay, T.F.C., (2009a). Co-regulated transcriptional networks contribute to natural genetic variation in *Drosophila* sleep. *Nat. Genet.* **41**: 371–375.
- Harbison, S.T., Mackay, T.F.C., and Anholt, R.R.H., (2009b). Understanding the neurogenetics of sleep: progress from *Drosophila*. *Trends Genet.* **25**: 262–269.
- Hardin, P.E., Hall, J.C., and Rosbash, M., (1992). Circadian oscillations in *period* gene mRNA levels are transcriptionally regulated. *Proc. Natl. Acad. Sci. USA* **89**: 11711–11715.
- Harmer, S.L., Panda, S., and Kay, S.A., (2001). Molecular bases of circadian rhythms. *Annu. Rev. Cell Dev. Biol.* **17**: 215–253.
- Hay, D.A., (1985). *Essentials of Behaviour-Genetics*, Blackwell, Palo Alto, CA.
- Heisenberg, M., (1997). Genetic approach to neuroethology. *Bioessays* **19**: 1065–1073.
- Heisenberg, M., (1998). What do the mushroom bodies do for the insect brain? An introduction. *Learn. Mem.* **5**: 1–10.
- Hendricks, J.C., Finn, S.M., Panckeri, K.A., Chavkin, J., Williams, J.A., Sehgal, A., and Pack, A.I., (2000). Rest in *Drosophila* is a sleep-like state. *Neuron* **24**: 129–138.
- Hoedjes, K.M., Kruidhof, H.M., Huigens, M.E., Dicke, M., Vet, L.E.M., and Smid, H.M., (2011). Natural variation in learning rate and memory dynamics in parasitoid wasps: opportunities for converging ecology and neuroscience. *Proc. R. Soc. London B* **278**: 889–897.
- Hogenesch, J.B., and Ueda, H.R., (2011). Understanding systems-level properties: timely stories from the study of clocks. *Nature Rev. Genetics* **12**: 407–416.
- Holman, G.M., Nachman, R.J., and Wright, M.S., (1990). Insect neuropeptides. *Annu. Rev. Entomol.* **35**: 201–217.
- Homberg, U., Christensen, T.A., and Hildebrand, J.G., (1989). Structure and function of the deutocerebrum in insects. *Annu. Rev. Entomol.* **34**: 477–501.
- Hoover, K., Grove, M., Gardner, M., Hughes, D.P., McNeil, J., and Slavicek, J., (2011). A gene for an extended phenotype. *Science* **333**: 1401.
- Hotta, Y., and Benzer, S., (1972). Mapping of behaviour in *Drosophila* mosaics. *Nature* **240**: 527–535.
- Huber, F., Moore, T.E., and Loher, W., Eds., (1989). *Cricket Behavior and Neurobiology*, Cornell University Press, Ithaca, NY.
- Huettel, M.D., Ed., (1986). *Evolutionary Genetics of Invertebrate Behavior. Progress and Prospects*, Plenum, New York.
- Huettel, M.D., and Bush, G.L., (1972). The genetics of host selection and its bearing on sympatric speciation in *Procecidochares* (Diptera: Tephritidae). *Entomol. Exp. Appl.* **15**: 465–480.
- Ikeno, T., Numata, H., and Goto, S.G., (2011). Circadian clock genes period and cycle regulate photoperiodic diapause in the bean bug *Riptortus pedestris* males. *J. Insect Physiol.* **57**: 935–938.
- Immonen, E., and Ritchie, M.G., (2012). The genomic response to courtship song stimulation in female *Drosophila melanogaster*. *Proc. R. Soc. B* **279**: 1359–11365.
- Ishida, N., Kaneko, M., and Allada, R., (1999). Biological clocks. *Proc. Natl. Acad. Sci. USA* **96**: 8819–8820.
- Ito, H., Sato, K., Koganezawa, M., Ote, M., Matsumoto, K., Hama, C., and Yamamoto, D., (2012). Fruitless recruits two antagonistic chromatin factors to establish single-neuron sexual dimorphism. *Cell* **149**: 1327–1338.

- Iwasaki, K., and Thomas, J.H., (1997). Genetics in rhythm. *Trends Genet.* **13**: 111–115.
- Jackson, F.R., Newby, L.M., and Kulkarni, S.J., (1990). *Drosophila* GABAergic systems: sequence and expression of glutamic acid decarboxylase. *J. Neurochem.* **54**: 1068–1078.
- Jaenike, J., (1986). Genetic complexity of host selection behavior in *Drosophila*. *Proc. Natl. Acad. Sci. USA* **83**: 2148–2151.
- Jarosch, A., Stolle, E., Crewe, R.M., and Moritz, R.F.A., (2011). Alternative splicing of a single transcription factor drives selfish reproductive behavior in honeybee workers (*Apis mellifera*). *Proc. Natl. Acad. Sci. USA* **108**: 15282–15287.
- Kalmring, K., and Elsner, N., Eds., (1985). *Acoustic and Vibrational Communication in Insects*, Paul Parey, Hamburg, Germany.
- Kernan, M.J., Kuroda, M.I., Kreber, R., Baker, B.S., and Ganetzky, B., (1991). *nap<sup>ts</sup>*, a mutation affecting sodium channel activity in *Drosophila*, is an allele of *mle*, a regulator of X chromosome transcription. *Cell* **66**: 949–959.
- Klun, J.A., and Huettel, M.D., (1988). Genetic regulation of sex pheromone production and response: interaction of sympatric pheromonal types of European corn borer, *Ostrinia nubilalis* (Lepidoptera: Pyralidae). *J. Chem. Ecol.* **13**: 2047–2061.
- Klun, J.A., and Maini, S., (1979). Genetic basis of an insect chemical communication system. *Environ. Entomol.* **8**: 423–426.
- Kolker, D.E., and Turek, F.W., (1999). The search for circadian clock and sleep genes. *J. Psychopharmacol.* **13** (4 Suppl. 1): S5–S9.
- Kostal, V., and Shimada, K., (2001). Malfunction of circadian clock in the non-photoperiodic-diapause mutants of the drosophilid fly, *Chymomyza costata*. *J. Insect Physiol.* **47**: 1269–1274.
- Krieger, J., and Breer, H., (1999). Olfactory reception in invertebrates. *Science* **286**: 720–728.
- Krieger, M.J.B., and Ross, K.G., (2002). Identification of a major gene regulating complex social behavior. *Science* **295**: 328–332.
- Kyriacou, C.P., (1990). The molecular ethology of the *period* gene in *Drosophila*. *Behav. Genet.* **20**: 191–211.
- Kyriacou, C.P., (1993). Molecular genetics of complex behavior: biological rhythms in *Drosophila*. In: Oakeshott, J., and Whitten, M.J., Eds., *Molecular Approaches to Fundamental and Applied Entomology*, Springer-Verlag, New York, pp. 329–356.
- Kyriacou, C.P., and Hall, J.C., (1989). Circadian rhythm mutations in *Drosophila melanogaster* affect short-term fluctuations in the male's courtship song. *Proc. Natl. Acad. Sci. USA* **77**: 6729–6733.
- Lakin-Thomas, P., (2000). Circadian rhythms. New functions for old clock genes? *Trends Genet.* **16**: 135–142.
- Liang, Z.S., Nguyen, T., Mattila, H.R., Rodriguez-Zas, S.L., Seeley, T.D., and Robinson, G.E., (2012). Molecular determinants of scouting behavior in honey bees. *Science* **235**: 1225–1228.
- Liu, X., Zwiebel, L.J., Hinton, D., Benzer, S., Hall, J.C., and Rosbash, M., (1992). The *period* gene encodes a predominantly nuclear protein in adult *Drosophila*. *J. Neurosci.* **12**: 2735–2744.
- Löfstedt, C., (1990). Population variation and genetic control of pheromone communication systems in moths. *Entomol. Exp. Appl.* **54**: 199–218.
- Löfstedt, C., Hansson, B.S., Roelofs, W., and Bengtsson, B.O., (1989). No linkage between genes controlling female pheromone production and male pheromone response in the European corn borer, *Ostrinia nubilalis* Hübner (Lepidoptera: Pyralidae). *Genetics* **123**: 553–556.
- Loudon, A.S.I., Semikhodskii, A.G., and Crosthwaite, S.K., (2000). A brief history of circadian time. *Trends Genet.* **16**: 477–481.
- Loughney, K., Kreber, R., and Ganetzky, B., (1989). Molecular analysis of the *para* locus, a sodium channel gene in *Drosophila*. *Cell* **58**: 1143–1154.
- Lucas, C., and Sokolowski, M.B., (2009). Molecular basis for changes in behavioral state in ant social behaviors. *Proc. Natl. Acad. Sci. USA* **106**: 6351–6356.

- Luo, L., Tully, T., and White, K., (1992). Human amyloid precursor protein ameliorates behavioral deficit of flies deleted for *Appl* gene. *Neuron* **9**: 595–605.
- Mackay, T.F.C., (2009). The genetic architecture of complex behaviors: lessons from *Drosophila*. *Genetics* **136**: 295–302.
- Majumdar, A., Cesario, W.C., White-Grindley, E., Jiang, H., Ren, F., and Khan, M., et al. (2012). Critical role of amyloid-like oligomers of *Drosophila* Orv2 in the persistence of memory. *Cell* **148**: 515–529.
- Maleszka, R., (2000). Molecules to behaviour in the honeybee—the emergence of comparative neurogenomics. *Trends Neurosci.* **23**: 513–514.
- Markov, A.V., Lazebny, O.E., Goryacheva, I.I., Antipin, M.I., and Kulikov, A.M., (2009). Symbiotic bacteria affect mating choice in *Drosophila melanogaster*. *Anim. Behav.* **77**: 1011–1017.
- Martel, V., and Boivin, G., (2011). Do choice tests really test choice? *J. Insect Behav.* **24**: 329–336.
- Masler, E.P., Kelly, T.J., and Menn, J.J., (1993). Insect neuropeptides: discovery and application in insect management. *Arch. Insect Biochem. Physiol.* **22**: 87–111.
- Masterman, R., Ross, R., Mesce, K., and Spivak, M., (2001). Olfactory and behavioral response thresholds to odors of diseased brood differ between hygienic and non-hygienic honey bees (*Apis mellifera* L.). *J. Comp. Physiol. A* **187**: 441–452.
- Matthews, R.W., and Matthews, J.R., (1978). *Insect Behavior*, Wiley Interscience, New York.
- McCabe, E.R.B., (1995). Modeling Lou Gehrig's disease in the fruit fly. *Proc. Natl. Acad. Sci. USA* **92**: 8533–8534.
- McGuire, T.R., (1984). Learning in three species of Diptera: the blow fly, *Phormia regina*, the fruit fly, *Drosophila melanogaster*, and the house fly, *Musca domestica*. *Behav. Genet.* **14**: 479–526.
- McGuire, T.R., and Hirsch, J., (1977). Behavior-genetic analysis of *Phormia regina*: conditioning, reliable individual differences, and selection. *Proc. Natl. Acad. Sci. USA* **74**: 5193–5197.
- McGuire, T.R., and Tully, T., (1987). Characterization of genes involved with classical conditioning that produce differences between bidirectionally selected strains of the blow fly *Phormia regina*. *Behav. Genet.* **17**: 97–107.
- Meller, V.H., and Davis, R.L., (1996). Biochemistry of insect learning: lessons from bees and flies. *Insect. Biochem. Mol. Biol.* **26**: 327–335.
- Menzel, R., (1999). Memory dynamics in the honey bee. *J. Comp. Physiol. A* **185**: 323–340.
- Merrow, M., and Roenneberg, T., (2001). The circadian cycle: is the whole greater than the sum of its parts? *Trends Genet.* **17**: 4–7.
- Miller, C., (1991). 1990: Annus mirabilis of potassium channels. *Science* **251**: 1092–1096.
- Min, K.T., and Benzer, S., (1997). Spongecake and eggroll: two hereditary diseases in *Drosophila* resemble patterns of human brain degeneration. *Curr. Biol.* **7**: 885–888.
- Min, K.T., and Benzer, S., (1999). Preventing neurodegeneration in the *Drosophila* mutant bubblegum. *Science* **284**: 1985–1988.
- Mollema, C., (1991). Heritability estimates of host selection behaviour by the parasitoid *Asobara tabida*. *Neth. J. Zool.* **41**: 174–183.
- Mombaerts, P., (1999). Seven-transmembrane proteins as odorant and chemosensory receptors. *Science* **286**: 707–711.
- Moore, M.S., DeZazzo, J., Luk, A.Y., Tully, T., Singh, C.M., and Heberlein, U., (1998). Ethanol intoxication in *Drosophila*: genetic and pharmacological evidence for regulation by the cAMP signaling pathway. *Cell* **93**: 997–1007.
- Moretto, G., Goncalves, L.S., and De Jong, D., (1993). Heritability of Africanized and European honey bee defensive behavior against the mite *Varroa jacobsoni*. *Rev. Bras. Genet.* **16**: 71–77.
- Moritz, R.F.A., (1988). A reevaluation of the two-locus model for hygienic behavior in honey bees (*Apis mellifera* L.). *J. Hered.* **79**: 257–262.
- Moritz, R.F.A., (1994). Selection for varroatosis resistance in honey bees. *Parasitol. Today* **10**: 236–238.

- Mutsuddi, M., and Nambu, J.R., (1998). Neural disease: *Drosophila* degenerates for a good cause. *Curr. Biol.* **8**: R809–R811.
- Nakayama, S., Sasaki, K., Matsumura, K., Lewis, Z., and Miyatake, T., (2012). Dopaminergic system as the mechanism underlying personality in a beetle. *J. Insect Physiol.* **58**: 750–755.
- Nassel, D.R., (1993). Neuropeptides in the insect brain: a review. *Cell Tissue Res.* **273**: 1–29.
- Nassel, D.R., (2002). Neuropeptides in the nervous system of *Drosophila* and other insects: multiple roles as neuromodulators and neurohormones. *Prog. Neurobiol.* **68**: 1–84.
- Nowak, M.A., Tarnita, C.E., and Wilson, E.O., (2010). The evolution of eusociality. *Nature* **466**: 1057–1062.
- O’Kane, C.J., and Gehring, W.J., (1987). Detection *in situ* of genomic regulatory elements in *Drosophila*. *Proc. Natl. Acad. Sci. USA* **84**: 9123–9127.
- Osborne, K.A., Robichon, A., Burgess, E., Butland, S., Shaw, R.A., Coulthard, A., Pereira, H.S., Greenspan, R.J., and Sokolowski, M.B., (1997). Natural behavior polymorphism due to a cGMP-dependent protein kinase of *Drosophila*. *Science* **277**: 834–836.
- Palmer, J.O., and Dingle, H., (1989). Responses to selection on flight behavior in a migratory population of milkweed bug (*Oncopeltus fasciatus*). *Evolution* **43**: 1805–1808.
- Papaj, D.R., and Prokopy, R.J., (1989). Ecological and evolutionary aspects of learning in phytophagous insects. *Annu. Rev. Entomol.* **34**: 315–350.
- Papazian, D.M., Schwarz, T.L., Tempel, B.L., Jan, Y.N., and Jan, L.Y., (1987). Cloning of genomic and complementary DNA from *Shaker*, a putative potassium channel gene from *Drosophila*. *Science* **237**: 749–753.
- Pavlidis, P., Ramaswami, M., and Tanouye, M.A., (1994). The *Drosophila* easily shocked gene: a mutation in a phospholipid synthetic pathway causes seizure, neuronal failure, and paralysis. *Cell* **79**: 23–33.
- Pereira, H.S., MacDonald, D.E., Hilliker, A.J., and Sokolowski, M.B., (1995). *Chaser* (*Csr*), a new gene affecting larval foraging behavior in *Drosophila melanogaster*. *Genetics* **141**: 263–270.
- Phillips, J.P., Tainr, J.A., Getzoff, E.D., Boulian, G.L., Kirby, K., and Hilliker, A.J., (1995). Subunit-destabilizing mutations in *Drosophila* copper/zinc superoxide dismutase: neuropathology and a model of dimer dysequilibrium. *Proc. Natl. Acad. Sci. USA* **92**: 8574–8578.
- Plomin, R., (1990). The role of inheritance in behavior. *Science* **248**: 183–188.
- Prestwich, G.D., (1996). Proteins that smell: pheromone recognition and signal transduction. *Bioorg. Med. Chem.* **4**: 505–513.
- Qiu, Y., Chen, C.N., Malone, T., Richter, L., Beckendorf, S.K., and Davis, R.L., (1991). Characterization of the memory gene *dunce* of *Drosophila melanogaster*. *J. Mol. Biol.* **222**: 553–565.
- Raina, A.K., and Menn, J.J., (1993). Pheromone biosynthesis activating neuropeptide: from discovery to current status. *Arch. Insect Biochem. Physiol.* **22**: 141–151.
- Ray, S., and Ferneyhough, B., (1999). Behavioral development and olfactory learning in the honeybee (*Apis mellifera*). *Dev. Psychobiol.* **34**: 21–27.
- Reaume, C.J., and Sokolowski, M.B., (2011). Conservation of gene function in behaviour. *Philos. Trans. R. Soc. B* **366**: 2100–2110.
- Reiser, M., (2009). The ethomics era? *Nat. Methods* **6**: 413–414.
- Renou, M., and Guerrero, A., (2000). Insect parapheromones in olfaction research and semiochemical-based pest control strategies. *Annu. Rev. Entomol.* **45**: 605–630.
- R’Kha, S., Capy, P., and David, J.R., (1991). Host-plant specialization in the *Drosophila melanogaster* species complex: a physiological, behavioral, and genetical analysis. *Proc. Natl. Acad. Sci. USA* **88**: 1835–1839.
- Richmond, R.C., Mane, S.D., and Tompkins, L., (1986). The behavioral effects of a carboxylesterase in *Drosophila*. In: Huettel, M.D., Ed., *Evolutionary Genetics of Invertebrate Behavior, Progress and Prospects*, Plenum, New York, pp. 223–236.

- Rinderer, T.E., Ed., (1986). *Bee Genetics and Breeding*, Academic, Orlando.
- Rinderer, T.E., and Collins, A.M., (1986). Behavioral genetics. In: Rinderer, T.E., Ed., *Bee Genetics and Breeding*, Academic, Orlando, FL, pp. 155–176.
- Rothenbuhler, W.C., (1964). Behavior genetics of nest cleaning in honey bees. IV. Responses of F1 and backcross generations to disease-killed brood. *Am. Zool.* **4**: 11–123.
- Robertson, H.M., and Wanner, K.W., (2006). The chemoreceptor superfamily in the honey bee, *Apis mellifera*: expansion of the odorant, but not gustatory, receptor family. *Genome Res.* **16**: 1395–1403.
- Rutzler, M., and Zwiebel, L.J., (2005). Molecular biology of insect olfaction: recent progress and conceptual models. *J. Comp. Physiol. A* **191**: 777–790.
- Rybak, J., and Menzel, R., (1993). Anatomy of the mushroom bodies in the honey bee brain: the neuronal connections of the alpha-lobe. *J. Comp. Neurol.* **334**: 444–465.
- Ryner, L.C., Goodwin, S.F., Castrillon, D.H., Anand, A., Villella, A., and Baker, B.S., et al. (1996). Control of male sexual behavior and sexual orientation in *Drosophila* by the *fruitless* gene. *Cell* **87**: 1079–1089.
- Salkoff, L., Butler, A., Wei, A., Scavarda, N., Giffen, K., and Ifune, C., et al. (1987). Genomic organization and deduced amino acid sequence of a putative sodium channel gene in *Drosophila*. *Science* **237**: 744–753.
- Salkoff, L., Baker, K., Butler, A., Covarrubias, M., Pak, M.D., and Wei, A., (1992). An essential 'set' of K<sup>+</sup> channels conserved in flies, mice and humans. *Trends Neurosci.* **15**: 161–166.
- Sancar, A., (2000). CRYPTOCHROME: The second photoactive pigment in the eye and its role in circadian photoreception. *Annu. Rev. Biochem.* **69**: 1–67.
- Sandrelli, F., Costa, R., Kyriacou, C.P., and Rosato, E., (2008). Comparative analysis of circadian clock genes in insects. *Insect Mol. Biol.* **17**: 447–463.
- Sato, Y., Oguchi, M., Menjo, N., Imai, K., Saito, H., Ideda, M., Isobe, M., and Amashita, O., (1993). Precursor polyprotein for multiple neuropeptides secreted from the suboesophageal ganglion of the silkworm *Bombyx mori*: characterization of the cDNA encoding the diapause hormone precursor and identification of additional peptides. *Proc. Natl. Acad. Sci. USA* **90**: 3251–3255.
- Saunders, D.S., (1990). The circadian basis of ovarian diapause regulation in *Drosophila melanogaster*: is the *period* gene causally involved in photoperiodic time measurement? *J. Biol. Rhythms* **5**: 315–331.
- Savarit, F., Sureau, G., Cobb, M., and Ferveur, J.F., (1999). Genetic elimination of known pheromones reveals the fundamental bases of mating and isolation in *Drosophila*. *Proc. Natl. Acad. Sci. USA* **96**: 9015–9020.
- Scharrer, B., (1987). Insects as models in neuroendocrine research. *Annu. Rev. Entomol.* **32**: 1–16.
- Schuett, W., Dall, S.R.X., Baeumer, J., Kloesener, M.H., Nakagawa, S., Beinlich, F., and Eggers, T., (2011). "Personality" variation in a clonal insect: the pea aphid, *Acyrtosiphon pisum*. *Dev. Psychobiol.* **53**: 631–640.
- Sehgal, A., Price, J., and Young, M.W., (1992). Ontogeny of a biological clock in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* **89**: 1423–1427.
- Shaw, P.J., and Franken, P., (2003). Perchance to dream: solving the mystery of sleep through genetic analysis. *J. Neurobiol.* **54**: 179–202.
- Sharon, G., Segal, D., Ringo, J.M., Hefetz, A., Zilber-Rosenberg, I., and Rosenberg, E., (2010). Commensal bacteria play a role in mating preference of *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* **107**: 20051–20056.
- Shaver, S.A., Varnam, C.J., Hilliker, A.J., and Sokolowski, M.B., (1998). The foraging gene affects adult but not larval olfactory-related behavior in *Drosophila melanogaster*. *Behav. Brain Res.* **95**: 23–29.
- Singh, C.M., and Heberlein, U., (2000). Genetic control of acute ethanol-induced behaviors in *Drosophila*. *Alcoholism: Clin. Exp. Res.* **24**: 1127–1136.

- Smadja, C., Shi, P., Butlin, R.K., and Robertson, H.M., (2009). Large gene family expansions and adaptive evolution for odorant and gustatory receptors in the pea aphid, *Acyrthosiphon pisum*. *Mol. Biol. Evol.* **26**: 2073–2086.
- Sokolowski, M.B., (2001). *Drosophila*: genetics meets behaviour. *Nat. Rev. Genet.* **2**: 879–890.
- Sokolowski, M.B., Pereira, H.S., and Hughes, K., (1997). Evolution of foraging behavior in *Drosophila* by density-dependent selection. *Proc. Natl. Acad. Sci. USA* **94**: 7373–7377.
- Soshnev, A.A., Ishimoto, H., McAllister, B.F., Li, X., Wehling, M.D., Kitamoto, T., and Geyer, P.K., (2011). A conserved long noncoding RNA affects sleep behavior in *Drosophila*. *Genetics* **189**: 455–468.
- Spivak, M., and Gilliam, M., (1993). Facultative expression of hygienic behaviour of honey bees in relation to disease resistance. *J. Apic. Res.* **32**: 147–157.
- Spivak, M., and Gilliam, M., (1998). Hygienic behaviour of honey bees and its application for control of brood diseases and varroa. II. Studies on hygienic behaviour since the Rothenbuhler era. *Bee World* **79**: 169–186.
- Strausfeld, N.J., (1976). *Atlas of an Insect Brain*, Springer-Verlag, Berlin, Germany.
- Strausfeld, N.J., Hansen, L., Li, Y., Gomez, R.S., and Ito, K., (1998). Evolution, discovery, and interpretations of arthropod mushroom bodies. *Learn. Mem.* **5**: 11–37.
- Suzuki, K., Juni, N., and Yamamoto, D., (1997). Enhanced mate refusal in female *Drosophila* induced by a mutation in the *spinster* locus. *Appl. Entomol. Zool.* **32**: 235–243.
- Takahashi, J.S., (1992). Circadian clock genes are ticking. *Science* **258**: 238–239.
- Takeda, M., and Skopik, S.D., (1997). Photoperiodic time measurement and related physiological mechanisms in insects and mites. *Annu. Rev. Entomol.* **42**: 323–349.
- Tauber, M.J., Tauber, C.A., and Masaki, S., (1986). *Seasonal Adaptations of Insects*, Oxford University Press, Oxford, UK.
- Tazima, Y., Doira, H., and Akai, H., (1975). The domesticated silkworm. In: King, R.C., Ed., *Handbook of Genetics*. Vol. 3. *Invertebrates of Genetic Interest* Plenum. New York, pp. 63–1124.
- Tillman, J.A., Seybold, S.J., Jurenka, R.A., and Blomquist, G.J., (1999). Insect pheromones—an overview of biosynthesis and endocrine regulation. *Insect Biochem. Mol. Biol.* **29**: 481–514.
- Ting, C.T., Takahashi, A., and Wu, C.I., (2001). Incipient speciation by sexual isolation in *Drosophila*: concurrent evolution at multiple loci. *Proc. Natl. Acad. Sci. USA* **98**: 6709–6713.
- Trpis, M., and Hausermann, W., (1978). Genetics of house-entering behaviour in East African populations of *Aedes aegypti* (L.) (Diptera: Culicidae) and its relevance to speciation. *Bull. Entomol. Res.* **68**: 521–532.
- Tully, T., (1991a). Genetic dissection of learning and memory in *Drosophila melanogaster*. In: Madden IV, J., Ed., *Neurobiology of Learning, Emotion and Affect*, Raven Press, New York, pp. 29–66.
- Tully, T., (1991b). Physiology of mutations affecting learning and memory in *Drosophila*--the missing link between gene product and behavior. *Trends Neurosci.* **14**: 163–164.
- Tully, T., (1996). Discovery of genes involved with learning and memory: an experimental synthesis of Hirschian and Benzerian perspectives. *Proc. Natl. Acad. Sci. USA* **93**: 13460–13467.
- Vanaphan, N., Dauwalder, B., and Zufall, R.A., (2012). Diversification of takeout, a male-biased gene family in *Drosophila*. *Gene* **491**: 142–148.
- Vanin, S., Bhutani, S., Montelli, S., Menegazzi, P., Green, E.W., and Pegoraro, M., (2012). Unexpected features of *Drosophila* circadian behavioural rhythms under natural conditions. *Nature* **484**: 371–375.
- Via, S., (1990). Ecological genetics and host adaptation in herbivorous insects: the experimental study of evolution in natural and agricultural systems. *Annu. Rev. Entomol.* **35**: 421–426.
- Villella, A., Gailey, D.A., Berwald, B., Ohshima, S., Barnes, P.T., and Hall, J.C., (1997). Extended reproductive roles of the *fruitless* gene in *Drosophila melanogaster* revealed by behavioral analysis of new *fru* mutants. *Genetics* **147**: 1107–1130.

- Vosshall, L.B., (2000). Olfaction in *Drosophila*. *Curr. Opin. Neurobiol.* **10**: 498–503.
- Vosshall, L.B., (2001). The molecular logic of olfaction in *Drosophila*. *Chem. Senses* **26**: 207–213.
- Vosshall, L.B., (2007). Into the mind of a fly. *Nature* **450**: 193–197.
- Vosshall, L.B.H., Amrein, P.S., Morozov, A., Rzhetsky, and Axel, R., (1999). A spatial map of olfactory receptor expression in the *Drosophila* antenna. *Cell* **96**: 725–736.
- Vosshall, L.B., Wong, A.M., and Axel, R., (2000). An olfactory sensory map in the fly brain. *Cell* **102**: 147–160.
- Waddell, S., and Quinn, W.G., (2001). Flies, genes and learning. *Annu. Rev. Neurosci.* **24**: 1283–1309.
- Wager-Smith, K., and Kay, S.A., (2000). Circadian rhythm genetics: from flies to mice to humans. *Nat. Genet.* **26**: 23–27.
- Walsh, B., (2009). Quantitative genetics, version 3.0: where have we gone since 1987 and where are we headed? *Genetica* **136**: 213–223.
- Warmke, J., Drysdale, R., and Ganetzky, B., (1991). A distinct potassium channel polypeptide encoded by the *Drosophila* eag locus. *Science* **252**: 1560–1562.
- Wheeler, D.A., Kyriacou, C.P., Greenacre, M.L., Yu, Q., Rutila, J.E., Rosbash, M., and Hall, J.C., (1991). Molecular transfer of a species-specific behavior from *Drosophila simulans* to *Drosophila melanogaster*. *Science* **251**: 1082–1085.
- Whitfield, C.W., Ben-Shahar, Y., Brillet, C., Leoncini, I., Crauser, D., Leconte, Y., Rodriguez-Zas, S., and Robinson, G.E., (2006). Genomic dissection of behavioral maturation in the honey bee. *Proc. Natl. Acad. Sci. USA* **103**: 16068–16075.
- Williams, J.A., and Sehgal, A., (2001). Molecular components of the circadian system in *Drosophila*. *Annu. Rev. Physiol.* **63**: 729–755.
- Wolf, M.E., (1999). Cocaine addiction: clues from *Drosophila* on drugs. *Curr. Biol.* **9**: R770–R772.
- Wolf, M., van Doorn, G.S., Leimar, O., and Weissing, F.J., (2007). Life-history trade-offs favour the evolution of animal personalities. *Nature* **447**: 581–584.
- Wolf, M., van Doorn, G.S., and Weissing, F.J., (2008). Evolutionary emergence of responsive and unresponsive personalities. *Proc. Natl. Acad. Sci. USA* **105**: 15825–15830.
- Wyatt, T.D., (1997). Methods in studying insect behaviour. In: Dent, D.R., and Walton, M.P., Eds., *Methods in Ecological and Agricultural Entomology*, CAB International, Wallingford, UK, pp. 27–56.
- Yamamoto, D., Jallon, J.M., and Komatsu, A., (1997). Genetic dissection of sexual behavior in *Drosophila melanogaster*. *Annu. Rev. Entomol.* **42**: 551–585.
- Yasukochi, Y., Miura, N., Nakano, R., Sahara, K., and Ishikawa, Y., (2011). Sex-linked pheromone receptor genes of the European corn borer, *Ostrinia nubilalis*, are in tandem arrays. *PLoS One* **6** (4): e18843.
- Young, M.W., (1998). The molecular control of circadian behavioral rhythms and their entrainment in *Drosophila*. *Annu. Rev. Biochem.* **67**: 124–152.
- Zars, T., (2000). Behavioral functions of the insect mushroom bodies. *Curr. Opin. Neurobiol.* **10**: 790–795.
- Zars, T., Fischer, M., Schulz, R., and Heisenberg, M., (2000). Localization of a short-term memory in *Drosophila*. *Science* **288**: 672–676.
- Zerr, D.M., Hall, J.C., Rosbash, M., and Siwicki, K.K., (1990). Circadian fluctuations of period protein immunoreactivity in the CNS and the visual system of *Drosophila*. *J. Neurosci.* **10**: 2749–2762.
- Zhong, Y., and Wu, C.F., (1991). Altered synaptic plasticity in *Drosophila* memory mutants with a defective cyclic AMP cascade. *Science* **251**: 198–201.

# Molecular Systematics and the Evolution of Arthropods

## Chapter Outline

**12.1 Overview** 522

**12.2 Introduction** 523

**12.3 Controversies in Molecular Systematics and Evolution** 527

12.3.1 Molecular versus Morphological Traits 527

12.3.2 The Molecular Clock 528

12.3.3 The Neutral (or Nearly Neutral) Theory of Evolution 530

12.3.4 Homology and Similarity 531

**12.4 Molecular Methods for Molecular Systematics and Evolution** 531

12.4.1 Protein Electrophoresis 533

12.4.2 Molecular Cytology 535

12.4.3 Restriction Fragment Length Polymorphism (RFLP) Analysis 536

12.4.4 DNA and Genome Sequencing 537

12.4.5 Fragment Analyses of Genomic DNA 538

**12.5 Targets of DNA Analysis** 538

12.5.1 Mitochondria 539

12.5.1.1 *The Barcode of Life Project* 542

12.5.2 Ribosomal RNA 544

12.5.3 Satellite DNA 546

12.5.4 Introns 546

12.5.5 Nuclear Protein-Coding Genes 546

12.5.6 Rare Genomic Changes 546

12.5.7 MicroRNAs 547

**12.6 Steps in Phylogenetic Analysis of DNA Sequence Data** 549

12.6.1 Gene Trees or Species Trees 549

12.6.2 Rooted or Unrooted Trees 549

12.6.3 Tree Types 549

12.6.4 Project Goals and Appropriate DNA Sequences 551

12.6.5 Sequence Comparisons with BLAST 553

12.6.6 Aligning Sequences 556

12.6.7 Constructing Phylogenies 556

12.6.8 Artifacts 561

12.6.9 Software Packages 562

<b>12.7 The Universal Tree of Life</b>	<b>564</b>
12.7.1 Two Domains	564
12.7.2 Three Domains	564
12.7.3 Origin of Eukaryota	565
<b>12.8 The Fossil Record of Arthropods</b>	<b>568</b>
<b>12.9 Molecular Analyses of Arthropod Phylogeny</b>	<b>571</b>
12.9.1 Evolution of the Ecdysozoa	571
12.9.2 Relationships among the Arthropoda	571
12.9.3 The Phylogeny of the Holometabola	571
12.9.4 Congruence Between Morphology- and Molecular-Based Trees	572
12.9.5 Genomes and Arthropod Phylogenies	572
<b>12.10 Molecular Evolution and Speciation</b>	<b>573</b>
12.10.1 Species Concepts	573
12.10.2 How Many Genes are Involved in Speciation?	575
12.10.3 Detecting Cryptic Species	577
<b>12.11 Some Conclusions</b>	<b>578</b>
<b>Relevant Journals</b>	<b>579</b>
<b>References Cited</b>	<b>579</b>

## 12.1 Overview

Systematics is the study of phylogeny and taxonomy. Taxonomy can be divided into descriptive taxonomy and identification. DNA and genome sequences are suitable for systematics studies because they provide the most direct analysis of the genetic material possible and are unlikely to be confounded by life-stage or environmentally induced variability in morphology. Molecular techniques used include restriction analyses of DNA sequences and DNA sequencing of genes and entire genomes. Each method has virtues and limitations in the amount and type of information provided, their technical difficulties, and costs.

There have been several significant controversies associated with using molecular techniques to study systematics and evolution. These include debates over the relative importance of molecular versus morphological data, the constancy of the molecular clock for evaluating time of divergence of taxa, the proper use of the terms homology and similarity, and the neutrality of DNA sequence variation. Another issue is how to resolve incongruencies between molecular- and morphology-based phylogenies and which phylogenetic method, which evolutionary model, and which DNA sequences should be used. The immense diversity of insects and their long evolutionary history provide challenges, especially with the analysis of deep evolutionary patterns. The availability of entire genomes of diverse arthropods will increasingly resolve many questions. With the use of molecular methods, evolutionary biologists and population geneticists are beginning to use common approaches to study both intraspecific and

interspecific genetic diversity. The methodology used to analyze data and to develop phylogenetic trees continues to improve. However, considerable controversy may surround specific conclusions because the use of different analysis tools or different DNA sequences can result in different phylogenetic trees.

In 2003, DNA barcoding was proposed as a method to obtain rapid and accurate identification of a broad range of organisms, including insects. The Consortium for the Barcode of Life was initiated in 2004, and in 2010 the International Barcode of Life project was established to develop an automated identification system based on a DNA barcode library, based on the COI mitochondrial sequences, of most eukaryotes. Projects for the Lepidoptera, Trichoptera, Formicidae, and Apidae were developed. The automated system requires advances in information technology for the huge datasets produced and one (Barcode of Life Database [BOLD]) is freely available to all scientists. However, even Barcoding has limitations and its validity has been questioned due to the discovery of nuclear copies of the mitochondrial genes (called numts), which can result in erroneous conclusions about the diversity of arthropod species.

## 12.2 Introduction

The methods and concepts used to classify arthropods, and other organisms, are themselves undergoing evolution. Classification started when Linnaeus established a formalized hierarchical system of binomial nomenclature in 1758. Carolus Linnaeus, a Swedish botanist, published *Systema Naturae* in 1758 and proposed basic principles for organizing newly described species into groups and for assigning these groups to specific taxonomic categories. This resulted in a ranking classification, typically ascending from species to genus, family, order, class, super class, subphylum, phylum, and kingdom. See [Table 12.1](#) for two outlines of the higher classification of the Phylum Arthropoda (others are available). One outline was presented in the second edition of *Insect Biology and Diversity* ([Daly et al. 1998](#)), and the third edition was just published by [Whitfield and Purcell \(2013\)](#), indicating how rapidly concepts can change with the addition of new, mostly molecular, data. The classification of the orders also changed between edition 2 and 3 ([Table 12.2](#)), with termites and cockroaches now included in the Blattodea and with Homoptera and Hemiptera now collapsed into Hemiptera.

The Linnaean system made no provision for naming and classifying organisms based on evolutionary relationships and Linnaeus assumed the living world was limited to  $\approx 10,000$  species. The notion that a classification should be based on phylogenetic, or evolutionary, relationships developed only after Darwin's publication of *The Origin of the Species* in 1859. The identification, description, and explanation of the diversity of organisms are known as systematics.

**Table 12.1: Two Recent Higher Classifications of the Phylum Arthropoda Illustrate Changing Views of the Evolutionary Relationships of the Arthropoda.**

Daly et al. 1998	Whitfield and Purcell 2013
Phylum Arthropoda Subphylum Trilobita Subphylum Chelicerata Class Merostomata Class Pycnogonida Class Arachnida Subphylum Crustacea Subphylum Labiata Superclass Myriapoda Class Diplopoda Class Chilopoda Class Pauropoda Class Symphyla Superclass Hexapoda Class Parainsecta (Protura and Collembola) Class Entognatha Class Insecta	Phylum Arthropoda Subphylum Trilobita Subphylum Chelicerata Class Merostomata Class Pycnogonida Class Arachnida Subphylum Crustacea Subphylum Myriapoda  Class Diplopoda Class Chilopoda Class Pauropoda Class Symphyla Subphylum Hexapoda Informal group Entognatha Order Protura Order Collembola Order Diplura Class Insecta

There is a broad overlap in the use of the terms **systematics** and **taxonomy**. [Mayr and Ashlock \(1991\)](#) define systematics as “the scientific study of the kinds and diversity of organisms and of any and all relationships among them” or the “science of the diversity of organisms” and **taxonomy** as “the theory and practice of classifying organisms.” Taxonomy can be divided into descriptive taxonomy and identification ([Post et al. 1992](#)). Systematics deals with populations, species, and higher taxa. It is concerned also with variation within taxa. Thus, DNA analysis is particularly suitable for systematics studies because it is the most direct analysis of the genetic material possible and is unlikely to show life stage or environmentally induced variability.

During the 19th century, after Darwin’s theory of evolution was proposed and numerous new organisms were found, more-extensive nomenclature rules were developed to accommodate the growing numbers of plant, animal and microbial species. So far, an estimated 1.4 million species have been identified, but these species may represent only 10% or so of the total species thought to live on this planet. [Mora et al. \(2011\)](#) recently calculated that ≈8.7 million eukaryotic species exist globally. (The number of unknown microbial species may exceed this number, and many may be important in the biology of arthropods.) Big surprises still occur, even with the relatively well-known insects. A new order

**Table 12.2: Classification of the Orders of the Subphylum Hexapoda (After Whitfield and Purcell 2013).**

Informal group	
Entognatha	Order Protura Order Collembola Order Diplura
<b>Class Insecta</b>	
Apterygota	Order Archeognatha Order Thysanura
Paleoptera	Order Odonata Order Ephemeroptera
Polyneoptera	Order Grylloblattodea Order Mantophasmatodea Order Embiidina Order Phasmida Order Zoraptera Order Dermaptera Order Orthoptera Order Blattodea Order Mantodea Order Plecoptera
Paraneoptera	Order Hemiptera Order Thysanoptera Order Psocidea
Holometabola	Order Strepsiptera Order Hymenoptera Order Megaloptera Order Neuroptera Order Raphidioptera Order Coleoptera Order Mecoptera Order Siphonaptera Order Diptera Order Lepidoptera Order Trichoptera

of insects (Mantophasmatodea) was discovered in 2002 in the mountains of Namibia in Africa (Klass et al. 2002), which was surprising because entomologists thought all insect orders had been detected. However, some now consider this group to be a suborder in a new insect order that contains grylloblattids, mantids, and phasmids, as well as mantophasmids (Cameron et al. 2006).

Under the Linnaean system, a taxonomist begins by assessing the physical (phenotypic) characteristics that a set of species shares, then selects the most representative species to be the “type” for each genus, then the most representative genus to be the type of the family, and so on. Individual specimens are deposited in museums to serve as a reference for that species and genus. When new specimens are found with similar traits, they are categorized as part of a known species, as a new species, or as a new genus (or family or order) depending on how closely the new specimens resemble the “type.” This reliance on types results in dramatic changes if a systematist reevaluates a group and decides that some members don’t belong. Removal of these individuals can mean that the group’s name must be changed, which often is disruptive to other biologists.

In the 1980s, an assessment of methods occurred after a classification method called **cladistics**, which is based on the evolutionary histories of organisms, was proposed. The cladistics approach is based on phylogeny, whereas traditional Linnaean methods (**phenetics**) are not. Most current systematists now take the phylogenetic approach with classification based on evolutionary relationships. However, the use of DNA-based methods created new concerns about appropriate methods of analysis and whether molecular and traditional morphological methods provide equivalent answers. The enormous amount of DNA sequence data also requires that new methods of handling large datasets be developed (Hall 2011, Fan and Kubatko 2011).

Systematists who have concluded that the fundamental Linnaean binomial system of nomenclature is obsolete have added additional ferment (De Queiroz and Gauthier 1994, Ereshefsky 2001). The **PhyloCode** was proposed as a solution to the perceived problems (De Queiroz and Gauthier 1994, Pennisi 2001, [www.ohio.edu/phylocode](http://www.ohio.edu/phylocode)). Advocates of PhyloCode want to replace the Linnaean system to make species names more stable. Under this system, genus names might be lost and species names might be shortened, hyphenated with their former genus name, or given a numeric identification. The debate over which is the better system has generated much heat and only time will tell which approach has the fewest shortcomings (Pennisi 2001, Nixon et al. 2003).

Systematics encompasses the study of both **phylogeny** and microevolutionary change. **Molecular evolution** encompasses 1) analyzing the evolution of DNA and proteins and the mechanisms responsible for such changes and 2) deciphering the evolutionary history of genes and organisms. A more recent topic, available only since the complete genomes of a variety of organisms have been sequenced, is **comparative genomics**. **Comparative genomics** compares the

overall structure and function of genomes. Molecular evolution and phylogeny are interrelated because phylogenetic knowledge is essential for determining the order of changes in the molecular characters being studied, whereas knowledge of the pattern and rate of change of a molecule is crucial in efforts to reconstruct the evolutionary history of a group of organisms (Li and Graur 1991, Graur and Li 2000). This chapter introduces several common molecular methods and describes their applications, limitations, and relative costs for systematic and evolutionary studies. First, key controversies are described.

## 12.3 Controversies in Molecular Systematics and Evolution

Significant controversies have been associated with using molecular tools, including: debates over the relative importance of molecular versus morphological data, the constancy of evolutionary rates (the molecular clock), the use of the terms homology and similarity, and the neutrality of DNA sequence variation. Other controversies surround the choice of evolutionary models for analysis of the data and the choice of which molecular data to use.

### 12.3.1 Molecular versus Morphological Traits

There was debate over whether morphological or molecular characters are *better* for constructing phylogenies (Adoutte et al. 2000). When comparisons were made, it appeared that morphological changes and molecular changes may be independent, responding to different evolutionary pressures or differently to evolutionary pressures. Hillis and Moritz (1990) noted that the real issue in choosing a technique to answer a hypothesis should be whether the 1) specific characters chosen exhibit the variation that is appropriate to the question posed, 2) characters have a genetic basis, and 3) data are collected and analyzed in such a way that it is possible to use both morphological and molecular information. Molecular and morphological data each have advantages and disadvantages.

DNA-sequence data have the advantage of having a clear genetic basis and the amount of data is limited only by the genome size and the time and funds of the scientist! Morphological data have the advantage that they can be obtained from fossils (if available) and preserved collections and can be interpreted in the context of ontogeny. Only limited amounts of DNA data can be obtained from preserved fossils by the polymerase chain reaction (PCR) due to degradation of DNA over time; see Chapter 8 for a discussion of PCR analyses of ancient DNA. Moritz and Hillis (1990) concluded the debate should not be either/or; studies that incorporate both molecular and morphological data provide better results than those using just one approach, and the recent trend is

to use as many different genes as possible rather than just a few hundred bp of a single protein-coding gene. Furthermore, Quental and Marshall (2010) argue that it is important to include the fossil record (if possible) in phylogenetic analyses to document change through time because molecular assumptions may be unrealistic (uniform changes in time may not occur). Assumptions made for fossils are that correct taxonomy and stratigraphy have been applied.

### 12.3.2 The Molecular Clock

Until the 1960s, the analysis of fossils was the only way to estimate the *time* when ancestors of extant organisms lived. Molecular studies in the 1960s provided a concept called the **molecular clock** that could be used to estimate the evolutionary history and time of divergence of organisms. The molecular clock was particularly useful for living species that have a poor fossil record, which includes a very high proportion of extant species.

The molecular-clock hypothesis was proposed after Zuckerkandl and Pauling (1965) examined amino-acid substitutions in hemoglobin and cytochrome c proteins from different vertebrates. They found the rate of molecular evolution was approximately constant over time in all vertebrate lineages examined and concluded that amino-acid sequences could be used to measure the evolutionary distance (time) between organisms by counting the number of accumulated changes (mutations).

The molecular clock is based on the assumption that basic processes such as DNA replication, transcription, protein synthesis, and metabolism are remarkably similar in all living organisms and the proteins and RNAs that carry out key “housekeeping functions” are highly conserved. Of course, over time, mutations in housekeeping genes occurred and DNA and protein sequences changed, although the changes tended to preserve the *function* of the gene rather than modify or improve it. Thus, changes in these fundamental genes should have minimal, or no, effect on function. For example, because the genetic code is degenerate, the third base in a codon often can be altered without affecting which amino acid is designated. Changes in the code also can occur without changing protein function if amino-acid changes occur in region(s) that do not affect function, or if one amino acid is replaced by a similar amino acid. The molecular-clock hypothesis assumes that mutations in the housekeeping genes that constitute the clock occur at a *constant* rate, thus providing a reliable method for measuring time.

Unfortunately, subsequent analyses of different protein sequences suggest that the rates of change *can* vary between proteins and lineages, indicating that

the molecular clock may tick at different rates in different lineages (Rodriguez-Trelles et al. 2001, Lanfear et al. 2010). For example, cytochrome c has an acceptable clockwise behavior for the original organisms studied. However, copper-zinc superoxide dismutase (SOD) behaves like an erratic clock (Ayala 1986). The average rate of amino-acid substitutions in SOD per 100 residues per 100 million years is a minimum of 5.5 when fungi and animals are compared. The rate of substitutions in SOD is 9.1 amino acids/100 residues/100 million years when comparisons are made between insects and mammals, and 27.8 when mammals are compared with each other (Ayala 1986, Fitch and Ayala 1994). Thus, the molecular clock should be calibrated with data that are independently derived, and preferably with fossil evidence, if the absolute time of divergence is desired. Wilson et al. (1987) pointed out that analyses by both morphological and molecular techniques of species with abundant fossil records have reduced the uncertainty in estimating the time of divergence by several orders of magnitude. However, the molecular clock is thought by some to be more useful in calculating *relative* times rather than absolute times of divergence, especially if fossil data are unavailable to validate the dates. Arbogast et al. (2002) reviewed the methods of estimating divergence times from molecular data and emphasized the importance of model testing and appropriate statistical tests. Warnock et al. (2012) explored the effect of using different methods for estimating divergence time using Bayesian methods and showed that calibration is important.

The molecular-clock approach was used by Moran et al. (1993) to determine when endosymbiont bacteria (*Buchnera*) colonized their aphid hosts. Moran et al. (1993) compared 16S ribosomal DNA sequences of aphids and *Buchnera* and found the clock was approximately constant. These symbiotic bacteria live within specialized aphid cells, are maternally inherited, and are essential for growth and reproduction of their hosts, indicating a long and intimate relationship. The 16S ribosomal DNA (rDNA) sequences indicate that the symbionts in diverse aphids are distinct and concordant with the phylogeny of their hosts, suggesting that the current distribution of *Buchnera* is due to vertical transfer from an ancestral aphid. The data also indicate that cospeciation occurred, with the aphids and their endosymbionts radiating synchronously. Moran et al. (1993) estimated the aphid and bacterial radiations occurred at a relatively constant rate, with 0.01–0.02 substitutions per site per 50 million years, suggesting that the association between aphids and endosymbionts began ≈160–280 million years ago (mya).

Nardi et al. (2010) evaluated the timing of the domestication of olive fly and a host shift to cultivated olives by using complete mitochondrial genomes.

### 12.3.3 The Neutral (or Nearly Neutral) Theory of Evolution

Another controversy involves the mechanism(s) of molecular evolution. At the core of the dispute is the **neutral theory of molecular evolution** (Kimura 1968, 1983, 1987, Ohta 1996, 2000b). The neutrality theory (or the modified “nearly neutral” theory) recognizes that for any gene a large proportion of all possible mutations (alleles) are deleterious and that these are eliminated or maintained at a very low frequency by natural selection. The evolution of morphological, behavioral, and ecological traits is governed largely by natural selection, because selection occurs on favorable alleles and against deleterious alleles. However, many mutations can result in alleles which are equivalent, or nearly so, to each other. Neutral mutations are *not* subject to selection because they do not affect the fitness of the individual carrying them. Neither do they affect their morphology, physiology, or behavior. The neutrality theory states that the majority of nucleotide substitutions in the course of evolution are the result of the gradual, random fixation of neutral changes, rather than the result of positive Darwinian selection. Neutral mutations can spread in a population because only a relatively small number of gametes are sampled each generation (random genetic drift). By chance, they can be transmitted to the next generation at a higher frequency (Kimura 1968, 1983).

Ohta (1996) concluded that the “strictly neutral theory has not held up as well as the nearly neutral theory, yet remains invaluable as a null hypothesis for detecting selection.” The main difference between the nearly neutral theory and the traditional selection theory is that “the nearly neutral theory predicts rapid evolution in small populations, whereas the latter predicts rapid evolution in large populations” (Ohta 1996). Kreitman (1996) noted that the neutral theory has been useful for organizing thinking about the nature of evolutionary forces acting on variation at the DNA level and has provided a set of testable predictions (acting as a useful null hypothesis). However, Kreitman (1996) argues “the neutral theory cannot explain key features of protein evolution nor patterns of biased codon usage in certain species.” Despite this, he concludes the neutral theory “is likely to remain an integral part of the quest to understand molecular evolution.” Finally, both Ohta (1996) and Kreitman (1996) agree that the “nearly neutral theory” is more compatible with the current data in explaining synonymous changes and the evolution of codon bias.

Why be concerned about neutrality or nearly neutral theories? The neutrality theory is a basic assumption of some methods of estimating phylogeny and also affects the molecular-clock hypothesis (Ohta 1996, 2000b, Kreitman 1996). Data indicate that many protein, chromosome, and DNA variations are

under selection. Data also support the hypothesis that much molecular variation is essentially (nearly) neutral. The debate thus is over how *many*, and *which*, molecular variants are selectively neutral or nearly neutral. Moritz and Hillis (1990) suggest that each molecular marker should be tested for neutrality. They also note that, because most departures from neutrality are locus-specific, selection will have relatively minor effects on analyses if many different loci are studied.

#### 12.3.4 Homology and Similarity

A fourth issue concerns terminology. **Homology** is an important concept in biology and historically has had the precise meaning of “having a common evolutionary origin” (Reeck et al. 1987). However, homology has been used in a more loose sense when comparing protein and nucleic-acid sequences. Protein and nucleic-acid sequences from different organisms have been called homologous when they are *similar*. According to the traditional definition of homology, amino-acid or nucleotide sequences are either homologous, or not. They cannot exhibit a “level of homology” or “percent homology.” Reeck et al. (1987) point out that using homology to mean similarity can cause three different problems. First, sequence similarities may be called homologies, but the sequences are *not evolutionarily related*, which is inconsistent. Second, similarities (again called homologies) are discussed but evolutionary origins are not, which can lead the reader to believe that coancestry is involved when it is not. Third, the similarities (called homologies) are used to support a hypothesis of evolutionary homology. The problem is that although similarity is easy to document, a common evolutionary origin usually is more difficult to establish, especially if fossil evidence is lacking. Several evolutionary processes other than homology could account for sequence similarities, including **convergent evolution**, that is the independent evolution of the same characteristic in separate branches of a phylogenetic tree. When in doubt, it is better to talk about “percentage similarity” of DNA sequences.

### 12.4 Molecular Methods for Molecular Systematics and Evolution

Systematics studies conducted prior to the 1960s primarily utilized morphological and behavioral attributes as characters, although cytogenetic characters were used in some cases (Mayr 1970, White 1973, 1978). In the 1960s, electrophoresis of proteins began to provide new characters after Lewontin and Hubby (1966) demonstrated that protein-coding genes often are polymorphic (have more than one allele) and that gel electrophoresis of proteins could reveal the

presence of functionally similar forms of enzymes (**isozymes**). Protein electrophoresis provides a relatively inexpensive method for analyzing several genes from individuals at the same time (Hames and Rickwood 1981, Pasteur et al. 1988, Murphy et al. 1996). This technique is useful for analyzing mating systems, heterozygosity, relatedness, geographic variation, hybridization, species boundaries, and phylogenetic analyses of divergences within the last 50 million years and is inexpensive and rapid.

Immunological analyses were used in the 1960s, but they are rarely used today. Immunological techniques provided qualitative or quantitative estimates of amino-acid sequence differences between homologous proteins (Maxon and Maxon 1990). Cytogenetic analyses of variation in chromosome structure and number have been used for studies of hybridization and species boundaries (White 1973). Specific DNA sequences can be localized by *in situ* hybridization and new staining techniques can reveal the fine structure of chromosomes by revealing banding patterns. However, cytogenetic analyses are less useful for phylogenetic analyses, gene evolution, heterozygosity, and relatedness studies (Hillis et al. 1996).

Differences in single-copy DNA sequences revealed by **DNA–DNA hybridization** have been used for analysis of phylogenies since the late 1960s (Powell and Caccone 1990, Werman et al. 1996). DNA–DNA hybridization is used for analyses of species and higher taxa relationships up to the family and order level but is used relatively infrequently for arthropods, so it is not discussed further.

Protocols of molecular methods for systematics and evolutionary studies were provided by Hillis and Moritz (1990), including guidelines for sampling, collection and storage of tissues, protocols for isozyme electrophoresis, immunological techniques, molecular cytogenetics, DNA–DNA hybridization, restriction site analysis, Sanger sequencing of nucleic acids, and analytical methods for intra-specific differentiation and phylogeny reconstruction. Pasteur et al. (1988) and Murphy et al. (1996) described protocols and methods of isozyme analysis. Weir (1990) provided guidelines on analyzing population structure, phylogeny construction, and diversity by using molecular and morphological data. Protocols are readily available for the PCR and nucleic-acid sequencing and data analysis (Howe and Ward 1989, Doolittle 1990, Hillis et al. 1990, 1996, Gribskov and Devereaux 1991, Palumbi 1996, Green 2001, Gibson and Muse 2002, Hall 2011). Updated methods for molecular phylogenetics analyses are available: Hall (2011) provides the fourth edition of *Phylogenetic Trees Made Easy: A How To Manual*, and Wiley and Lieberman (2011) provide the second edition of *Phylogenetics: Theory and Practice of Phylogenetic Systematics*.

### 12.4.1 Protein Electrophoresis

The term **isozyme** is a general designation for multiple forms of a single enzyme. Isozymes will catalyze the same reaction, but they may differ in properties such as the pH or substrate concentration at which they best function. Isozymes are complex proteins made up of paired polypeptide subunits; their subunits may be coded for by different loci. For example, protein Z could be a tetramer made up of two polypeptides, A and B. Five isozymes of protein Z could exist and be symbolized: AAAA, AAAB, AABB, ABBB, and BBBB. Isozymes may have different isoelectric points and be separated by gel electrophoresis.

The term **allozyme** refers to variant proteins produced by *allelic* forms of the *same locus*. Thus, A is now A'. A different mutation of A could produce A''. Allozymes are a *subset* of isozymes; allozymes may differ by net charge or size so they can be separated by electrophoresis.

The process of analyzing isozymes or allozymes can be divided: extraction of proteins, separation, staining, interpretation, and application. Proteins are more difficult to handle than DNA because they are more susceptible to degradation. Proteins must be frozen and stored at  $-70^{\circ}\text{C}$  but, even at those temperatures, some proteins can degrade within months.

Proteins are separated in an electric field on a gel. In gels with a single pH the proteins move through the gel at a continuous rate, but in gels with a pH gradient, they move until they reach their isoelectric point and then stop. The resultant electrophoretic bands are visualized by appropriate staining (May 1992, Murphy et al. 1996). If a general protein-detection system is used, only those proteins present in large quantities are detected, but more-specific stains can be used. Specific stains and buffer recipes are available for >50 enzymes (May 1992). The banding phenotypes observed on the gels can be interpreted in terms of genes and their alleles (Pasteur et al. 1988, May 1992).

Protein-coding genes are often codominant, with both alleles being expressed in heterozygous organisms. This makes it possible to relate a particular phenotype to a given genotype, if we assume that isozyme data reflect changes in the encoding DNA sequence. To interpret the banding patterns, the number of sub-units in the enzyme and the distribution of enzymes in particular cells or tissues should be known (May 1992).

Analyses of isozymes remain a cost-efficient and useful method for deciphering the systematics, population genetics, and evolution of insects (Table 12.3). Protein electrophoresis can be conducted using starch (horizontal or vertical

**Table 12.3: Applications of Selected Molecular Methods to Systematics Problems.**

Research problem <sup>a</sup>	Isozymes	RFLPs	RAPD-PCR	Microsatellite single-locus	Multilocus fingerprints	DNA/RNA sequencing	Genome sequencing
Gene evolution	M	M	I	M	I	A	I
Relatedness	M	M	M	A	M	\$	I
Hybridization	A	A	A	M	I	\$	I
Species boundaries	A	A	A	M	I	A	\$
<b>Phylogeny</b>							
0–5 mya	A	A	I	M	I	A	\$
50–500 mya	M	M	I	I	I	A	A
500–3500 mya	I	I	I	I	I	A	A

Modified from [Hillis et al. \(1996\)](#) and [Trautwein et al. \(2012\)](#).

<sup>a</sup>I, inappropriate use of the technique; M, marginally appropriate or appropriate under limited circumstances; \$, appropriate but probably not cost-effective; A, appropriate and effective method.

gel systems), polyacrylamide, agarose, and cellulose acetate gels as substrates ([Hames and Rickwood 1981](#)). Each has specific advantages and disadvantages ([Moritz and Hillis 1990](#)). However, isozyme or allozyme data are useful for estimating the evolution of only a portion of the genome: those genes coding for enzymes that have a different charge and size. The data also are most useful for analyzing relatively closely related taxa. Unfortunately, allozyme variation in some insects, such as aphids and Hymenoptera, is low and other molecular techniques are required.

An example of allozyme analysis illustrates an economically important application of the technique. Twenty-four populations consisting of three subspecies of *Culicoides variipennis* (Diptera: Ceratopogonidae) from different geographic regions were examined for genetic differences ([Tabachnick 1992](#)). Twenty-one loci were examined among the 24 populations of this vector of bluetongue virus, a disease that causes losses of US\$125 million annually to the U.S. livestock industry. The results were analyzed with a stepwise discriminant analysis and are consistent with the conclusion that there are three North American subspecies; the three subspecies appear to be sufficiently different to be considered species. Furthermore, geographic variation in bluetongue disease epidemiology is correlated with the distributions of the three subspecies. These, and subsequent, data support the hypothesis that one subspecies is more effective as a vector of the virus ([Tabachnick 1996](#), [Holbrook et al. 2000](#)). The results had implications for pest-management programs and had significant economic impacts. The areas inhabited by the two (nonvector) subspecies could be considered virus-free regions and animals raised in such areas would not have to undergo extensive

testing when livestock or germplasm from them are exported to regions without the disease.

#### 12.4.2 Molecular Cytology

Three breakthroughs in cytogenetic techniques revived this approach to systematic and evolutionary studies. The first was the discovery that hypotonic treatment spreads metaphase chromosomes, allowing more accurate counts of chromosome numbers and details of chromosome morphology. The second was the development of chromosome-banding techniques that allow the identification of specific types of DNA within homologous chromosomes. The third was the development of *in situ* hybridization techniques that allow specific DNA sequences to be localized to particular segments of the chromosomes.

*In situ* hybridization involves annealing single-stranded probe molecules to target DNA to form DNA duplexes. *In situ* hybridization is effective in locating satellite DNA, ribosomal gene clusters, or duplicated genes of polytene chromosomes and can even locate single-copy DNA on mitotic chromosomes. Chromosomal DNA is denatured in such a way that it will anneal with high efficiency to complementary ss nucleic-acid probes to form hybrid duplexes. Because chromosomal DNA is complexed with proteins and RNA, the efficiency of *in situ* hybridization is determined by how well the chromosomal DNA can be denatured, how much DNA is lost during fixation and treatment, and whether chromosomal proteins are present in the region of interest (Sessions 1996). Autoradiography is used to detect where hybridization between a radioactive probe and its target DNA occur. Nonradiographic labeling techniques such as biotinylation or fluorescence can be used, as well.

Chromosome morphology may be used as a taxonomic character. In many cases, chromosomes can be identified by their relative size, centromere position, and secondary constrictions. Many chromosomes, particularly polytene chromosomes, have complex patterns of bands or other markers that can be used to identify specific populations or to discriminate between closely related species. Q, G, or C banding produces distinctive patterns that identify chromosomes in most species.

**Q banding** is the simplest technique, and involves treating chromosome preparations with quinacrine mustard or quinacrine dihydrochloride, which produces fluorescent bands that are brightest in AT-rich regions of the chromosomes. Q banding is visible only with UV optics and the bands fade rapidly. **G banding** involves treating chromosome preparations with trypsin or NaOH and staining with Giemsa in a phosphate buffer, a process that yields alternating light and

dark bands. The dark bands are primarily AT-rich regions and thus correspond to most Q-bands. **C banding** requires a stringent extraction step that can result in loss of chromosomal DNA. During C-banding, chromosomes are treated with a strong base at a high temperature, incubated in a sodium citrate solution again at high temperature, and stained in a concentrated Giemsa solution. C banding extracts almost all of the non-C band chromatin, leaving only constitutive heterochromatin, which usually contains rapidly reassociating repeated DNA sequences (Sessions 1996).

Cytogenetic data provide information independent from morphological, biochemical, or behavioral data for phylogenetic analyses. Cytogenetic data can reveal differences or similarities that may not be obvious at the morphological level. Chromosome size, shape, number, and ploidy levels can provide insights into the genetic architecture of taxa. Banding studies reveal aspects of the structural organization of chromatin on individual chromosomes, whereas probes of DNA sequences with *in situ* hybridization can reveal finer details of chromosome anatomy in terms of spatial arrangement, and presence or absence of particular kinds of DNA sequences.

#### **12.4.3 Restriction Fragment Length Polymorphism (RFLP) Analysis**

Restriction-enzyme analyses are versatile, providing information on the nature, as well as the extent, of differences between sequences in nuclear or mitochondrial DNA (mtDNA) (Dowling et al. 1996, Table 12.3). RFLP analyses reveal variations within a species in the length of DNA fragments generated by a specific restriction endonuclease. RFLP variations are caused by mutations that create or eliminate recognition sites for the restriction enzymes.

RFLP analyses can be used effectively, and relatively economically, to analyze clonal populations, heterozygosity, relatedness, geographic variation, hybridization, species boundaries, and phylogenies ranging in age from 0 to 50 mya (Table 12.3). It is possible to analyze more loci per individual by RFLP analysis than by DNA sequencing because RFLPs are less time-consuming and expensive, although the information provided for each locus is less complete (Dowling et al. 1996, Hall 1998). Higher level systematics studies only rarely have used RFLPs.

More than 1400 restriction enzymes are known that cut DNA at a specific position within a specific recognition sequence. See Chapter 5 for a discussion of restriction digests, as well as Brown (1991), and catalogs from a variety of commercial producers. Most recognition sequences are 4–6 bp long, although they can be as large as 12 bp. The specificity of restriction enzymes means that a complete digestion will yield a reproducible array of DNA fragments. Changes in the

number and size of fragments can occur by changes in DNA sequence by rearrangements (inversions, tandem duplication, and inverted duplication), or addition, deletion, or substitution of specific bases.

Once the DNA is digested with a restriction enzyme, the fragments produced are sorted by size using agarose or polyacrylamide gel electrophoresis. DNA fragments of known length are run on each gel to serve as an internal standard and to allow the size of the experimental fragments to be estimated. The DNA fragments in the gel are visualized by several methods, including staining with ethidium bromide (if the DNA was previously amplified by the PCR), or by probing Southern blots with labeled probes. The detection technique used depends on the amount of DNA present in the gel.

Staining with ethidium bromide is simple and cheap, but least sensitive. The minimal amount of DNA in a band that can be detected by ethidium bromide is  $\approx 2\text{ ng}$ , so small fragments can be detected only if a large amount of DNA is present. DNA probes can be end labeled by adding  $^{32}\text{P}$ -labeled nucleotides to the ends of DNA fragments produced by the restriction enzymes. Intensity of labeling is independent of fragment size and is more sensitive than ethidium bromide (EtBr), with 1–5 ng of DNA easily visualized. If primers are available, DNA can be first amplified by the PCR, cut with a restriction enzyme, and labeled by ethidium bromide (PCR-RFLP).

If less DNA is available, radiolabeled DNA probes can be used to visualize fragments. Southern-blot hybridizations are highly sensitive and picogram quantities of DNA can be detected, although small fragments  $<50\text{ bp}$  in size are more difficult to detect. Southern blots require a suitable probe with sufficient sequence similarity to the target DNA that a stable hybrid can be formed at moderate to high stringency. The use of probes from other species (heterologous probes) makes interpretation of results more difficult.

#### **12.4.4 DNA and Genome Sequencing**

Sequences of proteins, RNA, and DNA have been obtained during the past 40 years. The first sequence information was obtained from proteins in the mid-1950s. RNA was sequenced in the mid-1960s, and DNA sequences were obtained in 1975 after Sanger sequencing methods were developed. More recently, NextGen sequencing methods have greatly reduced the time and costs of sequencing genomes and the third-generation sequencing technologies promise to further reduce costs and time, while increasing read lengths (see Chapter 7).

The use of the PCR makes DNA sequencing of single genes or several gene fragments rapid and relatively inexpensive for systematic studies. Core facilities

and commercial resources conduct automated Sanger sequencing so that few laboratories need to carry out their own sequencing reactions. Now that Next-Generation sequencing methods are available, sequencing of entire genomes can provide multiple DNA sequences and even allow comparisons of entire genomes of related species (comparative genomics).

DNA sequence data can be used to 1) construct molecular phylogenies to evaluate the evolution of particular genes or gene families, 2) evaluate evolutionary changes within species, and 3) construct phylogenies of different species. DNA sequences can be obtained for single-copy genes, mtDNA, ribosomal DNA, and entire genomes. Sequences can be used to study most systematics problems from intraspecific variability to phylogeny of all organisms ([Table 12.3](#)). Sequence data are appropriate for analysis of intraspecific variation, cryptic species, geographic variation, reproductive behavior, and heterozygosity estimates. However, DNA sequencing can be relatively expensive and time-consuming if multiple genes from very large numbers of individuals must be analyzed. Sequence analysis of nuclear or mtDNA sequences provides very large amounts of detailed data. The number of potential characters that can be examined theoretically is limited only by the number of nucleotides in the DNA of the organism.

#### **12.4.5 Fragment Analyses of Genomic DNA**

Fragment analyses, which include RAPD-PCR, single-locus microsatellites or multilocus DNA fingerprinting, can be used for some systematics problems ([Table 12.3](#)). The Random Amplified Polymorphic DNA (RAPD) method of the PCR (described in Chapter 8) has been used to discriminate between cryptic species. Multiple RAPD markers may have to be used to produce a banding pattern that can be analyzed by discriminant analysis, although the need to conduct multiple RAPD reactions would make RAPD-PCR more expensive and time-consuming. RAPD-PCR may be useful for examining hybridization and species boundaries, as well as clonal variation.

Single-locus microsatellites are potentially useful for analysis of population structure, mating systems, clonal boundaries, heterozygosity, paternity testing, relatedness, and geographic variation ([Table 12.3](#)). Multilocus DNA fingerprinting can be used for clonal detection, species delimitation, and paternity testing ([Lunt et al. 1998, 1999](#)).

### **12.5 Targets of DNA Analysis**

Sequence analyses of nuclear, mitochondrial, and ribosomal DNA have been used in systematics studies, as have microsatellites and introns ([Caterino et al.](#)

2000). The following discussion describes some of the attributes of these targets, which are relevant because specific assumptions may be essential for an appropriate phylogenetic-analysis method. More recently, microRNAs have been studied, especially for the analysis of deep evolutionary divergences.

### 12.5.1 Mitochondria

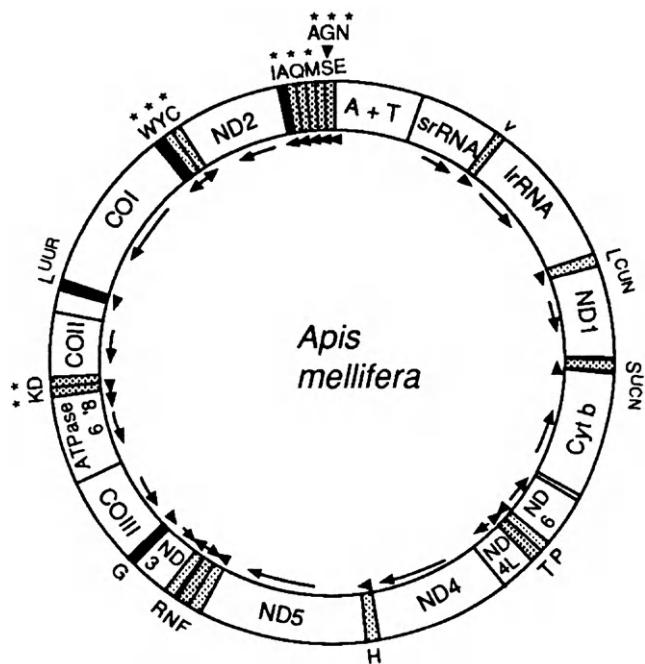
Mitochondria are the cell's respiratory power plant for the generation of ATP. Mitochondria are thought to have developed more than a billion years ago when a free-living eubacterium took up residence within another cell (Margulis 1970). Sequence analysis of modern mitochondrial DNA suggests that  $\alpha$ -Proteobacteria, such as *Rickettsia*, *Anaplasma* and *Ehrlichia*, are the closest contemporary relatives of that eubacterium (Gray et al. 1999, Lang et al. 1999). The relatively low gene content of mtDNA, compared with even the smallest eubacterial genome, indicates that loss or transfer of genetic information occurred at an early stage in the evolution of the "protomitochondrial genome" (Gray et al. 1999). What is not clear is whether mitochondria originated from a single endosymbiotic event or more than one event (Lang et al. 1999). Over the past billion years, mitochondria in Eukaryotes have evolved to the point that genome size and gene content vary among the major groups (Lang et al. 1999).

Mutation rates in mtDNA are variable, with mammalian mtDNA having a mutation rate at least 50 times greater than in plants (Moritz et al. 1987, Lang et al. 1999). The number and type of genes present in mitochondria differ, with perhaps seven to 10 independent losses having occurred for each gene. The evolution of mitochondrial genes over the past billion years has been complex; some genes apparently were transferred independently several times into the nuclear genome; some genes were lost without transfer to the nuclear genome due to gene substitution; and genes were acquired by lateral gene transfer as well (Gray et al. 1999). There is no evidence that, once mitochondrial genes were transferred to the nucleus, they were regained by the mitochondria and there is no evidence for widespread and substantial lateral transfer of genetic information into or between mitochondria (Lang et al. 1999). The predominantly vertical inheritance of genes is a prerequisite for phylogenetic analyses. Within mitochondrial genomes, there are regions that diverge rapidly, whereas other regions are highly conserved, making the different regions suitable for analysis of different taxonomic levels (Simon et al. 1991, Liu and Beckenbach 1992, Tamura 1992, Caterino et al. 2000).

Animal mitochondria are small, circular (16–20 kb in length), and lack introns, with the genes compactly arranged on both DNA strands. With a few exceptions, animal mtDNA contains the same 37 genes coding for small and large subunit

rRNAs, 13 proteins, and 22 transfer RNAs (tRNAs) arrayed in an order that is well conserved within phyla (Figure 12.1). All mtDNAs have at least one noncoding region, which contains regulatory elements for replication and transcription, but intergenic sequences are small or absent (Boore and Brown 1998). The control region containing the origin of replication is extremely A + T-rich in insects. Within the insects, the tRNA genes are known to vary in position (Crozier and Crozier 1993). There are thousands of mitochondria in each cell, so mtDNA is abundant and relatively easy to isolate, even from somewhat-degraded samples. By contrast, nuclear genes evolve more slowly, which makes it possible to extend analyses further into the past.

Mitochondrial DNA can be used in analyses of population structure and gene flow, hybridization, biogeography, and phylogenetic relationships (Avise et al. 1987, Lang et al. 1999). The small size, relatively rapid rate of evolutionary



**Figure 12.1** Map of mitochondrial DNA of the honey bee, *Apis mellifera*. Genes for tRNAs are denoted by the one-letter code for their corresponding amino acids. tRNA genes with asterisk are in a different position compared to the same genes in the mitochondrion of *Drosophila yakuba*. Protein-coding genes are denoted COI, COII, and COIII for the genes encoding subunits one, two, or three of cytochrome *c* oxidase, Cyt *b*, for the cytochrome *b* gene and ND4L for the genes encoding subunits 1–6 and 4L of the NADH dehydrogenase system. The AT-rich region containing the origin of replication is denoted A + T. The direction of transcription for each coding region is shown by arrows. (Redrawn from Crozier and Crozier 1993.)

change, and (usually) maternal inheritance of mtDNA make it suitable for examining population history and evolution among closely related taxa (Lansman et al. 1981, Gray 1989, Simon et al. 1991, Caterino et al. 2000), as well as deeper evolutionary relationships (Lang et al. 1999, Caterino et al. 2000). Molecular studies of mtDNA can be done using RFLPs or sequencing of specific regions of the mtDNA after cloning or amplification by the PCR (Satta and Takahata 1990, Pashley and Ke 1992, White and Densmore 1992). Gene order can be used as a phylogenetic tool (Boore and Brown 1998).

The lack of recombination in mitochondrial DNA means that fixation of an advantageous mutation by selection will cause fixation of all other polymorphisms by “genetic hitchhiking.” Even the quickly evolving noncoding origin of replication region may not have neutral-allele frequencies because this region is linked to the rest of the genome (Ballard and Kreitman 1995). Ballard and Kreitman (1995) point out that violation of the neutral-evolution assumption can have important phylogenetic implications. It violates a major assumption of one phylogenetic analysis method (UPGMA clustering, described in Section 12.6.7). Also, selection or parasite-induced sweeps can mimic the effects of inter-population migration or population bottlenecks (Mooers and Holmes 2000, Yang and Bielawski 2000).

Mitochondria have been used as “molecular clocks” to time the divergences of organisms from each other. However, dating of evolutionary events can be problematic when nonneutral evolution within species is combined with altered rates of evolution in the sister taxon (Ballard and Kreitman 1995). Furthermore, in many cases, it appears that mitochondrial molecular clocks tick at different rates in different lineages and at different times within a lineage.

There are difficulties in working with mtDNA. These include the lack of recombination, which makes mtDNA essentially a single heritable unit. Although recombination in mtDNA has not been observed in insects, it has been found in fungi when **heteroplasmy** (presence of two types of mitochondria) occurs (Saville et al. 1998). Lack of recombination potentially produces gene diversity estimates that have larger standard errors than those determined using nuclear loci that can recombine. Biparental inheritance of mitochondria occurs occasionally in insects, which can complicate population studies (Lansman et al. 1983, Kondo et al. 1990, Matsuura et al. 1991, Fontaine et al. 2007). Introgression of mitochondria between *Drosophila* species has been suggested as an explanation for the presence of mitochondria from a related species (Aubert and Solignac 1990). Fontaine et al. (2007) reviewed cases in which mitochondria were shown to be paternally inherited. Such “paternal leakage” appears to occur most often (or be detected most often) when closely related

species cross. In fact, [Fontaine et al. \(2007\)](#) showed that experimental crosses between three species of periodical cicadas resulted in paternal leakage and that “paternal leakage may be more common than previously estimated.”

The complete mitochondrial sequences of many insects, and other arthropods, are available in GenBank and other databases. A map of *Drosophila yakuba* mitochondria is shown in Chapter 3 (see Figure 3.6), and a map of the mtDNA of the honey bee is shown in [Figure 12.1](#). Honey bee mtDNA is 16,343 bp long and 11 of the tRNA genes are in altered positions compared with their positions in *D. yakuba* ([Crozier and Crozier 1993](#)).

#### *12.5.1.1 The Barcode of Life Project*

DNA barcoding is a method of identifying organisms based on a short, standardized fragment of genomic DNA and has been developed for use by taxonomists, ecologists, conservation biologists, regulatory agencies, and others. The technique was proposed by [Hebert et al. \(2003a,b, 2004\)](#) and involved using a primer set to amplify by the PCR a 648-bp region of the mitochondrial cytochrome c oxidase subunit 1 (COI) gene. Once the DNA is sequenced, it can be analyzed and the original claim was that it would allow identification of all animal species. Barcoding was initiated in 2004, and in 2010 the International Barcode of Life project was developed ([International Barcode of Life 2010](#)). [Miller \(2007\)](#) associated DNA barcoding with the renaissance of taxonomy, because it could lead to more and faster species-identification services and better information about biodiversity. This is relevant because taxonomy around the world is receiving less funding and fewer and fewer scientists can identify species. Barcoding was seen as having the potential to “accelerate our discovery of new species, improve the quality of taxonomic information and make this information readily available to nontaxonomists and researchers outside of major collection centers” ([Miller 2007](#)). [Hebert et al. \(2003a\)](#) suggested the COI sequences would be relevant to identifying all animals, and discriminating between even closely related species. [Zahariev et al. \(2009\)](#) provided algorithms for discovering DNA barcodes from sequence databases. However, some related animal species cannot be identified using the COI sequence and other sequences have been tried to clarify these relationships ([Taylor and Harris 2012](#)).

Many entomologists have used DNA barcoding ([Folmer et al. 1994](#), [Jinbo et al. 2011](#)), and [Taylor and Harris \(2012\)](#) concluded that more arthropods have been subjected to barcoding than any other taxonomic group. There is a Barcode of Life project for the Formicidae, Trichoptera, and Lepidoptera. At least 79,320 insects have been barcoded ([Jinbo et al. 2011](#)). The most useful results have been to accurately identify larval stages of species and to identify small

organisms that require microscopic examination for morphological traits or that have a variety of life stages. [Smith et al. \(2006\)](#) reported that parasitoid tachinid flies have cryptic host specificity when subjected to DNA barcoding.

However, COI sequences in many organisms (algae, protists, and plants) are not informative and indicate that COI is *not* a universal barcode for all life. Furthermore, the method used to separate species with barcode data has been criticized ([DeSalle et al. 2005](#)). The original expectation was that intraspecific divergences in COI sequences are less than interspecific divergences and that there is a “barcoding gap” between species. However, there clearly is not such a gap between all animal species, and the degree of difference in distance that is needed is debated. The use of barcoding to identify new species is particularly controversial if it is the only attribute used to identify the new species. DNA barcoding should create the hypothesis that a population is a new species and additional data (morphological, molecular, behavioral) should be developed to support the hypothesis.

[Yassin et al. \(2010\)](#) used 68 species of *Drosophila* as a model to test DNA barcoding and tree- and character-based methods of species identification. They analyzed 1058 COI sequences of the 68 species and found that “DNA barcoding of *Drosophila* shows no reason to alter the 250 year old tradition of character-based taxonomy.” By contrast, [Ball and Armstrong \(2006\)](#) found that DNA barcodes provided a highly accurate means of identifying 20 lymantriid species. [Derocles et al. \(2012\)](#) studied 50 species of parasitoids that attack aphids in northwestern Europe. They used both COI and a nuclear gene (long wavelength rhodopsin) for their analyses and found that some species were indistinguishable on the basis of their COI sequences, whereas the nuclear gene failed to discriminate between other species and concluded that “no unique locus but a combination of two genes should be used to accurately identify members of Aphidiinae.”

To work well, a library of DNA barcodes needs to be developed so that identifications can be made. This requires developing a large-scale database and the ability to access voucher specimens for confirmation of identifications ([Borisenko et al. 2009](#)). Thus, specimens are collected, prepared, sequence data obtained, data are entered into databases, and the collection managed so that data and specimens can be compared.

DNA barcoding is a “method of identifying previously described taxa” and “reference sequences lie at the very heart of the DNA barcoding initiative” because “Without verified reference sequences from voucher specimens that have been authenticated by qualified taxonomists, there is no reliable library for newly generated query sequences to be compared with” ([Taylor and Harris](#)

2012). Thus, in the view of Taylor and Harris (2012), DNA barcoding has actually reinforced the need for qualified taxonomists by producing sequence data that needs to be paired with a verified morphological type specimen. Taylor and Harris (2012) suggested that Next-Generation sequencing will determine the future of barcoding, because inexpensive sequencing machines may allow the development of more data that can be used to identify species, and that “DNA barcoding could soon become obsolete.”

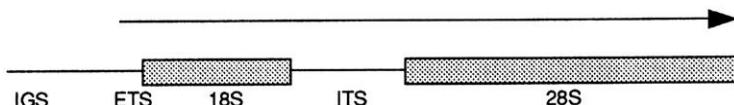
Another concern about DNA barcodes is the discovery of **nuclear mitochondrial pseudogenes (numts)**. Numts are nonfunctional copies of mtDNA in the nucleus that can be found in a number of eukaryotes and can be coamplified with mtDNA when using conserved universal PCR primers. Numts can be highly divergent from the mtDNA sequences in the individual and sequencing of numts, using the standard of 3% sequence divergence, can result in erroneous estimates of species diversity. More than 82 eukaryotes are known to have numts, and it appears that numts may be more common than realized, providing a challenge to DNA barcoding as an accurate method for species identification (Song et al. 2008). It was probably unrealistic to expect that a single gene fragment could resolve all species-identification issues.

### 12.5.2 Ribosomal RNA

Ribosomes are a major component of cells that are involved in translating messenger RNA (mRNA) into proteins. Ribosomes consist of ribosomal RNA (rRNA) plus proteins. All ribosomes can be dissociated into two subunits, each containing rRNA and proteins. Ribosomal RNAs are used frequently to evaluate evolutionary relationships among species because they are universally present and abundant in cells. Ribosomal RNAs contain regions that are “conserved” and regions that are more variable, so rRNAs can serve as both slow and fast clocks.

In eukaryotes, the genes encoding the 18S (small subunit) and 28S (large subunit) rRNAs are clustered as tandem repeats in the nucleolus-organizing regions of the nuclear chromosomes (Figure 12.2), but two ribosomal genes are found in mitochondria (Figure 12.1). In most animals, there are 100–500 copies of rDNA in the nuclear genome in tandemly repeated transcription units. Ribosomal gene copy number ranges from as few as 45 in the fly *Sciara coprophila* to >3000 in the grasshopper *Locusta migratoria*. In a survey of 30 species of mosquitoes, copy numbers ranged from 39 to 1023 (Kumar and Rai 1990).

The repeated transcription unit of rRNAs is composed of a leader promoter region known as the **External Transcribed Spacer (ETS)**, an 18S rRNA coding region, an **Internal noncoding Transcribed Spacer** region (**ITS**), a 28S



**Figure 12.2** A simplified diagram of the ribosomal DNA repeat unit of eukaryotes. IGS is the intergenic spacer region. ETS is the external transcribed spacer, 28S is the large subunit rRNA gene. The arrow indicates the direction of transcription. Most insects have hundreds of ribosomal RNA genes in tandem array. Some contain R1 and R2 retrotransposable elements (not shown) in specific locations. Genes with R1 and R2 elements produce nonfunctional product.

rRNA coding region, and an **InterGenic nontranscribed Spacer segment (IGS)** (Figure 12.2). Different portions of the repeated transcription unit evolve at different rates. In general, a higher degree of polymorphism has been found in the *noncoding segments* (ETS, ITS, and IGS). The most variable part of the repeated unit is the intergenic spacer (IGS); it typically contains reiterated subrepeats ranging from about 50 to several hundred bp in length (Cross and Dover 1987). The coding regions of the repeated unit change relatively little, and can be used for systematic studies of higher-level taxa or for ancient lineages. Highly conserved regions are no doubt important for maintaining the characteristic secondary and tertiary structure of rRNA molecules (Simon et al. 1991, Van de Peer et al. 1993, Caterino et al. 2000).

Ribosomal RNA genes undergo concerted evolution so that the sequence similarity of members of an rRNA family is expected to be greater within a species than between species. Unequal crossing over, gene conversion, and illegitimate recombination are responsible for concerted evolution. Ribosomal gene families are considered to be “quite uniform” (Ohta 2000a). Two retrotransposons, called R1 and R2, have been found in the 28S rRNA genes of most insects (Eickbush 2002). These elements have been associated with arthropods for >500 million years, and they usually are located at the same position within the 28S gene. Most R2 elements are located  $\approx$ 74 bp upstream from the site of R1 insertions. R1 and R2 elements lack long terminal repeats (LTRs) and block the production of functional rRNA (Eickbush 2002). The insect survives because it contains hundreds of rRNA genes and the R2 elements are kept from invading too many of them by unknown mechanisms. Surprisingly, most R1 and R2 elements have not accumulated mutations that would make them inactive. Some species have more than one family of R1 or R2 elements, and sequence identity between the different families can be low, suggesting either that each insertion family is able to maintain its copy number without eliminating other families or that there has been horizontal transfer of R1 and R2 elements between species. A phylogenetic analysis of R2 elements and arthropods suggests that multiple

lineages of R2 elements evolved in arthropods and these have been maintained ([Eickbush 2002](#)).

### **12.5.3 Satellite DNA**

Satellite DNA may comprise a large fraction of the total DNA in an arthropod. Microsatellites are usually species specific, perhaps because this DNA evolves at a very high rate. There are only a few cases in which the same satellite sequences have been found throughout an entire genus. Satellite DNA can be used for species diagnoses or analyses of populations ([Bachmann et al. 1993](#), [Lunt et al. 1999](#), [Caterino et al. 2000](#), [Lumley and Sperling 2011](#)). Satellite DNA has most often been used in population ecology and is discussed in Chapter 13.

### **12.5.4 Introns**

Introns within single-copy nuclear genes are perceived to be highly variable and can be used in taxonomic studies ([Caterino et al. 2000](#)). The use of introns to resolve origins of invasive populations of the Mediterranean fruit fly is described in Chapter 13.

### **12.5.5 Nuclear Protein-Coding Genes**

A variety of protein-coding loci have been used in molecular systematics ([Friedlander et al. 1992](#), [Caterino et al. 2000](#)). Nuclear genes exhibit a wide range of evolutionary rates, and the availability of entire genome sequences for many arthropod species has dramatically increased the number and type of genes that can be used in phylogenetic analyses.

Problems with nuclear DNA sequences used for phylogenetic analysis include the fact that they may be heterozygous; they also are present in low copy number, which may make them difficult to amplify by the PCR. Furthermore, many genes contain large introns that make it difficult to amplify more than one exon unless **Reverse Transcriptase-PCR (RT-PCR)** is carried out on mRNA. Caution also is warranted: many single-copy loci actually are present in more than one copy. Furthermore, pseudogenes (inactive forms of a gene) may create problems if comparisons are made inadvertently between genes and pseudogenes.

### **12.5.6 Rare Genomic Changes**

DNA sequence data are used most often to construct phylogenies. However, [Rokas and Holland \(2000\)](#) suggest that single-nucleotide substitutions may not always be informative and argue that rare genomic changes such as **indels** (an insertion or deletion), retroposon integrations, signature sequences,

mitochondrial and chloroplast-gene order changes, or gene duplications and genetic-code changes provide useful information with “enormous potential for molecular systematics.”

As an example, [Rokas and Holland \(2000\)](#) reviewed research conducted to resolve the relationship of the Strepsiptera, Diptera, and Coleoptera. Strepsipteran forewings resemble the hind wing balancing organs (halteres) of Diptera. Under one scenario, dipteran (hind wing) halteres could be homologous to the front wings of Strepsiptera if a homeotic mutation reversed the position of the structures in Strepsiptera. By contrast, some would place the Strepsiptera closer to the Coleoptera because both use the hind wings for flight. Analysis of 18S rDNA sequence data did not resolve the question. However, a unique intron insertion was found in the homeobox of the *engrailed* gene of Diptera and Lepidoptera, that is absent from other insects and other outgroups. If Strepsiptera had the intron, it would support a sister-group relationship with Diptera, but its absence would not. Cloning of the strepsipteran homolog of *engrailed* found the intron was absent, indicating that halteres of Strepsiptera and Diptera are more likely a case of convergent evolution. [Wiegmann et al. \(2009\)](#) concluded that Strepsiptera are closely related to the Coleoptera (rather than the Diptera) using both molecular and morphological data. As noted by [Minelli \(2009\)](#), the study by [Wiegmann et al. \(2009\)](#) demonstrated the importance of integrating phylogeny and knowledge of the evolvability of developmental mechanisms. [Minelli \(2009\)](#) recommended that phylogenetics be combined in the future with evolutionary developmental biology or “phylo-evo-devo” to resolve other phylogenies.

Another example in which a rare genomic change may provide useful phylogenetic information involves the gene order in mitochondria of insects, Crustacea, and Myriapoda ([Boore et al. 1998](#)). The mitochondria of both crustaceans and insects share a changed gene order, suggesting that myriapods are an outgroup.

### 12.5.7 MicroRNAs

[Erwin et al. \(2011\)](#) pointed out that studies of comparative genomics and developmental patterning have changed our perception of the early evolution of animals. “First whole-genome sequencing of dozens of metazoans has demonstrated that any animal requires only about 20,000 protein-coding genes for the production of its essential morphologic architecture. Second, much of this protein-coding repertoire ... is conserved throughout all metazoans and is even found today among single-celled opisthokonts.” [Erwin et al. \(2011\)](#) further state, “the last common ancestor of metazoans ... was a genetically complex animal possessing all of the families of protein-coding genes used during

development. ... Consequently, the morphological simplicity of basal animals, and the great differences in morphology between sponges and arthropods or vertebrates, cannot be due to the absence of these protein-coding gene families but instead must involve differences in the temporal and spatial deployment of these genes and their regulation." One possible mechanism is the "continual evolutionary addition of microRNAs, which code for 22-nucleotide regulatory RNAs that affect translation of target mRNAs, resulting in homeostasis and cell identity. MicroRNAs have been continuously added to eumetazoan genomes through time with very little loss in most taxa. When losses did occur, the loss seems to be associated with morphological simplification."

[Dolgin \(2012\)](#) reported on three analyses of evolutionary biology based on analyses of microRNAs. [Sempere et al. \(2006\)](#), [Peterson et al. \(2009\)](#), and [Janvier \(2010\)](#) analyzed animal phylogeny using microRNA sequences. MicroRNAs "are either there or they aren't," and "once gained, microRNAs usually remain functional, which means that their signal stays intact for hundreds of millions of years" ([Dolgin 2012](#)). The hypothesis is that microRNAs hold the secret to morphological complexity. The farther away from the base of the evolutionary tree the animals are, the more microRNAs they have accumulated. This fact could provide "a brand new way to do phylogeny, using a set of rare genomic characters that no one had ever considered before" ([Dolgin 2012](#)). According to an analysis by [Peterson et al. \(2009\)](#), "778 microRNA families have arisen during the 600 million or so years of animal evolution, and only 48 have been lost. This pattern of inheritance leaves an easy-to-follow evolutionary trail for phylogenetic sleuths."

This approach to deep phylogeny is controversial and, in the case of mammalian evolution, contradicts the tree developed using other genetic data. [Campbell et al. \(2011\)](#) used microRNAs to analyze the relationships of the Tardigrada and Arthropoda. Morphological data group Tardigrada (water bears) and Onychophora (velvet worms) and Arthropoda into a monophyletic group known as the Panarthropoda. Molecular data do not support the inclusion of tardigrades within the Panarthropoda. However, analysis of microRNAs and ESTs support a monophyletic Panarthropoda that includes tardigrades.

[Rota-Stabelli et al. \(2011\)](#) used a combination of microRNAs and morphology to support the monophyletic status of the Mandibulata. Morphology indicates that myriapods, insects and crustaceans form a monophyletic group, the Mandibulata, but other studies indicate other interpretations. They studied 198 protein-coding genes and microRNAs to indicate that the Mandibulata is monophyletic. Two microRNAs are present and expressed in all mandibulates studied, but not in chelicerates.

## 12.6 Steps in Phylogenetic Analysis of DNA Sequence Data

An ongoing need in molecular systematics and evolution studies is to resolve which genes are informative for which questions. The wealth of information obtained from DNA sequences can provide insights into evolution and speciation, but how does one choose appropriate genes for a specific problem? How can estimates of genetic distance be used to make judgments about species status or date of speciation? Several concepts are important in using various phylogenetic programs to resolve the relationships of different taxa (Gibson and Muse 2002).

### 12.6.1 Gene Trees or Species Trees

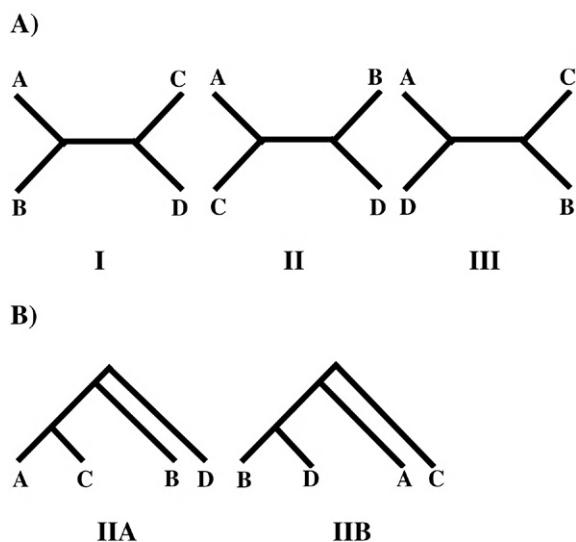
Phylogenetic analysis of a particular genetic locus may not agree with the species phylogeny (Caterino et al. 2000). The incongruence may be due to the horizontal movement of genes, duplication and extinction of one of the genes, or to lineage sorting (deep coalescence). For example, the twospotted spider mite, *Tetranychus urticae*, has *fungal* carotenoid genes in its genome (Altincicek et al. 2012), as do aphids (Moran and Jarvik 2010), presumably through horizontal transfer. Mitochondrial genes may be more reliable than nuclear genes for evaluating some recent divergences (Caterino et al. 2000).

### 12.6.2 Rooted or Unrooted Trees

Many phylogenetic methods produce unrooted trees. Information regarding evolutionary rates or the most ancient relationships is needed to root the inferred trees. A comparison of the two concepts is shown in Figure 12.3. For any four taxa (the tips of the branches), there are three distinct unrooted trees (I, II, and III). Each unrooted tree can be rooted on any of its five branches; two of the possible rooted trees for the center, unrooted, tree are shown (IIA and IIB).

### 12.6.3 Tree Types

The immense diversity of insects and their long evolutionary history provides a challenge for systematists. Due to mutation, high reproductive rates, natural selection and stochastic events, populations change through time. A process of gradual change through thousands of years can result in a different species; a change within a single lineage is called **phytic speciation**. Speciation also can occur through **cladogenic speciation**, in which two populations of a species become isolated and diverge genetically due to independent mutation, natural selection, and genetic drift. Other models for speciation include speciation through hybridization and polyploidy, or by modification of regulatory genes. The attributes of an organism used by systematists to establish their relationship



**Figure 12.3** For any four taxa (A, B, C, D; the tips of the branches), there are three different unrooted trees (I, II and III at left of figure). Each unrooted tree can be rooted on any of its five branches; two of the possible five rooted trees for unrooted tree II are shown at the right (IIA, IIB). Rooted trees IIC, IID, and IIE are not shown. (Redrawn from Gibson and Muse 2002.)

to other organisms are called **characters**. Characters can be based on morphology, physiology, ecology, behavior, biochemistry, or genetics.

There are several approaches to developing classifications using these characters. Unfortunately, debates remain over which approach is more objective, appropriate, or practical. **Phenetic systematics** focuses on overall similarities among organisms, involves all possible characters, and calculates average similarities with all characters assumed to be equally useful. In some cases, classifications based on phenetic similarities may reflect the phylogeny of taxa because those that are most similar may well have shared a most-recent ancestor, but this need not be so because of convergent evolution.

**Cladistic (phylogenetic) systematics** uses only cladistic relationships as a basis for constructing classifications (Hennig 1966). The rate or amount of change is not considered and only monophyletic taxa are allowed. This approach focuses on the order of origin of lineages and also takes into account the amount and nature of evolutionary change that occurs after cladogenesis. Characters are not assumed to be equal, and are weighted accordingly. One of the major difficulties in any reconstruction of phylogeny is to determine which character is primitive or ancestral (**plesiomorphic**), and which is derived (**apomorphic**).

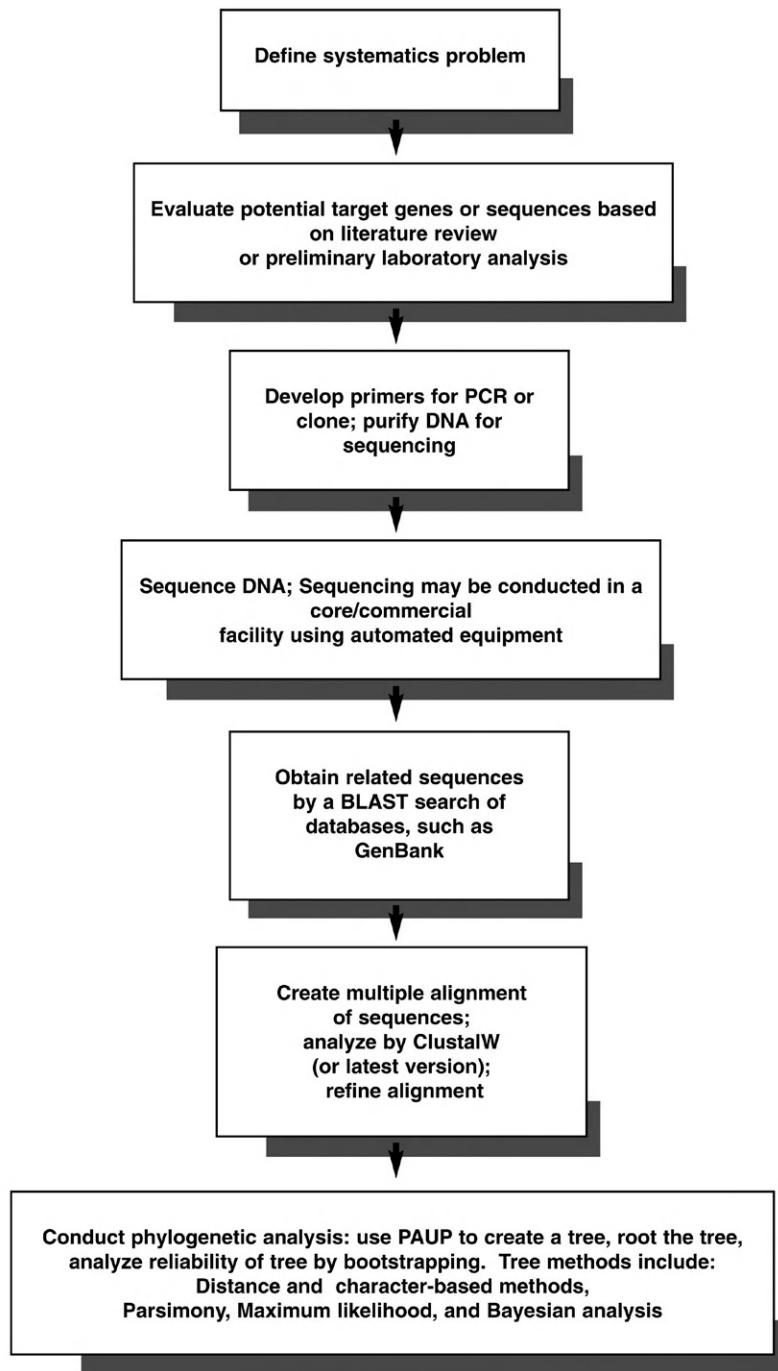
Classifications are often represented in graphical forms as tree-like dichotomous branching graphs or **dendograms**. A dendrogram produced from phenetic information is called a **phenogram**. A phenogram shows how similar the group is, but it does not provide information about probable lines of descent. When a dendrogram is produced from cladistic information it is called a **cladogram**. A cladogram shows the sequence of origin of clades and indicates the times at which the various cladogenic events have taken place. If the dendrogram includes both phenetic and phylogenetic data it is a phylogram, or **phylogenetic tree**, and indicates the cladistic branching, as well as the relative amount of change that has occurred. Those species that show the closest relationship are grouped together into larger, more-inclusive groups or genera. Genera are grouped into families, and families into orders, classes, and phyla.

#### 12.6.4 Project Goals and Appropriate DNA Sequences

The first step is to consider carefully the project goals and to evaluate published information as to which genes/DNA sequences may be most appropriate to answer the question (Figure 12.4). Analysis of the evolution of orders will require different approaches than analysis of species within a genus. Whether a specific DNA sequence is appropriate for a particular project can be difficult to predict in advance unless a survey has been, or can be, conducted to determine whether the appropriate level of variability is present.

Analyses of different DNA sequences provide information about different levels of phylogenetic analyses over a broad range of taxa. For example, ribosomal genes are widely used because they are highly conserved, but have regions that change rapidly and regions that change slowly. Once the target gene or other DNA sequence has been chosen, primers must be designed or made to amplify the target sequence by the PCR or a cloning strategy developed. Cloning of target DNA may be required if inadequate sequence information is available in the literature or GenBank and if “universal primers” are unavailable.

Once insects have been obtained by collecting or from museums, DNA must be extracted. As noted in Chapter 8, PCR results vary with the preservation method. Old, dried museum specimens are likely to contain degraded DNA, so using DNA sequences that are present in multiple copies (such as mitochondrial or ribosomal DNA) may be more appropriate than using single-copy nuclear genes. DNA extraction results are easier with freshly collected, frozen (at  $-80^{\circ}\text{C}$ ), or alcohol-preserved (95% ethanol [EtOH]) insects. However, a method developed for isolating DNA from fossils without destroying the specimen has been adapted to arthropods (Hoss and Paabo 1993, Rowley et al. 2007, Jeyaprakash and Hoy 2010). This method involves soaking intact specimens in a guanidinium



**Figure 12.4** Steps in the phylogenetic analysis of DNA sequence data.

thiocyanate (GuSCN) or guanidinium hydrochloride (GuHCl) buffer, which destabilizes nucleases. By using a silica matrix, which binds nucleic acids, specimens can yield DNA for subsequent PCR amplification yet are suitable for examination of their morphology (Jeyaprakash and Hoy 2010).

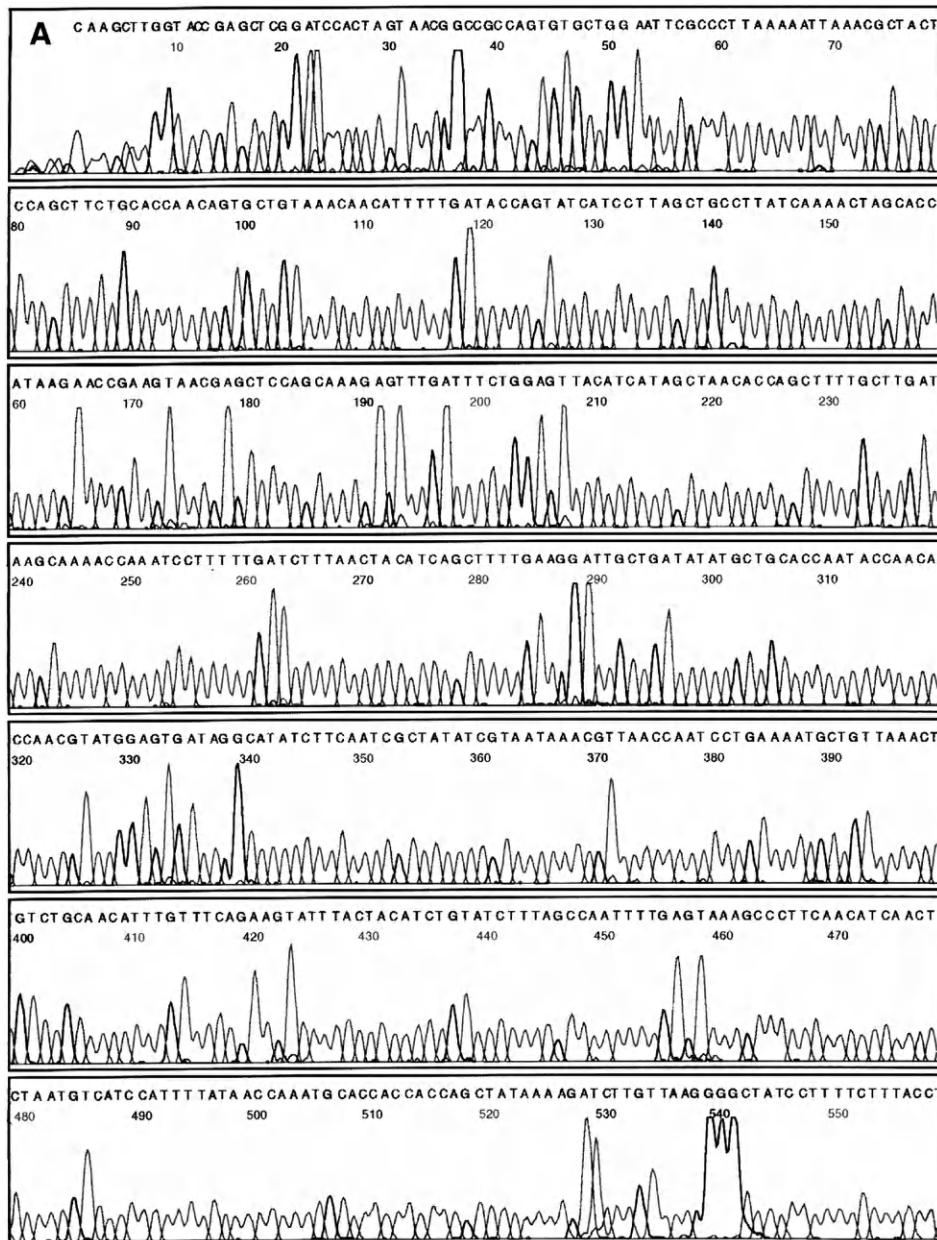
The sequencing outcome depends upon the purity of the DNA used and the fidelity of the sequencing procedures. It is important to sequence both strands to detect errors. Once sequence information has been obtained, analysis can provide several types of information, including possible structure, function, and characteristics of the protein. The similarity of the sequences to sequences obtained from other organisms can be compared. The tasks of collating, assembling, and correcting the sequence data are usually performed with the help of a variety of computer programs (Gribskov and Devereux 1991, Eernisse 1998, Fortna and Gardiner 2001, Mount 2001, Felsenstein 2004, Knowles and Kubatko 2010, Hall 2011).

Analysis of sequence data involves several steps (Figure 12.4). Figure 12.5A shows the form in which Sanger sequence data is obtained from a sequencing laboratory. Figure 12.5B shows the data after they have been analyzed to show the open reading frame (ORF) of the coding strand (the line with the •), the codons, and the region at the end of the sequence that represents the vector DNA, which should be excluded from the subsequent analysis.

#### 12.6.5 Sequence Comparisons with BLAST

Once sequences have been obtained, the scientist usually wishes to compare them with other sequences in the databases (Figure 12.4). Genetic sequences are stored in three major databases: GenBank in the United States, the EMBL Data Library in Europe, and the DNA Data Bank of Japan. The amount of sequence information available has grown exponentially and continues to accumulate at an ever-faster pace. The importance of submitting DNA sequence data to public databases is recognized and most journals require that sequences be submitted before, or simultaneous with, publication.

Sequences can be obtained from these databases using a computer program called **Basic Local Alignment Search Tool (BLAST)**. BLAST (and its sub-types) is one of the most widely used tools in phylogenetic analysis. BLAST is used to search large databases of DNA (or amino-acid) sequences, returning sequences that have regions of similarity to the sequence of interest provided by the user (query sequence) (Fortna and Gardiner 2001, Gibson and Muse 2002). The goal is to find regions in sequence pairs that have high levels of similarity. The results of a BLAST search orders the sequences and provides an e-value.



**Figure 12.5** A) Sanger DNA sequence data are obtained from an automated sequencing facility in this format. The different peaks representing A, T, C, and G are printed in different colors but are shown in black and white here. This sequence is the noncoding sequence, thus the sequence to be analyzed must be transformed into the complementary coding sequence.

**B ORF → Coding Strand**

18            27            36            45            54

5' AGG TAA AGA AAA GGA TAG CCC CTT AAC AAG ATC TTT TAT AGC TGG TGG TGC

Arg \*\*\* ... \*\*\* ...

- Gly Lys Glu Lys Asp Ser Pro Leu Thr Arg Ser Phe Ile Ala Gly Gly Ala Val Lys Lys Arg Ile Ala Pro \*\*\* ... \*\*\* ...

63            72            81            90            99            108

ATT TGG TTA TAA ATT GGA TGA CAT TAG AGT TGA TGT TGA AGG GCT TTA CTC AAA

... \*\*\* ... \*\*\* ... \*\*\* ... \*\*\* ... \*\*\* ...

- Phe Gly Tyr Lys Met Asp Asp Ile Arg Val Asp Val Glu Gly Leu Tyr Ser Lys Met Thr Leu Glu Leu Met Leu Lys Gly Phe Thr Gln Asn

117            126            135            144            153            162

ATT GGC TAA AGA TAC AGA TGT AGT AAA TAC TTC TGA AAC AAA TGT TGC AGA CAG

... \*\*\* ... \*\*\* ... \*\*\* ... \*\*\* ...

- Leu Ala Lys Asp Thr Asp Val Val Asn Thr Ser Glu Thr Asn Val Ala Asp Ser Trp Leu Lys Ile Gln Met \*\*\* \*\*\* ... ... Met Leu Gln Thr Val

171            180            189            198            207            216

TTT AAC AGC ATT TTC AGG ATT GGT TAA CGT TTA TTA CGA TAT AGC GAT TGA AGA

... \*\*\* ... \*\*\* ... \*\*\* ... \*\*\* ...

- Leu Thr Ala Phe Ser Gly Leu Val Asn Val Tyr Tyr Asp Ile Ala Ile Glu Asp \*\*\* ...

225            234            243            252            261            270

TAT GCC TAT CAC TCC ATA CCT TCG TGT TCG TAT TCG TGC AGC ATA TAT CAG CAA

... ... ... ... ... ...

- Met Pro Ile Thr Pro Tyr Val Gly Val Gly Ile Gly Ala Ala Tyr Ile Ser Asn

279            288            297            306            315            324

TOC TTC AAA AGC TGA TGT AGT TAA AGA TCA AAA AGG ATT TGG TTT TGC TTA TCA

... \*\*\* ... \*\*\* ... \*\*\* ... \*\*\* ...

- Pro Ser Lys Ala Asp Val Val Lys Asp Gln Lys Gly Phe Gly Phe Ala Tyr Gln Met \*\*\* ...

333            342            351            360            369            378

AGC AAA AGC TGG TGT TAG CTA TGA TGT AAC TCC AGA AAT CAA ACT CTT TGC TGG

... \*\*\* ... \*\*\* ... \*\*\* ... \*\*\* ...

- Ala Lys Ala Gly Val Ser Tyr Asp Val Thr Pro Glu Ile Lys Leu Phe Ala Gly Met Met \*\*\* ...

387            396            405            414            423            432

AGC TCG TTA CTT CGG TTC TTA TGG TGC TAG TTT TGA TAA GGC AGC TAA GGA TGA

... \*\*\* ... \*\*\* ... \*\*\* ... \*\*\* ...

- Ala Arg Tyr Phe Gly Ser Tyr Gly Ala Ser Phe Asp Lys Ala Ala Lys Asp Asp Met Val Leu Val Ile Arg Gln Leu Arg Met Ile

441            450            459            468            477            486

TAC TGG TAT CAA AAA TGT TGT TTA CAG CAC TGT TGG TGC AGA AGC TGG AGT AGC

... ... ... ... ... ...

- Thr Gly Ile Lys Asn Val Val Tyr Ser Thr Val Gly Ala Glu Ala Gly Val Ala Leu Val Ser Lys Met Leu Phe Thr Ala Leu Leu Val Gln Lys Leu Glu \*\*\* ...

495 Vector Sequence → 513            522            531            540

GTT TAA TTT TTA AGG CGG AAT TCC AGC ACA CTG GCG GCC GIT ACT AGT GGA TCC

... \*\*\* ... \*\*\* ... \*\*\* ... \*\*\* ...

- Phe Asn Phe \*\*\* ...

549            558

GAG CTC GGT ACC AAG CTT G 3'

... ... ... ...

**Figure 12.5 (Continued) B)** Once the coding sequence (● marked) is obtained, the sequence must be aligned by a program such as MacDNASIS and the coding strand analyzed. The open reading frame (ORF) is designated and the beginning of the cloning vector sequence is shown at the end.

The e-value is the number of hits with the same level of similarity that would be found by chance if there were no true matches in the database; thus, an e-value of 0.01 would occur once every 100 searches even when there is no true match in the database. BLAST searches can be run over the web through the National Center for Biotechnology Information (NCBI), the European Biotechnology Institute (EBI), or the DNA database of Japan (DDBJ). Once the sequences have been obtained with which the data are to be compared, they need to be aligned.

### **12.6.6 Aligning Sequences**

Sequences can be aligned either with other sequences obtained in the project or with sequences obtained from databases such as GenBank ([Figure 12.4](#)). Aligning the sequences usually involves computer analyses of the sequences using one of three major methods for comparing sequence similarity: matrix plots, global alignments, and local alignments ([Hillis et al. 1990, 1996](#)). Both alignment and phylogenetic inferences involve assumptions and subjective decisions ([Hillis et al. 1990, 1996](#); [Howe and Ward 1989](#); [Gribskov and Devereaux 1991](#); [Hall 2011](#)). The alignments usually are made based on the assumption of parsimony. **Parsimony** dictates that an alignment of sequences is based on the minimal number of changes needed to transform one sequence into the other.

**ClustalW** is a commonly used program that aligns DNA (or amino-acid) sequences in such a way as to maximize the number of residues that match by introducing gaps or spaces into one or the other sequence. These gaps are assumed to be due to insertions or deletions that occurred as the sequences diverged from a common ancestor over evolutionary time ([Thompson et al. 1994](#), [Hall 2011](#); [Larkin et al. 2007](#)).

### **12.6.7 Constructing Phylogenies**

What is a tree? It is a method to illustrate relationships among organisms, and trees can be portrayed in several ways ([Page 2011](#)). The number of species in the tree may be in the tens or tens of thousands. The tree should provide a pattern of ancestry, divergence and descent, using branches that merge at points that represent common ancestors, each of which is connected through more-distant ancestors. The more ancestors that two species share, the more closely related they are. If two species share a common ancestor but that ancestor is not shared by any other taxa on the tree, these species are known as “sister taxa.” If a species is not linked to any of the other species (other than by a distant ancestor), it is considered an “out group” ([Gregory 2008](#)).

The primary methods of phylogeny construction are parsimony, distance, likelihood, and Bayesian, with variants within each of these broad categories. The goal of all methods is to identify the relationships (topology of a tree) that

are most congruent with the observed data. However, all will be *estimates* of the “true” tree; we cannot be sure that our estimate is truly accurate in depicting the evolutionary relationships. Many reviews of phylogenetic methods are available (Felsenstein 1988, 2004; Swofford and Olson 1990; Stewart 1993; Hillis et al. 1996; Swofford et al. 1996; Pagel 1999; Shoemaker et al. 1999; Fox et al. 1999; Steel and Penny 2000; Fortna and Gardiner 2001; Huelsenbeck et al. 2001; Arbogast et al. 2002; Gibson and Muse 2002; Swofford 2002; Bergsten 2005; Hall 2011; Ronquist and Deans 2010), and providing detailed procedures for constructing phylogenies is beyond the scope of this chapter. Only a brief outline of the different approaches is provided. Details of phylogenetic methods should be obtained from the above-cited reviews, books, and “how-to” manuals. Be aware that the methods for inferring phylogenetic relationships from molecular data continue to evolve. Phylogenetic analysis involves knowledge of statistics, computers, and mathematics, including calculus and matrix algebra; previous exposure to the theory of quantitative genetics is useful (Felsenstein 2004).

Inferring a phylogeny is an estimation procedure and is based on incomplete information. Any study of DNA sequences sampled from different species or different individuals in a population is likely to start with a phylogenetic analysis. Thus, phylogenetic analysis is now common in biology, but a novice will be frustrated by the fact that there are so many different approaches and differences among experts.

The selection of one or more trees from among the set of possible phylogenies is based on one of two approaches: 1) defining a specific sequence of steps, an algorithm, for constructing the best tree; or 2) defining a criterion for comparing alternative phylogenies to one another and deciding whether they are equally good, or one is better. Some methods of phylogeny construction are based on different explicit evolutionary assumptions, whereas others are not.

Phylogenetic trees represent evolutionary pathways and there is a difference between species trees and gene trees (Goldstein and Harvey 1999). Branches in a species tree join extant species to an ancestral species and represent the time since those species diverged. The data used to construct the tree often represent a single region of the genome of those species. A gene tree constructed from a short region of the genome may not be the same as the species tree. Two species may carry genes that diverged before the species split, or introgression or transposition may have resulted in genes having diverged after the species split.

Phylogenies are presented as rooted or unrooted trees. A **rooted tree** conveys the temporal ordering of the species or genes on a tree, but an **unrooted tree** reflects the distances between units with no notion of which was ancestral to which. Some

analytical techniques result in an unrooted tree or unrooted phylogeny, one in which the earliest point in time is unidentified (Figure 12.3). In molecular phylogenies, **branch length** is the average number of nucleotide substitutions per site. If a branch length is 0.2 then, on average, the site has undergone 0.2 changes. Because a nucleotide changes or it doesn't, this average is based on 0 or 1 change.

Molecular data used to construct trees are either discrete characters or similarities (**distances**). Examples of discrete molecular characters include DNA sequences, allozyme frequencies, or restriction-map data. Most methods assume independence and homology among discrete characters. Distance data specify a relationship between pairs of taxa or molecules. Sequence, restriction map, and allozyme data must be transformed to produce distance data. Once data have been gathered and transformed into appropriate values there are four broad categories of methods to estimate phylogeny. These include distance-matrix methods, maximum-parsimony methods, and maximum-likelihood methods, which are discussed in detail by Swofford and Olson (1990), Weir (1990), Hillis et al. (1996), Nei (1996), Huelsenbeck and Rannala (1997), Steel and Penny (2000), Whelan et al. (2001), and Hall (2011). A more recent addition to phylogenetic analysis involves Bayesian inference (Shoemaker et al. 1999, Huelsenbeck et al. 2001).

**Distance-matrix methods** are based on the set of distances calculated between each pair of species and this is the oldest family of phylogenetic reconstruction methods. The computations are relatively simple and the quality of the resulting tree depends on the quality of the distance measure. Using distances to group the taxonomic units into a phenetic grouping usually uses clustering.

Several methods of clustering can be used, but the most widely used is called **Unweighted Pair-Group Method using an Arithmetic Average (UPGMA)**. It defines the intercluster distance as the average of all the pairwise distances for members of two clusters. The results of the clustering can be presented in a dendrogram, in which the branch points are placed midway between two sequences or clusters. The distance between a pair of sequences is the sum of the branch lengths. The UPGMA method often is used for distance matrices, and it generally performs well when the mutation rates are the same along all branches of the tree. However, the assumption of nearly equal mutation rates (or that a molecular clock is operating) is crucial for the UPGMA method.

For situations in which the assumptions of the molecular clock are inappropriate, the Fitch-Margoliash algorithm can be used (Weir 1990). If information for an out-group is available, the resultant tree can be rooted. The Fitch and Margoliash method allows for the possibility that the tree found is incorrect and recommends that other trees be compared based on a measure of goodness

of fit. The best tree will have the smallest percentage standard deviation. The Fitch–Margoliash and UPGMA methods should result in very similar trees if a molecular clock is operating.

**Maximum-parsimony methods** focus on the character values observed for each species, rather than working with the distances between sequences that summarize differences between character values. These methods *minimize* the numbers of changes in sequences between species over the tree, usually making the assumption that there have been approximately constant rates of change. Branch lengths usually are not obtained. Maximum parsimony is widely used and works well when change is rare or branches are short (Pagel 1999). Parsimony methods can work poorly when rates of character evolution are high and the phylogeny includes some long branches because it tends to underestimate the amount of change in long branches. In some circumstances, maximum-likelihood and parsimony methods can provide equivalent results.

For each possible parsimony tree, the sequences at each node are inferred to be those that require the least number of changes to give each of the two sequences for the immediate descendants. The total number of changes required over the whole tree is found, and the tree with the minimum number of changes is the most parsimonious. Parsimony methods assume that genetic changes are improbable. However, if there are large amounts of change, parsimony methods can yield estimated trees that are inaccurate (Swofford and Olson 1990). Stewart (1993) pointed out that parsimony analysis can be problematic for two general reasons: 1) failure to find the shortest tree and 2) the shortest tree is not the correct phylogeny. Failure to find the shortest tree can occur if too many taxa or too few informative data are used.

**Likelihood methods** of analyzing DNA data rely on genetic models, and can provide a basis for statistical inference. Likelihood is an amount proportional to the probability of observing the data, given a model. Likelihood methods are more difficult to compute than the methods described above (Weir 1990). Maximum-likelihood methods of tree construction assume the form of the tree and then choose the branch length to maximize the likelihood of the data given that tree. These likelihoods then are compared over different possible trees and the tree with the greatest likelihood is considered to be the best estimate.

Unfortunately, the number of possible trees increases very rapidly as the number of taxa under consideration increases in likelihood methods. Thus, if three species are being compared, the number of possible unrooted trees is one, with four species it is three trees, with six species it is 105 trees, and with eight

species, it is 10,395 trees. Maximum-likelihood methods provide consistent estimates of branch lengths, indicating that the estimates approach the true values as the amount of data increases. To estimate the likelihood that a particular tree estimate is the true tree, **bootstrapping** techniques can be used. Bootstrapping involves repeated sampling, with replacement, of artificial data sets to produce an estimate of the variance. The name of this statistical method was derived from the term “pull one’s self up by your bootstraps,” and the method allows statistical distributions to be generated from very little data.

Methods for analyzing molecular data are still undergoing development. The immense amount of DNA sequence data that is becoming available makes it difficult to use maximum-likelihood methods unless very powerful computers are used. Maximum-likelihood algorithms have been developed to build trees from pairwise distances, but they use only a summary of the data and information is thus lost. Parsimony methods are fast, but may be appropriate only for very slow rates of evolutionary change.

Another approach to analyzing evolutionary processes and phylogeny is Bayesian inference ([Shoemaker et al. 1999](#), [Huelsenbeck et al. 2001](#), [Ronquist and Deans 2010](#), [Fan and Kubatko 2011](#)). Bayesian inference uses the same models of evolution as other methods and can be used to infer phylogeny, evaluate uncertainty in phylogenies, detect selection, compare trees, evaluate divergence times, and test the molecular clock ([Huelsenbeck et al. 2001](#)). Bayesian inference of phylogeny is based on a quantity called the “posterior probability of a tree” and uses Bayes’s theorem:

$$\text{Pr}[\text{Tree} \mid \text{Data}] = \frac{\text{Pr}[\text{Data} \mid \text{Tree}] \times \text{Pr}[\text{Tree}]}{\text{Pr}[\text{Data}]}$$

In this theorem the vertical bar should be read as “given” and is used to “combine the prior probability of a phylogeny ( $\text{Pr}[\text{Tree}]$ ) with the likelihood ( $\text{Pr}[\text{Data} \mid \text{Tree}]$ ) to produce a posterior-probability distribution on trees ( $\text{Pr}[\text{Tree} \mid \text{Data}]$ ). The posterior probability of a tree is the probability that the tree is correct. Inferences about the history of the group are then based on the posterior probability of trees and the tree with the highest posterior probability might be chosen as the best estimate of phylogeny” ([Huelsenbeck et al. 2001](#)). The likelihood is calculated using one of a number of standard Markov models of character evolution. A Markov process is a mathematical model of infrequent changes of discrete states (nucleotides or amino acids) over time, in which future events occur by chance.

Phylogenetic analysis can be difficult because a large number of trees potentially could describe the relationships of a group of species. Evaluating which of

these trees is the best approximation of the “true” tree can be difficult when rates of DNA substitution are high; multiple substitutions at a site can make it difficult to resolve true relationships, producing the “wrong tree.” Methods that explicitly deal with multiple substitutions can overcome the statistical problems, but the most powerful methods (maximum likelihood) can be used only on relatively small data sets and many of the faster methods do not take advantage of all the data contained in the DNA sequences.

Bayesian inference makes it possible to analyze large data sets more easily. Instead of searching for the optimal tree, trees are sampled according to their posterior probabilities. Once such a sample is available, features that are common among these trees can be discerned and a consensus tree can be constructed. “This is roughly equivalent to performing a maximum likelihood analysis with bootstrap resampling, but much faster” ([Huelsenbeck et al. 2001](#)).

[Shoemaker et al. \(1999\)](#) noted a “common criticism of the Bayesian approach is that the choice of the prior distribution is too subjective.” Thus, researchers using the same data could reach different conclusions if they used different prior distributions. Furthermore, implementation of Bayesian methods can be “very complex.” Bayesian methods may be especially useful for analyzing complex evolutionary models (including horizontal gene transfer), and accommodating phylogenetic uncertainty.

Most analyses are conducted using at least two methods (maximum likelihood and Bayesian) and if the results obtained are similar the tree estimates are considered adequate.

#### 12.6.8 Artifacts

Inaccuracies in trees may occur for a variety of reasons ([Adoutte et al. 2000](#)). Alignments of corresponding sequences must be carried out carefully. If unambiguous alignments of sequences cannot be obtained, different relationships may be estimated. Poor alignments may result in a lack of strong statistical support for a particular tree. Another factor that affects phylogenies is the species chosen to represent each group. Use of different species within a group can result in different trees. Increasing the number of species analyzed may resolve this problem, but the increased number of species increases the computational time required to find the best tree to represent the relationships. For example, if five species are studied there are just 15 possible unrooted trees, but if 50 species are included, there are  $3 \times 10^{74}$  potential trees to analyze.

Other issues include determining whether sequences are orthologous, paralogous, or xenologous. An **ortholog** is a homologous sequence produced by

speciation; orthologs are derived from a common ancestor and tend to have similar functions. **Paralogs** are homologous sequences produced by gene duplication and represent genes that duplicated within an organism and then diverged, often having different functions. **Xenologs** result from the horizontal transfer of a gene between two organisms, with the function being variable, although often it is similar. Sometimes divergent lineages are morphologically similar (**homoplasy**). Homoplasy can occur due to a reversal to an ancestral trait in a lineage or to independent evolution (convergence) or to parallelism (similarity resulting from the same developmental genetic mechanisms) (Wake et al. 2011).

Inaccurate trees also may occur due to a phenomenon called **long-branch attraction** (Bergsten 2005). When long branches are in close proximity to short branches on evolutionary trees, maximum parsimony will recover the wrong tree because the long branches tend to group together or “attract each other.” Methods to avoid long-branch attraction include excluding faster evolving third-codon positions, excluding long-branch taxa, and sampling more taxa to break up long branches, as well as sampling more characters.

The statistical methods used for developing phylogenetic trees involve scoring for optimality in some way. Maximum parsimony (MP) and maximum likelihood (ML) methods attempt to find the tree or trees that optimize tree length or likelihood of the DNA sequences evolving along the trees. Bayesian methods consider many possible trees. A problem with all these trees when using large data sets, including many gene sequences that have been concatenated or combined or the use of entire genomes, is that many possible trees can be produced. Other concerns include the problems of discriminating between homology and convergence or parallel evolution (Wake et al. 2011). How to best handle enormous datasets remains to be resolved, although research is being conducted to do so (Sanderson et al. 2011, Xia and Yang 2011).

Another issue that can complicate phylogenetic analysis is missing data (Sanderson et al. 2011). Missing data can make the tree building more complex, producing “tree terraces.” The solution to such terraces is to focus on filling key data gaps to improve the effectiveness of tree-building programs.

### 12.6.9 Software Packages

Numerous software packages for phylogenetic analyses are available and supported (Swofford and Olsen 1990, Eernisse 1998). Software evolves rapidly and several software packages make it easy to conduct analyses using multiple methods. If the results are compared and there is concordance among the different

analysis methods, a particular tree is more likely to be correct (Caterino et al. 2000).

**MrBayes** is available if you wish to carry out a Bayesian analysis of your data. **Phylogeny Inference Package (PHYLIP)** is available from Joseph Felsenstein, Department of Genetics, University of Washington, Seattle, WA. It is a collection of  $\approx 30$  independent programs implementing maximum likelihood, parsimony, compatibility, distance, and invariant methods. Some of the programs provide bootstrap methods for estimating confidence limits. **Phylogenetic Analysis Using Parsimony (PAUP)** is available commercially (<http://paup.csit.fsu.edu/downl.html>). PAUP performs parsimony analysis under a variety of models, and bootstrapping routines are available. Also available is BIOSYS-2, which includes cluster analysis and distance Wagner routines for gene frequency data. **Hennig86** is a small, fast, and effective program for parsimony analysis under the Wagner and Fitch models. **MacClade**, written by W. P. Maddison and D. R. Maddison (2000), is useful in the analysis of character evolution and the testing of phylogenetic hypotheses under the same parsimony models described for PAUP plus additional ones.

Other software for phylogenetic analysis includes PALM, which assists in choosing an appropriate substitution model (Chen et al. 2009). PAML 4 is a package of programs for phylogenetic analyses using maximum likelihood methods (Yang 2007), and Molecular Evolutionary Genetics Analysis (MEGA3) facilitates analysis of DNA and protein sequence variation (Kumar et al. 2004). **Bayesian Evolutionary Analysis by Sampling Trees (BEAST)** (Drummond and Rambaut (2007) and **Bayesian Analysis of Trees with Internal Node Generation (BATWING)** (Wilson et al. 2003) are available, as well as others.

Geneious, a DNA-sequence analysis software package for biologists, allows DNA, RNA, and protein sequence alignment, assembly, and analysis. It integrates bioinformatic and molecular biology tools with a user-friendly interface, allows custom BLAST searches, sequence searches, and literature searches. Alignment tools include ClustalW, MUSCLE, and MAFFT, translation and profile alignment, and more. Geneious organizes data, stores data, and provides graphical outputs; it can provide visualizations of three-dimensional structures, annotations, SNPs, restriction sites, dot plots, and RNA secondary structures. It contains Mr. Bayes, PAUP, Neighbor Joining, UPDMA, bootstrapping and consensus tree programs, as well as primer-design programs ([www.geneious.com](http://www.geneious.com)). A simplified version is available free ([www.geneious.com/basic](http://www.geneious.com/basic), Kearse et al. 2012).

Although there are concerns about the difficulty of calculating large phylogenies (Degnan and Rosenberg 2009), Goloboff et al. (2009) analyzed 13 genes

and the morphology of **73,060** eukaryotic species by a parsimony method. Their results indicated a high degree of congruence with the major taxonomic groups, with only a small number of misplaced species. They concluded that tree-calculation algorithms can retrieve phylogenetic trees for very large data sets.

## 12.7 The Universal Tree of Life

As the following sections show, the evolution of life, particularly early life, remains controversial and difficult to study. Because arthropods evolved early and diverged rapidly, it has been especially difficult to resolve their early evolutionary history ([Trautwein et al. 2012](#)).

### 12.7.1 Two Domains

The traditional view was that life is divided into animals and plants, and the study of bacteria and fungi often took place in departments of botany. Later, it was realized that organisms could be divided into prokaryotes or eukaryotes (organisms without or with nuclei bounded by nuclear membranes and containing mitochondria). It became clear, however, that fungi are not plants (molecular data indicate they are actually more closely related to animals). All single-celled eukaryotes initially were placed into the phylum Protista, but this was found to be a heterogeneous group consisting of algae (formerly plants), protozoa (formerly animals), water molds (formerly fungi), and others ([Mayr 1998](#)). The grouping of living organisms was modified as more information became available.

### 12.7.2 Three Domains

A more recent view is that there are three primary domains of life (Archaeabacteria, Eubacteria, and Eukaryota) ([Woese et al. 1990](#), [Doolittle 1999](#), [Woese 2000](#)). The “Bacteria” were separated into two groups on the basis of variation in the small subunit rRNA: the traditional bacteria (called Eubacteria) and a previously unrecognized group (Archaeabacteria or Archaea) that contain members that inhabit extreme environments such as hot springs, sulfur springs, and deep vents ([Woese 1987](#)). [Woese \(1987\)](#) considered the Archaea might have been the first organisms on earth because they inhabit such extreme environments, although now we know Archaea are present in less-extreme habitats ([Brochier-Armanet et al. 2011](#)). However, separation of “bacteria” into two domains of a rank equal to that of the Eukaryota was justified because they were assumed to have evolved independently from a precursor group ([Figure 12.6](#)). The Archaeabacteria (sometimes called Archaea) were considered to be as different from the Eubacteria as the Eubacteria were from the Eukaryota on a molecular basis.

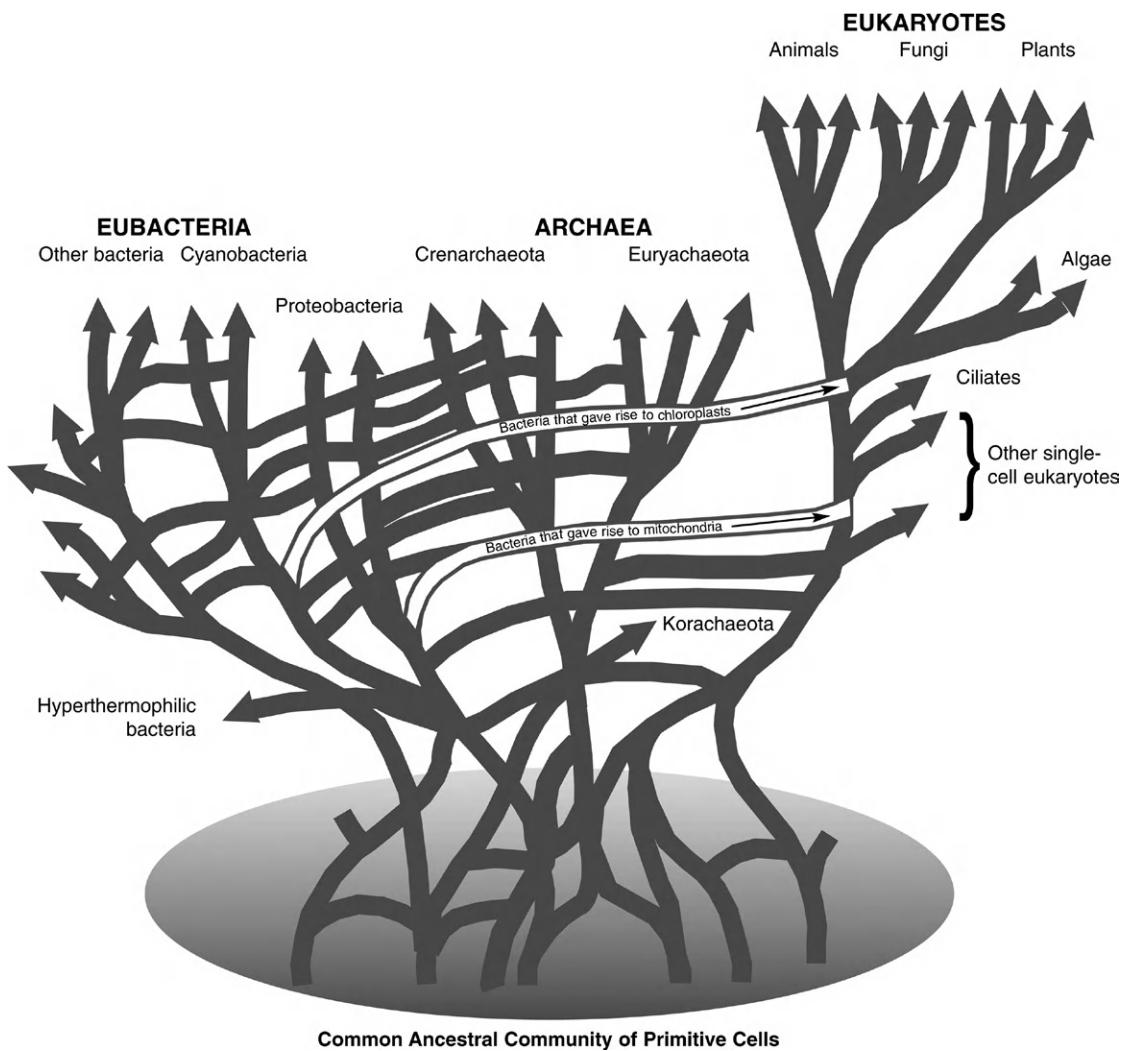
Ernst Mayr (1998) argued that the three-domain arrangement was unjustified because the number of groups known for Archaeabacteria and Eubacteria was far exceeded by the number of eukaryotic species, and the *phenotypic* diversity is orders of magnitude less for the Archaeabacteria and Eubacteria than the Eukaryota. However, separation of life into three domains has become less controversial (Cavalier-Smith 1998, Mayr 1998, Brochier-Armanet et al. 2011, Reynaud and Devos 2011). More than 100 archaeal genomes had been sequenced by 2011, and their analysis “provides an increasingly complex picture of archaeal phylogeny and evolution ...” and supports the three-domain arrangement (Brochier-Armanet et al. 2011, Fournier et al. 2011). However, the evolutionary relationship between the three domains of life remains difficult to resolve (Reynaud and Devos 2011). Several hypotheses have been proposed as to the origin of the Eukaryota.

### 12.7.3 Origin of Eukaryota

Molecular, geological, and paleontological evidence suggests that the Eukaryota originated a very long time ago (Margulis 1970, 1996; Katz 1998, 1999; Guy and Ettema 2011; Parfrey et al. 2011; Figure 12.6). Estimates of eukaryotic origins, however, vary (Parfrey et al. 2011), with the fossil record suggesting eukaryotes were present  $\approx$ 1800 million years ago and possible biomarkers as old as 2700 million years old have been found. Parfrey et al. (2011) provide molecular-clock estimates that suggest the origin of eukaryotes occurred 1866–1679 million years ago, which they claim is congruent with the fossil record.

Eukaryotic cells developed from an ancestor that contained an endosymbiont that originated from an  $\alpha$ -proteobacterium and that microorganism evolved to become a mitochondrion. This suggests that bacteria evolved before the Last Universal Common Ancestor (LUCA) of all eukaryotes and that the host cell that engulfed the bacterium could have been an archaeabacterium (Guy and Ettema 2011, Kelly et al. 2011). The rRNA genealogy and analyses of ancient gene duplications of protein-coding genes support the sister status of Archaeabacteria and Eukaryota (Katz 1998). However, this view had to be modified after complete genome sequences were obtained from a variety of organisms. It is believed that eukaryotic genomes are **chimeric**, derived from both archaeabacterial and eubacterial lineages (Golding and Gupta 1995, Rivera et al. 1998, Lang et al. 1999, Roger 1999, Kelly et al. 2011, Reynaud and Devos 2011).

The three domains share an RNA polymerase, ribosomes, membrane-protein systems, and a common genetic code, indicating that they are derived from a common ancestor (Reynaud and Devos 2011). However, each domain has unique features. Archaeabacteria and Eubacteria both have a circular chromosome



**Figure 12.6** Molecular phylogenies support the concept of a “web of life” in which gene exchange and horizontal gene transfer have had significant effects on the evolution of the Archaeabacteria, Eubacteria and Eukaryota. The evolution of life has involved multiple events of horizontal gene transfer between the domains, including incorporation of Eubacteria that gave rise to chloroplasts and mitochondrial, as well as transmittal of vertically transmitted archaeal genes. A “linear” view of the evolution of life had to be modified once genome analyses indicated the relationships between the domains were more complex. (Modified from Doolittle 2000.)

with genes arranged in operons. Archaeabacteria and the Eukaryota have gene-expression machinery with shared traits. Chimeric genomes could have developed only if lateral transmission of genes (or genomes) took place across species boundaries (Katz 1999). Doolittle (1998) proposes that gene swapping (leading

to chimerism) among early organisms may have occurred when primitive eukaryotes picked up genes from their food, and he suggested that "You are what you eat." According to this scenario, the ancestor of the Eukaryota was archaeal, but many eukaryotic nuclear genes today are of eubacterial origin because horizontal transfer occurred. Some eubacterial genes could have moved horizontally from mitochondria, but this is unlikely because the nuclear genes of mitochondrial origin are few and limited to proteins that are reimported into the mitochondria. Doolittle (1998) speculates that eubacterial genes with other functions could have moved into the eukaryotic genome when phagocytic unicellular eukaryotes fed on an  $\alpha$ -proteobacterium. DNA from these food bacteria would have moved repeatedly into the nuclear genome. Doolittle (1998) argued, "all genes that can be replaced by food-derived [eubacterial] genes will be, in the fullness of time. We should not think of such gene replacement as idiosyncratic or exceptional, but as the normal course. It is, instead, the persistence of some genes of archaeal ancestry that requires special explanation."

Forterre (2011) suggests another origin of the Eukaryota, in which an archaean was engulfed by a eubacterium followed by invasions of viruses, which provided many proteins. Poole and Neumann (2011) suggest an archaean engulfed a mitochondrial ancestor. Cotton and McInerney (2010) report that archaeabacterial-derived genes are more essential to yeast viability, more highly expressed and more connected and central to the yeast protein-interaction network, indicating that genes of archaeabacterial origin are more important to yeast metabolism than genes of eubacterial origin and speculated that archaeabacterial genes originated in the ancestral nuclear component of the eukaryotic genome. Clearly, the debate continues as to the origin of the Eukaryota due to the difficulty of analyzing such deep evolutionary events.

A consequence of lateral (horizontal) gene transfer is that phylogenetic analyses of different genes can result in conflicting phylogenies, causing confusion (Katz 1998, Bushman 2002). Evidence from sequences of 66 protein-coding genes from members of the three domains suggests that some eukaryotic genes are more similar to archaeal genes, whereas others appear to share ancestry with eubacterial genes (Brown and Doolittle 1997, Katz 1998). Archaeabacteria and the Eukaryota share genes involved in the genetic machinery of the cell, whereas Eubacteria and the Eukaryota share genes that regulate metabolic processes. These analyses "challenge the traditional view that vertical transmission of genetic material from one generation to the next is the predominant force in evolution" (Katz 1998).

The origin of the Eukaryota continues to be studied and debated (Kelly et al. 2011). Key characters involved in the emergence of eukaryotes include the

presence of a nucleus, microtubules, mitochondria, and a chimeric genome (Katz 1998). How all these parts were assembled remains controversial; some hypothesize there was a single endosymbiosis event and others suggest that there were two or more endosymbiosis events. It is possible that the original event that gave rise to mitochondria occurred in the ancestor of all extant eukaryotes, which could explain both the chimeric nuclear genome and the origin of mitochondria.

**Genome duplication** is thought to be an important component of the evolution of eukaryotic genomes (Ohno 1970, Wagner 1998, Sankoff 2001). Genome or gene duplication is thought to be a common method by which new gene functions can evolve, despite the fact that the vast majority of duplicate genes are expected to become pseudogenes through mutations. The loss of a duplicate gene is expected, because as long as one gene functions normally the other can accumulate deleterious mutations. However, many protein-coding genes belong to multigene families, which likely evolved by gene duplication (Friedman and Hughes 2001). Analysis of the complete genomes of *Drosophila melanogaster*, the nematode *Caenorhabditis elegans*, and the yeast *Saccharomyces cerevisiae* showed that duplication of genomic blocks has occurred, although the duplications did not all occur at the same time (Friedman and Hughes 2001). Some blocks could have been due to an ancient polyploidization event, whereas others are more recent and could have involved duplications of chromosome segments. However, as much as one-third of eukaryotic genomes may consist of unique genes with no apparent homologues in other organisms, suggesting that novel genes can evolve as well.

## 12.8 The Fossil Record of Arthropods

Insects have a relatively extensive fossil record beginning in the Cambrian, with 1263 families of fossil insects known (Labandeira and Sepkoski 1993, Budd and Telford 2009, Edgecombe 2010). Labandeira and Sepkoski (1993) found 472 references on fossils covering 1263 insect families, with all of the commonly recognized extant orders of insects represented as fossils. Edgecombe (2010) indicates that the earliest arthropod body fossils can be found in Stage 3 of the Cambrian and that arthropods have been the dominant component of animal species diversity for the past 520 million years, although Grimaldi (2010) suggests that the earliest divergences of hexapods were perhaps in the Late Silurian. Schaefer et al. (2010) argue that oribatid soil mites (Arthropoda: Chelicera: Arachnida: Acari) originated in the Precambrian (571 mya) and that the radiation of basal groups coincides with the gap in the terrestrial fossil record between the Cambrian explosion and the earliest fossil records on land. These authors suggest that the colonization of land started >150 million years earlier than the oldest fossils that are from terrestrial ecosystems.

Although only a few fossil insects (such as *Collembola*) are known from the lower Devonian, a massive radiation began sometime during the early Carboniferous, >325 million years ago, and the pterygotes radiated into stem groups of all major lineages, including ephemeroids, odonatoids, plecopteroids, orthopteroids, blattoids, hemipteroids, and endopterygotes. Insects continued to increase in diversity during the late Carboniferous and middle Permian (Table 12.4).

Insects are highly diverse and ancient arthropods (Budd and Telford 2009). The Crustacea are considered the sister group of the Tracheata (=Myriapoda+Hexapoda or Insecta). Relatively advanced Crustacea are found in the Cambrian (600 mya), so it is assumed that tracheates were present by this time as well (Kukalova-Peck 1991). Labandeira et al. (1988) showed that a bristletail (Archaeognatha) from the Early Devonian resembles modern archaeognathans. Arthropods apparently have been found on land since Devonian times (Table 12.4). Two *Collembola* species found in the lower Devonian (400 mya) resemble recent Isotomidae and Neanuridae, suggesting that terrestrial arthropods already had radiated in the Ordovician ( $\approx$ 500 mya).

Several extinct and extant orders of primitive insects have been found in a diverse late-Paleozoic fauna (Table 12.4). During the Carboniferous, which began 360 mya, a diverse array of extinct and extant insects were present, including the Diplura, Monura, Diaphanopterodea, Palaeodictyoptera, Megasecoptera, Permothemistida, Ephemeroptera (mayflies), Protodonata, Paraplectoptera, Plecoptera (stoneflies), Orthoptera (grasshoppers and crickets), Blattodea (cockroaches), Caloneurodea, Blattinopsodea, and Miomoptera.

During the Permian, which began  $\approx$ 285 mya, additional extinct and extant (underlined) insect groups are found in the fossil record, including Plecoptera (stone flies), Embioptera (web spinners), Protelytroptera, Glosselytrodea, Thysanoptera (thrips), Hemiptera (bugs and leafhoppers), Antliophora, Mecoptera (scorpion flies), Diptera (true flies), Asmphiesmenoptera, Neuroptera (lacewings, ant lions), Megaloptera (dobsonflies), and Coleoptera (beetles).

By the Triassic (245 mya), nearly all of the modern orders of insects are found in the fossil record, including Lepidoptera (butterflies and moths), Trichoptera (caddisflies), and Hymenoptera (bees and wasps). By the Jurassic (210 mya), many recent families are present. Tertiary insects (65 mya) are essentially modern and include genera nearly indistinguishable from living fauna.

There are >700,000 living species of living insects (estimates vary, with some asserting that up to two million insects are currently present). The orders Coleoptera (>300,000 named species), Lepidoptera (>120,000 species),

**Table 12.4: Geological Time Scale in Millions of Years and Types of Fossil Insects Found.**

Era	Period	Epoch	Began mya <sup>a</sup>	Extinct and extant fossil insect orders first found
Cenozoic	Quaternary	Recent		Protura, Zoraptera, and Phthiraptera first appeared in fossil record
		Pleistocene	1.6	
	Tertiary	Pliocene	5	
		Miocene	25	
		Oligocene	35	
		Eocene	60	Mantodea first appeared in fossil record
		Paleocene	65	
Mesozoic	Cretaceous		145	Isoptera first appeared in fossil record
	Jurassic		210	Dermaptera first appeared in fossil record
	Triassic		245	Odonata, <i>Titanoptera</i> , Grylloblattodea, Tricoptera, Lepidoptera, and Hymenoptera first appeared in the fossil record
Paleozoic	Permian		285	<i>Permethemistida</i> , Plecoptera, Embioptera, <i>Proteolyptoptera</i> , <i>Glosselytrodea</i> , Psocoptera, Thysanoptera, Hemiptera, <i>Antliophora</i> , Mecoptera, Diptera, <i>Amphiesmenoptera</i> , Neuroptera, Megaloptera, and Coleoptera first appeared in the fossil record
	Carboniferous		360	Pterygotes radiated into stem groups of all major lineages, with seven surviving to modern times (ephemeroids, odonatoids, plecopterooids, orthopteroids, blattoids, hemipteroids, and endopterygotes). <i>Diplura</i> , <i>Monura</i> , <i>Thysanura</i> , <i>Diaphanopterodea</i> , <i>Megasecoptera</i> , <i>Permethemistida</i> , <i>Protodonata</i> , <i>Paraplectoptera</i> , Orthoptera, Blattodea, <i>Caloneurodea</i> , <i>Blattinopsodea</i> , and <i>Miomoptera</i> were present
			400	<i>Collembola</i> ( <i>Rhyniella precursor</i> ) and <i>Archaeognatha</i> , and Opiliones
			440	Mites, opilionids, scorpions, pseudoscorpions, centipedes,
			500	pycnogonids, and spiders found in pre-Devonian strata
	Cambrian		600	Lobopodians ( <i>Panarthropoda?</i> ) Split into Chelicerata and Mandibulata (635–542 mya)

(Modified from Kukalova-Peck 1991, Giribet and Edgecombe 2012.)

<sup>a</sup>Dates given in this table and in the text may vary because different date estimates are used by different authors.

Hymenoptera (>120,000 species), and Diptera (>150,000 species) contain the most species. Insects are diverse, numerous, and ancient (Daly et al. 1998, Whitfield and Purcell 2013). Mora et al. (2011) recently suggested there are at least 8.7 million species of eukaryotes on earth, and several have suggested that

there are 3.7 million arthropod species (Strain 2011), so there will be systematics work required for a long time to come. An understanding of arthropod systematics and phylogeny requires the combined use of the fossil record, traditional morphological data, and molecular data and many relationships within the Arthropoda remain to be resolved (Giribet and Edgecombe 2012).

## 12.9 Molecular Analyses of Arthropod Phylogeny

### 12.9.1 *Evolution of the Ecdysozoa*

The superphylum Ecdysozoa is an evolutionary clade that includes the Insecta, Crustacea, Myriapoda, Chelicerata, Onychophora, Tardigrada, and five phyla of worms, including the Nematoda. The Ecdysozoa is the largest clade of species and occupies the greatest diversity of ecological niches, with an estimated total of >4.5 million living species (Telford et al. 2008). Telford et al. (2008) reviewed the molecular data supporting the Ecdysozoa as a natural, monophyletic group.

### 12.9.2 *Relationships among the Arthropoda*

Regier et al. (2010) evaluated 75 arthropod species using 62 single-copy nuclear protein-coding genes. The data were analyzed using likelihood, Bayesian, and parsimony methods. The 75 species included every major arthropod lineage plus five species of tardigrades and onychophorans. The results strongly supported the Pancrustacea (Hexapoda plus Crustacea) hypothesis and the morphology-based Mandibulata (Myriapoda plus Pancrustacea).

### 12.9.3 *The Phylogeny of the Holometabola*

Wiegmann et al. (2009) examined the evolutionary relationships of 11 orders of holometabolous insects. These include the Neuroptera, Raphidioptera, Megaloptera, Trichoptera, Lepidoptera, Diptera, Strepsiptera, Mecoptera, Siphonaptera, Coleoptera, and Hymenoptera, all of which are thought to have originated in the late Carboniferous (318–300 mya). Six nuclear protein-coding genes were analyzed using maximum likelihood and Bayesian methods. The results indicated that all orders are monophyletic, and the Hymenoptera are the basal-most lineage of the Holometabola, originating approximately when the Holometabola diverged. The Strepsiptera were found to be a sister group to the Coleoptera, resolving the controversial placement of this order as a sister group of the Neuroptera. Fleas (Siphonaptera) recently have been hypothesized to be members of the Mecoptera, but the data from Wiegmann et al. (2009) do not support collapsing the Siphonaptera into the Mecoptera. Based on molecular data, the holometabolous orders diverged rapidly, between 274 and 213 mya (although the fossil record does not always reflect these dates).

#### 12.9.4 Congruence Between Morphology- and Molecular-Based Trees

The fossil record is incomplete, so inferences made about lineages of organisms are based on what the scientist can observe and measure. Many scientists are concerned about using a single method, such as sequence data or morphological traits, to deduce evolutionary patterns. Often, inferences concerning phylogenetic relationships based on molecular data alone or morphological data alone may not reflect the historical relationship of the taxa from which the data were obtained, producing a “gene tree/species tree” problem (Telford and Copley 2011). For example, Powell (1991) pointed out that molecular studies on the *Drosophila pseudoobscura* group can lead to conclusions of monophyly, paraphyly and polyphyly, depending upon the data used to construct the trees. The different sets of data used are presumed to be accurate and thus neither tree is “wrong,” but reflects different aspects of the history of the same taxa when different data are considered. Morphological and molecular data can lead to different conclusions in some cases, but can produce congruent results in others.

Telford and Copley (2011) note, “To the outsider, phylogenetic debates can seem particularly fractious. If different investigators can reach strongly supported but entirely different conclusions by analyzing the same data, how does a consensus emerge and should it be trusted?” They conclude, “The falling cost of DNA sequencing means that, in the near future, phylogenetic questions will be approached with greatly expanded molecular datasets, both in terms of sampled taxa and quantity of data.” However, much remains to be learned about how to analyze that data. Errors in phylogeny construction can occur if we do not understand how different traits evolve and accurate models of evolution are not used.

#### 12.9.5 Genomes and Arthropod Phylogenies

Telford and Copley (2011) review what has been done to understand the evolution of animals using whole-genome sequences. They explore the following question: Is it possible that complete-genome sequences might provide an understanding of what happened  $\approx$ 530 million years ago in the Cambrian explosion? Genomes contain a vast amount of evolutionary information, but understanding how genomes have evolved requires the comparisons of genomes at different branches of the evolutionary tree. The use of genomic data is complicated by the fact that, during evolutionary time, multiple changes can occur and convergence and loss of traits can occur in genomes. Furthermore, the use of different models of molecular evolution can result in different trees. Problems include long-branch attraction, which is the introduction of systematic error in phylogeny reconstruction that results in unrelated long branches to cluster together, which could occur if evolution in specific taxa occurred through more rapid rates

of evolution or because no close relatives were sampled for particular species on the tree. As noted by [Telford and Copley \(2011\)](#), when the long-branch attraction issue was resolved using a better model, nematodes and arthropods were found to be closely related, supporting the Ecdysozoa hypothesis.

[Huerta-Cepas et al. \(2011\)](#) provide a repository for alignments based on whole-genome sequences, which allows additional analyses to be conducted using the alignments used by others. [Rokas and Abbott \(2009\)](#) discuss the experimental design and concepts associated with analysis of whole-genome sequences. [Rokas and Abbott \(2009\)](#) concluded that Next-Generation sequencing will soon “transform ecology and evolution by fundamentally changing the ranges and types of questions that can be addressed.”

## 12.10 Molecular Evolution and Speciation

The concept of species is inherent to the study of evolution and to understanding the evolution of life. Species concepts, definitions, and origins, however, remain controversial ([Hey 2001](#), [Margulis and Sagan 2002](#), [de Meeus et al. 2003](#), [Hey 2006](#), [Winker et al. 2007](#), [Coleman 2009](#), [Nosil et al. 2009](#), [Faria and Navarro 2010](#), [Marie Curie SPECIATION Network 2012](#)). [Winker et al. \(2007\)](#), “maintain that the inherent subjectivity within all species concepts is likely to ensure continued disagreement on where to place species limits.” [de Meeus et al. \(2003\)](#) note, “Species are entities that can be discriminated from one another, following criteria that seem the most appropriate at one place in time and space, and for one class of organisms.”

### 12.10.1 *Species Concepts*

One of the central questions of biology is how a continuous process of evolution can produce species ([Coyne 1992](#), [Rice and Hostert 1993](#), [O’Hara 1994](#), [Hollocher 1998](#), [Hey 2001](#), [Noor 2002](#), [de Meeus et al. 2003](#)). There are multiple views of “species” ([Table 12.5](#)). The **biological species concept** indicates that, “Species are groups of actually or potentially interbreeding natural populations, which are reproductively isolated from other such groups” ([Mayr 1970](#)). Reproductive isolation is achieved by **prezygotic isolating factors** (mating discrimination, different habitat preferences) and **postzygotic isolating factors** (hybrid inviability, sterility). Reproductive isolation, in concert with selection and genetic drift, creates and expands the morphological differences between species living in the same area.

Physical isolation (**allopatry**) leads inevitably to evolutionary change through natural selection or drift, and pre- or postmating reproductive isolation mechanisms evolve as a by-product of the genetic changes. Any resultant hybrid

**Table 12.5: Some Species Concepts or Definitions.<sup>a</sup>**

Biological	Members of populations that actually or potentially interbreed in nature, not according to similarity of appearance. Appearance may be helpful in identifying species, but it does not define them.
Chromosomal	Chromosomal rearrangements result in infertile hybrids, resulting in speciation.
Chrono species	Species at different stages in the same evolving lineage that existed at different points in time.
Ecological	Populations have distinct ecological roles but may still exchange genes.
Evolutionary	Symbionts or behavioral imprinting render populations reproductively isolated without ecological divergence.
Incipient	Populations that are in the process of becoming a species.
Phenetic	A set of organisms that are phenotypically similar and that look different from other sets of organisms.
Phylogenetic	The smallest set of organisms that share an ancestor and can be distinguished from other such sets. Under this definition, a ring species is a single species that encompasses a lot of phenotypic variation.
Recognition species	A species is a set of organisms that can recognize each other as potential mates.
Ring species	A species with a geographic distribution that forms a ring and overlaps at the ends.

Derived in part from the University of California Museum of Paleontology web page on Evolution 101 (2011).

<sup>a</sup>Additional definitions have been proposed.

Inviability could be due to the development of divergent developmental systems. Reproductive isolation may be increased if incompletely isolated populations become **sympatric** (live together in the same area) so that selection would fix the alleles that reduce interspecific mating. The process of increasing isolation is called reinforcement, but how often this process occurs is debated. Likewise, the extent of **sympatric speciation**, in which reproductive isolation occurs without geographic isolation, remains controversial, although the sympatric host races of the tephritid *Rhagoletis pomonella* represent a well-documented example (Feder et al. 1988, 1997). The biological species concept, in the view of some, overemphasizes reproductive isolation between populations.

The **evolutionary species concept** emphasizes the continuity of populations through time. An “evolutionary species is a lineage (ancestral-descendant sequence of populations) evolving separately from others with its own unitary evolutionary role and tendencies” (Simpson 1961). This definition focuses more on time than the biological species concept and has been criticized as being vague and not subject to observational test (O’Hara 1994).

The **phylogenetic species** is, “... the smallest diagnosable cluster of individual organisms within which there is a parental pattern of ancestry and descent” (Cracraft 1983). Critics of this concept note that the definition of diagnosability is vague; if examined carefully, characters can be found to diagnose virtually

any population, especially if molecular techniques are used. This would tend to greatly increase the number of species (O'Hara 1994).

Other species concepts include the cladistic species, cohesion species, composite species, ecological species, genetic species, genotypic cluster concept, Hennigian species, phenetic species, taxonomic species, and more (Hey 2001). Hey (2001) suggests that the problem is that "evolutionary biologists try... to find a way to have the taxa be the same as the evolutionary groups." Thus, depending on which species definition and assumptions are used, species are not necessarily equivalent. O'Hara (1994) noted many species are easy to delimit and,

*"in those cases where they are not, the difficulty that arises illustrates well the special historical character of the evolutionary process.... Because evolutionary history is something we are still in the midst of, it will not always be possible for us to determine which varieties--which distinctive populations in nature--are temporary and which are permanent, and so our counts of species across space and through time will always have some measure of ambiguity in them that we cannot escape."*

As a result, controversy will continue with regard to which genes or other traits are most suitable for analysis, which model of evolution is more appropriate, and which analysis method and software is best. Whether a phylogeny is based on molecular or phenetic data, the decision ultimately is affected by the scientist's biases.

### 12.10.2 How Many Genes are Involved in Speciation?

Nosil and Schluter (2011) defined a "speciation gene" as any gene contributing to the evolution of reproductive isolation, which means that the gene affects a component of reproductive isolation, that divergence at the locus occurred before the speciation event occurred, and that the amount of reproductive isolation caused by its divergence can be quantified. Other definitions include any gene that reduces hybrid fitness, and genes that result in hybrid inviability, sterility, or behavioral aberration.

The genetic basis of speciation is assumed to be due to changes in more than one gene (Harrison 1991, Coyne 1992, Hollocher 1998), with estimates ranging from 18 to 191 genes in the case of different species of *Drosophila* (Nosil and Schluter 2011). Changes in more than two genes have been considered the minimum required for reproductive isolation. Changes in segments of the genome, such as inversions or translocations, can result in reproductive isolation as well. A variety of characters can contribute to speciation, including hybrid sterility, hybrid inviability, interspecific mate discrimination, and interspecific divergence in secondary sexual characters (Hollocher 1998).

Speciation may occur rather rapidly under some circumstances. Higgle et al. (2000) showed that artificial selection can produce the kind of isolation that separates species in the wild, and that it can do so within nine generations under laboratory conditions. *Drosophila serrata* and *D. birchii* are sibling species in Australia that are very similar in morphology and can produce viable and fertile hybrid progeny in the laboratory. In the field, these two species can be found in the same area, but rarely interbreed. Where their geographic ranges overlap, the two differ in the mix of hydrocarbons found on their cuticle, which is important in mate choice. Populations of *D. serrata* found in regions of Australia where *D. birchii* occur have a different set of hydrocarbons. This suggests that selection to reduce mating between the species has occurred where the two populations overlap (sympatry). Under laboratory conditions, cuticular hydrocarbons of allopatric *D. serrata* populations collected from the field evolved within nine generations to resemble those of the sympatric populations when held with populations of *D. birchii*. This experiment in artificial sympatry indicates how rapidly mate-recognition systems can change. However, the experiment does not indicate whether selection on mate recognition was a component of the actual speciation event resulting in *D. serrata* and *D. birchii*.

*Drosophila melanogaster* populations from Zimbabwe and populations from other continents were shown to be reproductively isolated. In the presence of males of their own kind, females from Zimbabwe do not mate with males from elsewhere; reciprocal mating is reduced as well. The genes for this behavior apparently are found on autosomes II and III (Wu et al. 1995). The data suggest these populations are in the “early stages of speciation” and that it is “driven by sexual selection” (Wu et al. 1995). Because *D. melanogaster* is so well known genetically, analyzing speciation should be especially tractable (Buckley et al. 1997).

A speciation gene, called *Odysseus*, was cloned, sequenced and compared between two closely related *Drosophila* species (*simulans* and *mauritiana*) (Ting et al. 2000). Ting et al. (2000) were testing the hypothesis that genomes may contain ancient polymorphisms, or gene introgression could have occurred, so that molecular phylogenies may not reflect reproductive isolation accurately. Rather, speciation genes may be better indicators of phylogenetic history. *Odysseus* is the cause of hybrid male sterility in the *D. simulans* clade; Ting et al. (2000) compared *Odysseus* and microsatellite sequences from *simulans*, *mauritiana*, and *sechelia* (with *D. melanogaster* as an out group). The results indicated the genome (as indicated by 39 microsatellite loci) can “indeed be a mosaic of regions of different genealogies among closely related species, because of shared ancient polymorphism and/or introgressions.” The sequences of *Odysseus*, by contrast, provided a clear resolution of species because there were extensive amino-acid differences. The

authors raise the possibility that “diverging species that remain incompletely isolated reproductively (such as *D. simulans* and *D. mauritiana*) may be permeable to introgression over a large portion of their genomes.” Because only a small region near each “speciation locus” is impermeable, the exchange may continue for some time until reproductive isolation is complete.

Michel et al. (2010) evaluated the genetic diversity of *Rhagoletis pomonella*, a species known to have speciated sympatrically by moving from hawthorn to apple hosts in North America within the past 250 years. They found “widespread divergence throughout the *Rhagoletis* genome, with the majority of loci displaying host differences, latitudinal clines, associations with adult eclosion time, and within-generation responses to selection.”

Many questions about speciation remain unanswered (Howard and Berlocher 1999, Mallet 2006, Schlüter and Conte 2009, Via 2009, Johnson 2010, Cutter 2012). How important are conventional gene mutations compared to novel genetic elements, such as repeated sequences, symbiotic microorganisms such as *Wolbachia*, or transposable elements? How often are “speciation” genes altered in their coding sequence compared to changes in noncoding regulatory regions? How often is reproductive isolation based on polyploidy, or on chromosomal rearrangements of chromosomes? If transposable elements, polyploidy, or infectious microorganisms such as *Wolbachia* cause speciation, would they produce a rapid change without significant genetic change in the arthropod genome? Cloning and characterizing genes important in speciation may provide information on how reproductive isolation occurs at the molecular level.

### 12.10.3 Detecting Cryptic Species

One can argue that morphological methods are faster, easier, and cheaper than molecular methods for many taxonomic studies. However, molecular methods often provide the only method for detecting cryptic species. The ability to detect cryptic species may have ecological and economic importance.

RAPD-PCR may provide an inexpensive method for detecting cryptic species. For example, two populations of the encyrtid parasitoid *Ageniaspis citricola* were imported into the United States from Taiwan and Thailand as part of a classical biological control program directed against an invasive pest of citrus, *Phyllocnistis citrella* (Hoy et al. 2000). Slight differences in the biology and behavior of the two populations led us to evaluate them with RAPD-PCR and the results indicated the two populations were genetically distinct (Hoy et al. 2000). Subsequently, analysis of two highly conserved *actin* genes confirmed the distinctiveness of these populations. Analysis of ribosomal ITS2 sequences

also indicated the two populations are different (Alvarez and Hoy 2002). Interestingly, multiple clones of the ITS2 region were sequenced from individuals and the intraindividual sequence variation observed was sometimes greater than sequence variation between individuals. This variability in sequence and length of the ITS2 region in the *Ageniaspis* populations suggests that concerted evolution has not homogenized all copies of the rRNA genes within these individuals and populations. Yet, despite this variability, the ITS2 region was informative phylogenetically (Alvarez and Hoy 2002).

Another economically important example of cryptic species involves a mite, *Varroa*, that is a parasite of honey bees (Oldroyd 1999, Anderson and Trueman 2000). *Apis mellifera* originally was restricted to Europe, whereas *A. cerana* was found in Asia. These sibling species came into contact after 1905 when the trans-Siberian railroad was completed. *Varroa* moved onto *A. mellifera* sometime in the past century from *A. cerana* and created serious problems for beekeepers in many locations around the world. *Varroa* parasitizes *Apis cerana*, but the effects are relatively mild. It was believed that *Varroa* is a homogeneous species (called *jacobsoni*), but molecular studies (RAPD-PCR and mitochondrial DNA analyses) indicated there is considerable hidden genetic variability within *Varroa* populations (Kraus and Hunt 1995; de Guzman et al. 1997, 1998). In fact, molecular data suggested there were multiple introductions of *Varroa* into the Western Hemisphere (de Guzman et al. 1999). Furthermore, molecular data suggested that *A. cerana* is, in fact, attacked by at least two cryptic species of *Varroa* (*jacobsoni* and *underwoodi*). In the Western Hemisphere, two introduced "strains" of *Varroa* coexist. One is highly destructive to European bees (*A. mellifera*) and one relatively benign (Oldroyd 1999). Bee breeders have been genetically selecting strains of *A. mellifera* that are resistant to *Varroa*. Conducting such selections in an effective manner requires that the "virulent" *Varroa* strains (or cryptic species) be used during the selection process. Thus, identifying cryptic *Varroa* species may explain differences in the effects of "Varroa" on bee populations.

Havill et al. (2006) analyzed mitochondrial DNA from the hemlock wooly adelgid (*Adelges tsugae*) that was introduced into eastern North America and found differences indicating cryptic speciation, as well as the possible sources of the introduction.

## 12.11 Some Conclusions

Molecular tools have provided, and will continue to provide, taxonomic and phylogenetic answers to both basic and applied problems. The use of molecular methods remains relatively young and is still undergoing evolution itself. New

techniques and analysis methods continue to be developed. Time will resolve which are most useful for particular problems. One problem (How different does a group need to be to be considered a different species [or a family or an order or a class?]), however, remains a matter of judgment by the scientist studying the group. There is no uniform answer to that question.

What are the costs of describing the entire animal kingdom? [Carbayo and Marques \(2011\)](#) noted that ≈1.4 million of the estimated 6.8 million animal species have been described. They estimated that an average researcher describes 24.8 species on average during his/her career, with an average cost of US\$97,000/year (with careers lasting from 1 year to 46 years). This suggests that the total cost to describe the remaining animal species is US\$263 billion. Furthermore, at the current rate of describing 16,000 species per year with the current generation of trained taxonomists, it will take 360 years to fully catalogue animal diversity. [Carbayo and Marques \(2011\)](#) noted the “future need for greater numbers of taxonomists working on invertebrate groups (particularly insects) is enormous.” Clearly, there remains much work to be done to describe the world’s biota and, most especially, the arthropods, which have been estimated to consist of 3.7 million species ([Strain 2011](#)).

## Relevant Journals

- Annual Review of Ecology and Systematics*, Annual Reviews, Palo Alto, CA.  
*BMC Evolutionary Biology*, BioMed Central, <http://www.biomedcentral.com/bmcevolbiol/about>  
*Cladistics*, Wiley Publishers, Malden, MA.  
*Journal of Evolutionary Biology*, Wiley Publishers, Malden, MA.  
*Journal of Molecular Evolution*, Springer, New York.  
*Molecular Biology and Evolution*, Oxford University Press/Highware Press, Cary, NC.  
*Molecular Phylogenetics and Evolution*, Academic Press, San Diego, CA.  
*Trends in Ecology and Evolution*, Cell Press, Elsevier, London, UK.  
*Trends in Genetics*, Cell Press, Cambridge, MA.

## References Cited

- Adoutte, A., Balavoine, G., Lartillot, N., Lespinet, O., Prudhomme, B., and de Rosa, R., (2000). The new animal phylogeny: reliability and implications. *Proc. Natl. Acad. Sci. USA* **97**: 4453–4456.
- Altincicek, B., Kovacs, J.L., and Gerardo, N.M., (2012). Horizontally transferred fungal carotenoid genes in the two-spotted spider mite *Tetranychus urticae*. *Biol. Lett.* **8**: 253–257.
- Alvarez, J.M., and Hoy, M.A., (2002). Evaluation of the ribosomal ITS2 DNA sequences in separating closely related populations of the parasitoid *Ageniaspis* (Hymenoptera: Encyrtidae). *Ann. Entomol. Soc. Am.* **95**: 250–256.
- Anderson, D.L., and Trueman, J.W.H., (2000). *Varroa jacobsoni* (Acari: Varroidae) is more than one species. *Exp. Appl. Acarol.* **24**: 165–189.
- Arbogast, B.S., Edwards, S.V., Wakeley, J., Beerli, P., and Slowinski, J.B., (2002). Estimating divergence times from molecular data on phylogenetic and population genetic timescales. *Annu. Rev. Ecol. Syst.* **33**: 707–740.

- Aubert, J., and Solignac, M., (1990). Experimental evidence for mitochondrial DNA introgression between *Drosophila* species. *Evolution* **44**: 1272–1282.
- Avise, J.C., Arnold, J., Ball, R.M., Bermingham, E., Lamb, T., and Neigel, J.E., et al. (1987). Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annu. Rev. Ecol. Syst.* **18**: 489–522.
- Ayala, F.J., (1986). On the virtues and pitfalls of the molecular evolutionary clock. *Heredity* **57**: 226–235.
- Bachmann, L., Schibell, J.M., Raab, M., and Sperlich, D., (1993). Satellite DNA as a taxonomic marker. *Biochem. System. Ecol.* **21**: 3–11.
- Ball, S.L., and Armstrong, K.F., (2006). DNA barcodes for insect pest identification: a test case with tussock moths (Lepidoptera: Lymantriidae). *Can. J. For. Res.* **36**: 337–350.
- Ballard, J.W.O., and Kreitman, M., (1995). Is mitochondrial DNA a strictly neutral marker? *Trends Ecol. Evol.* **10**: 485–488.
- Bergsten, J., (2005). A review of long-branch attraction. *Cladistics* **21**: 163–193.
- Borisenko, A.V., Sones, J.E., and Hebert, P.D.N., (2009). The front-end logistics of DNA barcoding: Challenges and prospects. *Mol. Ecol. Res.* **9** (Suppl. 1): 27–34.
- Boore, J.L., and Brown, W.M., (1998). Big trees from little genomes: mitochondrial gene order as a phylogenetic tool. *Curr. Opin. Gen. Dev.* **8**: 668–674.
- Boore, J.L., Lavrov, D., and Brown, W.M., (1998). Gene translocation links insects and crustaceans. *Nature* **392**: 667–668.
- Brochier-Armanet, C., Forterre, P., and Gribaldo, S., (2011). Phylogeny and evolution of the Archaea: one hundred genomes later. *Curr. Opin. Microbiol.* **14**: 274–281.
- Brown, T.A., Ed., (1991). *Molecular Biology LabFax*, Bios Scientific Publishers, Oxford, UK.
- Brown, J.R., and Doolittle, W.F., (1997). Archaea and the prokaryote-to-eukaryote transition. *Microbiol. Mol. Biol. Rev.* **61**: 456–502.
- Buckley, S.H., Tregenza, T., and Butlin, R.K., (1997). Speciation and signal trait genetics. *Trends Ecol. Evol.* **12**: 299–301.
- Budd, G.E., and Telford, M.J., (2009). The origin and evolution of arthropods. *Nature* **457**: 812–817.
- Bushman, F., (2002). *Lateral DNA Transfer. Mechanisms and Consequences*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Cameron, S.L., Barker, S.C., and Whiting, M.F., (2006). Mitochondrial genomics and the new insect order Mantophasmatodea. *Mol. Phylogenet. Evol.* **38**: 274–279.
- Campbell, L.I., Rota-Stabelli, O., Edgecombe, G.D., Marchioro, T., Longhorn, S.J., and Telford, M.J., et al. (2011). MicroRNAs and phylogenomics resolve the relationships of Tardigrada and suggest that velvet worms are the sister group of Arthropoda. *Proc. Natl. Acad. Sci. USA* **108**: 15920–15924.
- Carbayo, F., and Marques, A.C., (2011). The costs of describing the entire animal kingdom. *Trends Ecol. Evol.* **26**: 154–156.
- Caterino, M.S., Cho, S., and Sperling, F.A.H., (2000). The current state of insect molecular systematics: a thriving Tower of Babel. *Annu. Rev. Entomol.* **45**: 1–54.
- Cavalier-Smith, T., (1998). A revised six-kingdom system of life. *Biol. Rev.* **73**: 203–266.
- Chen, S.H., Su, S.Y., Lo, C.Z., Chen, K.H., Huang, T.J., and Kuo, B.H., et al. (2009). PALM: A paralleled and integrated framework for phylogenetic inference with automatic likelihood model selectors. *PLoS ONE* **4** (12): e8116.
- Coleman, A.W., (2009). Is there a molecular key to the level of 'biological species' in eukaryotes? A DNA guide. *Mol. Phylogenet. Evol.* **50**: 197–203.
- Cotton, J.A., and McInerney, J.O., (2010). Eukaryotic genes of archaeabacterial origin are more important than the more numerous eubacterial genes, irrespective of function. *Proc. Natl. Acad. Sci. USA* **107**: 17252–17255.
- Coyne, J.A., (1992). Genetics and speciation. *Nature* **355**: 511–515.
- Cracraft, J., (1983). Species concepts and speciation analysis. *Curr. Ornithol.* **1**: 159–187.

- Cross, N.C.P., and Dover, G.A., (1987). Tsetse fly rDNA: an analysis of structure and sequence. *Nucleic Acids Res.* **15**: 15–30.
- Crozier, R.H., and Crozier, Y.C., (1993). The mitochondrial genome of the honeybee *Apis mellifera*: complete sequence and genome organization. *Genetics* **133**: 97–117.
- Cutter, A.D., (2012). The polymorphic prelude to Bateson-Dobzhansky-Muller incompatibilities. *Trends Ecol. Evol.* **27**: 209–218.
- Daly, H.V., Doyen, J.T., and Purcell III, A.H., (1998). *Introduction to Insect Biology and Diversity*, 2nd Ed., Oxford University Press, Oxford, UK.
- Degnan, J.H., and Rosenberg, N.A., (2009). Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends Ecol. Evol.* **24**: 3332–3340.
- de Guzman, L.I., Rinderer, T.E., and Stelzer, J.A., (1997). DNA evidence of the origin of *Varroa jacobsoni* Oudemans in the Americas. *Biochem. Genet.* **35**: 327–335.
- de Guzman, L.I., Rinderer, T.E., Stelzer, J.A., and Anderson, D., (1998). Congruence of RAPD and mitochondrial markers in assessing *Varroa jacobsoni* genotypes. *J. Apicult. Res.* **37**: 49–51.
- de Guzman, L.I., Rinderer, T.E., and Stelzer, J.A., (1999). Occurrence of two genotypes of *Varroa jacobsoni* Oud. in North America. *Apidologie* **30**: 31–36.
- de Meeus, T., Durand, P., and Renaud, F., (2003). Species concepts: what for? *Trends Parasitol.* **19**: 425–427.
- De Queiroz, K., and Gauthier, J., (1994). Toward a phylogenetic system of biological nomenclature. *Trends Ecol. Evol.* **9**: 27–31.
- Deroches, S.A.P., Le Ralec, A., Plantegenest, M., Chaubet, B., Cruaua, C., Cruaud, A., and Rasplus, J.Y., (2012). Identification of molecular markers for DNA barcoding in the Aphidiinae (Hym. Braconidae). *Mol. Ecol. Res.* **12**: 197–208.
- DeSalle, R., Egan, M.G., and Siddal, M., (2005). The unholy trinity: taxonomy, species delimitation and DNA barcoding. *Philos. Trans. R. Soc. B* **360**: 1905–1916.
- Dolgin, E., (2012). Rewriting evolution. Tiny molecules called microRNAs are tearing apart traditional ideas about the animal family tree. *Nature* **486**: 460–463.
- Doolittle, R.F., Ed., (1990). *Molecular Evolution: Computer Analysis of Protein and Nucleic Acid Sequences*, Methods in Enzymology, Volume **183**, Academic Press, New York.
- Doolittle, W.F., (1998). You are what you eat: a gene transfer ratchet could account for bacterial genes in eukaryotic nuclear genomes. *Trends Genet.* **14**: 307–311.
- Doolittle, W.F., (1999). Phylogenetic classification and the universal tree. *Science* **284**: 2124–2128.
- Dowling, T.E., Moritz, C., Palmer, J.D., and Rieseberg, L.H., (1996). Nucleic acids III: Analysis of fragments and restriction sites. In: Hillis, D.M., Moritz, C., and Mable, B.K., Eds., *Molecular Systematics*, 2nd Ed., Sinauer, Sunderland, MA, pp. 249–320.
- Drummond, A.J., and Rambaut, A., (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* **7**: 214.
- Edgecombe, G.D., (2010). Arthropod phylogeny: an overview from the perspectives of morphology, molecular data and the fossil record. *Arthrop. Struct. Dev.* **39**: 74–87.
- Eernisse, D.J., (1998). A brief guide to phylogenetic software. *Trends Genet.* **14**: 473–475.
- Eickbush, T.H., (2002). R2 and related site-specific non-long terminal repeat retrotransposons. In: Craig, N.L., Ed., *Mobile DNA II*, ASM Press, Washington, DC, pp. 813–835.
- Ereshefsky, M., (2001). *The Poverty of the Linnaean Hierarchy: A Philosophical Study of Biological Taxonomy*. Cambridge University Press, Cambridge, UK.
- Erwin, D.H., Laflamme, M., Tweedt, S.M., Sperling, E.A., Pisani, D., and Peterson, K.J., (2011). The Cambrian conundrum: early divergence and later ecological success in the early history of animals. *Science* **334**: 1091–1096.
- Fan, H.H., and Kubatko, L.S., (2011). Estimating species trees using approximate Bayesian computation. *Mol. Phylogenet. Evol.* **59**: 354–363.
- Faria, R., and Navarro, A., (2010). Chromosomal speciation revisited: rearranging theory with pieces of evidence. *Trends Ecol. Evol.* **25**: 660–669.

- Feder, J.L., Chilcote, C., and Bush, G.L., (1988). Genetic differentiation between sympatric host races of the apple maggot fly *Rhagoletis pomonella*. *Nature* **336**: 61–64.
- Feder, J.L., Roethle, J.B., Wlazlo, B., and Berlocher, S.H., (1997). Selective maintenance of allozyme differences among sympatric host races of the apple maggot fly. *Proc. Natl. Acad. Sci. USA* **94**: 11417–11421.
- Felsenstein, J., (1988). Phylogenies from molecular sequences: inference and reliability. *Annu. Rev. Genet.* **22**: 521–565.
- Felsenstein, J., (2004). *Inferring Phylogenies*. Sinauer, Sunderland, MA.
- Fitch, W.M., and Ayala, F.J., (1994). The superoxide dismutase molecular clock revisited. *Proc. Natl. Acad. Sci. USA* **91**: 6802–6807.
- Folmer, O., Black, M., Hoeh, W., Lutz, R., and Vrijenhoek, R., (1994). DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Mol. Marine Biol. Biotechnol.* **3**: 294–299.
- Fontaine, K.M., Cooley, J.R., and Simon, C., (2007). Evidence for paternal leakage in hybrid periodical cicadas (Hemiptera: *Magicicada* spp.). *PLoS ONE* **2** (9): e892.
- Forterre, P., (2011). A new fusion hypothesis for the origin of Eukarya: better than previous ones, but probably also wrong. *Res. Microbiol.* **162**: 77–91.
- Fortna, A., and Gardiner, K., (2001). Genomic sequence analysis tools: a user's guide. *Trends Genet.* **17**: 158–164.
- Fournier, G.P., Dick, A.A., Williams, D., and Gogarten, J.P., (2011). Evolution of the archaea: emerging views on origins and phylogeny. *Res. Microbiol.* **162**: 92–98.
- Fox, D.L., Fisher, D.C., and Leighton, L.R., (1999). Reconstructing phylogeny with and without temporal data. *Science* **284**: 1816–1819.
- Friedlander, T.P., Regier, J.C., and Mitter, C., (1992). Nuclear gene sequences for higher level phylogenetic analysis: 14 promising candidates. *Syst. Biol.* **41**: 83–490.
- Friedman, R., and Hughes, A.L., (2001). Gene duplication and the structure of eukaryotic genomes. *Genome Res.* **11**: 373–381.
- Gibson, G., and Muse, S.V., (2002). *A Primer of Genome Science*. Sinauer, Sunderland, MA.
- Giribet, G., and Edgecombe, G.D., (2012). Reevaluating the arthropod tree of life. *Annu. Rev. Entomol.* **57**: 167–186.
- Goloboff, P.A., Catalano, S.A., Mirande, J.M., Szumik, C.A., Arias, J.S., Kallersjo, M., and Farris, J.S., (2009). Phylogenetic analysis of 73,060 taxa corroborates major eukaryotic groups. *Cladistics* **25**: 211–230.
- Golding, G.B., and Gupta, R.S., (1995). Protein-based phylogenies support a chimeric origin for the eukaryotic genome. *Mol. Biol. Evol.* **12**: 1–6.
- Goldstein, D.B., and Harvey, P.H., (1999). Evolutionary inference from genomic data. *Bioessays* **21**: 148–156.
- Graur, D., and Li, W.H., (2000). *Fundamentals of Molecular Evolution*, 2nd Edition, Sinauer, Sunderland, MA.
- Gray, M.W., (1989). Origin and evolution of mitochondrial DNA. *Annu. Rev. Cell Biol.* **5**: 25–50.
- Gray, M.W., Burger, G., and Lang, B.F., (1999). Mitochondrial evolution. *Science* **283**: 1476–1481.
- Green, E.D., (2001). Strategies for the systematic sequencing of complex genomes. *Nat. Rev. Genet.* **2**: 573–583.
- Gregory, T.R., (2008). Understanding evolutionary trees. *Evol. Edu. Outreach* **1**: 121–137.
- Gribskov, M., and Devereux, J., Eds., (1991). *Sequence Analysis Primer*, Stockton Press, New York.
- Grimaldi, D.A., (2010). 400 million years on six legs: on the origin and early evolution of Hexapoda. *Arthropod. Struct. Develop.* **39**: 191–203.
- Guy, L., and Ettema, T.J.G., (2011). The archaeal 'TACK' superphylum and the origin of eukaryotes. *Trends Microbiol.* **19**: 580–586.
- Hall, B.G., (2011). *Phylogenetic Trees Made Easy. A How-To Manual*, Fourth Edition, Sinauer, Sunderland, MA.

- Hall, H.G., (1998). Restriction fragment length polymorphism (RFLP) analyses. In: Clapp, J.P., Ed., *Methods in Molecular Biology 50: Species Diagnostics Protocols: PCR and other Nucleic Acid Methods*, Humana, Totowa, NJ, pp. 333–366.
- Hames, B.D., and Rickwood, D., Eds., (1981). *Gel Electrophoresis of Proteins. A Practical Approach*, IRL Press, Oxford, UK.
- Harrison, R.G., (1991). Molecular changes at speciation. *Annu. Rev. Ecol. Syst.* **22**: 281–308.
- Havill, N.P., Montgomery, M.E., Yu, G., Shiyake, S., and Caccone, A., (2006). Mitochondrial DNA from hemlock wooly adelgid (Hemiptera: Adelgidae) suggests cryptic speciation and pinpoints the source of the introduction to eastern North America. *Ann. Entomol. Soc. Am.* **99**: 195–203.
- Hebert, P.D.N., Ratnasingham, S., and deWaard, J.R., (2003a). Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proc. R. Soc. Lond. B* **270**: 596–599.
- Hebert, P.D.M., Cywinski, A., Ball, S.L., and deWaard, J.R., (2003b). Biological identifications through DNA barcodes. *Proc. R. Soc. Lond. B* **270**: 313–321.
- Hebert, P.D.N., Penton, E.H., Burns, J.M., Janzen, D.H., and Hallwachs, W., (2004). Ten species in one: DNA barcoding reveals cryptic species in the Neotropical skipper butterfly *Astraptes fulgerator*. *Proc. Natl. Acad. Sci. USA* **101**: 14812–14817.
- Hennig, W., (1966). *Phylogenetic Systematics*. University of Illinois Press, Urbana, IL.
- Hey, J., (2001). The mind of the species problem. *Trends Ecol. Evol.* **16**: 326–329.
- Hey, J., (2006). On the failure of modern species concepts. *Trend Ecol. Evol.* **21**: 447–450.
- Higgle, M., Chenoweth, S., and Blows, M.W., (2000). Natural selection and the reinforcement of mate recognition. *Science* **290**: 519–520.
- Hillis, D.M., and Moritz, C., (1990). An overview of applications of molecular systematics. In: Hillis, D.M., and Moritz, C., Eds., *Molecular Systematics* Sinauer, Sunderland, MA, pp. 502–515.
- Hillis, D.M., Larson, A., Davis, S.K., and Zimmer, E.A., (1990). Nucleic Acids III: sequencing. In: Hillis, D.M., and Moritz, C., Eds., *Molecular Systematics* Sinauer, Sunderland, MA, pp. 318–370.
- Hillis, D.M., Morita, C., and Mable, B.K., Eds., (1996). *Molecular Systematics*, 2nd Ed., Sinauer, Sunderland, MA.
- Holbrook, F.R., Tabachnick, W.J., Schmidtmann, E.T., McKinnon, C.N., Bobian, R.J., and Grogan, W.C., (2000). Sympatry in the *Culicoides variipennis* complex (Diptera: Ceratopogonidae): A taxonomic reassessment. *J. Med. Entomol.* **37**: 65–76.
- Hollocher, H., (1998). Reproductive isolation in *Drosophila*: how close are we to untangling the genetics of speciation. *Curr. Opin. Genet. Dev.* **8**: 709–714.
- Hoss, M., and Paabo, S., (1993). DNA extraction from Pleistocene bones by a silica-based purification method. *Nucleic Acids Res.* **31**: 3913–3914.
- Howard, D.J., and Berlocher, S.T., Eds., (1999). *Endless Forms: Species and Speciation*, Oxford University Press, Oxford, UK.
- Howe, C.J., and Ward, E.S., Eds., (1989). *A Practical Approach*, IRL Press, New York.
- Hoy, M.A., Jeyaprakash, A., Morakote, R., Lo, P.K.C., and Nguyen, R., (2000). Genomic analyses of two populations of *Ageniaspis citricola* (Hymenoptera: Encyrtidae) suggest that a cryptic species may exist. *Biocontrol* **17**: 1–10.
- Huelsenbeck, J.P., and Rannala, B., (1997). Phylogenetic methods come of age: testing hypotheses in an evolutionary context. *Science* **276**: 227–232.
- Huelsenbeck, J.P., Ronquist, F., Nielsen, R., and Bollback, J.P., (2001). Bayesian inference of phylogeny and its impact on evolutionary biology. *Science* **294**: 2310–2314.
- Huerta-Cepas, J., Capella-Gutierrez, S., Pryszcz, L.P., Denisov, I., Kormes, D., and Marcet-Houben, M., et al. (2011). PhylomeDB v3.0: an expanding repository of genome-wide collections of trees, alignments and phylogeny-based orthology and paralogy predictions. *Nucleic Acids Res.* **39**: D556–560.
- International Barcode of Life (2010). International Barcode of Life. Available at: <<http://ibol.org/>>

- Janvier, P., (2010). microRNAs revive old views about jawless vertebrate divergence and evolution. *Proc. Natl. Acad. Sci. USA* **107**: 19137–19238.
- Jeyaprakash, A., and Hoy, M.A., (2010). A DNA extraction procedure that allows mite specimens to be slide mounted: phytoseiid species evaluated as a model. *Exp. Appl. Acarol.* **52**: 131–140.
- Jinbo, U., Kato, T., and Ito, M., (2011). Current progress in DNA barcoding and future implications for entomology. *Entomol. Sci.* **14**: 107–124.
- Johnson, N.A., (2010). Hybrid incompatibility genes: remnants of a genomic battlefield? *Trends Genet.* **26**: 317–325.
- Katz, L.A., (1998). Changing perspectives on the origin of eukaryotes. *Trends Ecol. Genet.* **13**: 493–497.
- Katz, L.A., (1999). The tangled web: gene genealogies and the origin of Eukaryotes. *Am. Nat. (Suppl.)* **154**: S137–S145.
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., and Sturrock, S., et al. (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinform. Appl. Note* **28**: 1647–1649.
- Kelly, S., Wickstead, B., and Gull, K., (2011). Archaeal phylogenomics provides evidence in support of a methanogenic origin of the Archaea and a thaumarchaeal origin for the eukaryotes. *Proc. R. Soc London B* **278**: 1009–1018.
- Kimura, M., (1968). Evolutionary rate at the molecular level. *Nature* **217**: 624–626.
- Kimura, M., (1983). *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge, UK.
- Kimura, M., (1987). Molecular evolutionary clock and the neutral theory. *J. Mol. Evol.* **26**: 24–33.
- Klass, K.D., Zompro, O., Kristensen, N.P., and Adis, J., (2002). Mantophasmatodea: a new insect order with extant members in the Afrotropics. *Science Express*: 069397. <[www.scienceexpress.org](http://www.scienceexpress.org)> 18 April 2002, 10.1126 Science1.
- Knowles, L.L., and Kubatko, L.S., Eds., (2010). *Estimating Species Trees: Practical and Theoretical Aspects*, Wiley-Blackwell, Hoboken, NJ.
- Kondo, R., Satta, Y., Matsuura, E.T., Ishiwa, H., Takahata, N., and Chigusa, S.I., (1990). Incomplete maternal transmission of mitochondrial DNA in *Drosophila*. *Genetics* **126**: 657–663.
- Kraus, B., and Hunt, G., (1995). Differentiation of *Varroa jacobsoni* Oud. populations by random amplification of polymorphic DNA (RAPD). *Apidologie* **26**: 283–290.
- Kreitman, M., (1996). The neutral theory is dead. Long live the neutral theory. *Bioessays* **18**: 678–683.
- Kukalova-Peck, J., (1991). Fossil history and the evolution of hexapod structures, 2nd Ed., *The Insects of Australia*, Vol. I Melbourne University Press, Melbourne, Australia, Pp. 141–179.
- Kumar, A., and Rai, K.S., (1990). Chromosomal localization and copy number of 18S + 28S ribosomal RNA genes in evolutionarily diverse mosquitoes (Diptera: Culicidae). *Hereditas* **113**: 277–289.
- Kumar, S., Tamura, K., and Nei, M., (2004). MEGA3: integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief. Bioinform.* **5**: 150–163.
- Labandeira, C.C., and Sepkoski Jr., J.J., (1993). Insect diversity in the fossil record. *Science* **261**: 310–315.
- Labandeira, C.C., Beall, B.S., and Hueber, F.M., (1988). Early insect diversification: evidence from a lower Devonian bristletail from Quebec. *Science* **242**: 913–916.
- Lanfear, R., Welch, J.J., and Bromham, L., (2010). Watching the clock: studying variation in rates of molecular evolution between species. *Trends Ecol. Evol.* **25**: 495–503.
- Lang, B.F., Gray, M.W., and Burger, G., (1999). Mitochondrial genome evolution and the origin of the Eukaryotes. *Annu. Rev. Genet.* **33**: 351–397.
- Lansman, R.A., Shade, R.O., Shapiro, J.F., and Avise, J.C., (1981). The use of restriction endonucleases to measure mitochondrial sequence relatedness in natural populations. III. Techniques and potential applications. *J. Mol. Evol.* **17**: 214–226.

- Lansman, R.A., Avise, J.C., and Huettel, M.D., (1983). Critical experimental test of the possibility of "paternal leakage" of mitochondrial DNA. *Proc. Natl. Acad. Sci. USA* **80**: 1969–1971.
- Larkin, M.A., et al. (2007). Clustal W and Clustal X version 2.0. *Bioinform. Appl. Note* **23**: 2947–2948.
- Lewontin, R.C., and Hubby, J., (1966). A molecular approach to the study of genic heterozygosity in natural populations. II. Amounts of variation and degree of heterozygosity in natural populations of *Drosophila pseudoobscura*. *Genetics* **54**: 595–609.
- Li, W.H., and Graur, D., (1991). *Fundamentals of Molecular Evolution*. Sinauer, Sunderland, MA.
- Liu, H., and Beckenbach, A.T., (1992). Evolution of the mitochondrial cytochrome oxidase II gene among 10 orders of insects. *Mol. Phylogen. Evol.* **1**: 41–52.
- Lumley, L.M., and Sperling, F.A.H., (2011). Utility of microsatellites and mitochondrial DNA for species delimitation in the spruce budworm (*Choristoneura fumiferana*) species complex (Lepidoptera: Tortricidae). *Mol. Phylogen. Evol.* **58**: 232–243.
- Lunt, D.H., Whipple, L.E., and Hyman, B.C., (1998). Mitochondrial DNA variable number tandem repeats (VNTRs): utility and problems in molecular ecology. *Mol. Ecol.* **7**: 1441–1455.
- Lunt, D.H., Hutchinson, W.F., and Carvalho, G.R., (1999). An efficient method for PCR-based isolation of microsatellite arrays (PIMA). *Mol. Ecol.* **8**: 891–894.
- Maddison, D.R., and Maddison, W.P., (2000). *MacClade: analysis of phylogeny and character evolution*. Sinauer Assoc., Sunderland, MA, Available at <<http://macclade.org/macclade/html/>>.
- Mallet, J., (2006). What does *Drosophila* genetics tell us about speciation? *Trends Ecol. Evol.* **21**: 386–393.
- Margulis, L., (1970). *Origin of Eukaryotic Cells*. Yale University Press, New Haven, CT.
- Margulis, L., (1996). Archaeal-eubacterial mergers in the origin of Eukarya: phylogenetic classification of life. *Proc. Natl. Acad. Sci. USA* **93**: 1071–1076.
- Margulis, L., and Sagan, D., (2002). *Acquiring Genomes: A Theory of the Origin of Species*. Basic Books, New York.
- Marie Curie SPECIATION Network, (2012). What do we need to know about speciation? *Trends Ecol. Evol.* **27**: 27–39.
- Matsuura, E.T., Fukuda, H., and Chigusa, S.I., (1991). Mitochondrial DNA heteroplasmy maintained in natural populations of *Drosophila simulans* in Reunion. *Gen. Res. Camb.* **57**: 123–126.
- Maxon, L.R., and Maxon, R.D., (1990). Proteins II: Immunological techniques. In: Hillis, D.M., and Moritz, C., Eds., *Molecular Systematics*, Sinauer, Sunderland, MA, pp. 127–155.
- May, B., (1992). Starch gel electrophoresis of allozymes. In: Hoelzel, A.R. Ed., *Molecular Genetic Analysis of Populations. A Practical Approach* IRL Press. Oxford, UK, pp. 1–27.
- Mayr, E., (1970). *Populations, Species, and Evolution*. Harvard University Press, Cambridge, MA.
- Mayr, E., (1998). Two empires or three? *Proc. Natl. Acad. Sci. USA* **95**: 9720–9723.
- Mayr, E., and Ashlock, P.D., (1991). *Principles of Systematic Zoology*, 2nd Ed., McGraw-Hill, New York.
- Michel, A.P., Sim, S., Powell, T.H.Q., Taylor, M.S., Nosil, P., and Feder, J.L., (2010). Widespread genomic divergence during sympatric speciation. *Proc. Natl. Acad. Sci. USA* **107**: 9724–9729.
- Miller, S.E., (2007). DNA barcoding and the renaissance of taxonomy. *Proc. Natl. Acad. Sci. USA* **104**: 4775–4776.
- Minelli, A., (2009). Phylo-evo-devo: combining phylogenetics with evolutionary developmental biology. *BMC Biol.* **7**: 36. doi: [10.1186/1741-7007-7-36](https://doi.org/10.1186/1741-7007-7-36).
- Mooers, A.O., and Holmes, E.C., (2000). The evolution of base composition and phylogenetic inference. *Trends Ecol. Evol.* **15**: 365–369.
- Mora, C., Tittensor, D.P., Adl, S., Simpson, A.G.B., and Worm, B., (2011). How many species are there on earth and in the ocean? *PLoS Biol.* **9** (8): e1001127. doi: [10.1371/journal.pbio.1001127](https://doi.org/10.1371/journal.pbio.1001127).
- Moran, N.A., and Jarvik, T., (2010). Lateral transfer of genes from fungi underlies carotenoid production in aphids. *Science* **328**: 624–627.

- Moran, N.A., Munson, M.A., Baumann, P., and Ishikawa, H., (1993). A molecular clock in endosymbiotic bacteria is calibrated using the insect hosts. *Proc. R. Soc. Lond. B* **253**: 167–171.
- Moritz, C., and Hillis, D.M., (1990). Molecular systematics: context and controversies. In: Hillis, D.M., and Moritz, C., Eds., *Molecular Systematics*, Sinauer, Sunderland, MA, pp. 1–12.
- Moritz, C., Dowling, T.E., and Brown, W.M., (1987). Evolution of animal mitochondrial DNA: relevance for population biology and systematics. *Annu. Rev. Evol. Syst.* **18**: 269–292.
- Mount, D.W., (2001). *Bioinformatics. Sequence and Genome Analysis*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Murphy, R.W., Sites, J.W., Buth, D.G., and Haufner, C., (1996). Proteins: Isozyme electrophoresis. In: Hillis, D.M., Moritz, C., and Mable, B.K., Eds., *Molecular Systematics*, 2nd Ed., Sinauer, Sunderland, MA, pp. 51–120.
- Nardi, F., Carapelli, A., Boore, J.L., Roderick, G.K., Dallai, R., and Frati, F., (2010). Domestication of olive fly through a multi-regional host shift to cultivated olives: comparative dating using complete mitochondrial genomes. *Mol. Phylogenet. Evol.* **57**: 678–686.
- Nei, M., (1996). Phylogenetic analysis in molecular evolutionary genetics. *Annu. Rev. Genet.* **30**: 371–403.
- Nixon, K.C., Carpenter, J.M., and Stevenson, D.W., (2003). The PhyloCode is fatally flawed, and the “Linnaean” System can easily be fixed. *Bot. Rev.* **69**: 111–120.
- Noor, M.A.F., (2002). Is the biological species concept showing its age? *Trends Ecol. Evol.* **17**: 153–154.
- Nosil, P., and Schlüter, D., (2011). The genes underlying the process of speciation. *Trends Ecol. Evol.* **26**: 160–167.
- Nosil, P., Harmon, L.J., and Seehausen, O., (2009). Ecological explanations for (incomplete) speciation. *Trends Ecol. Evol.* **24**: 145–156.
- O’Hara, R.J., (1994). Evolutionary history and the species problem. *Am. Zool.* **34**: 12–22.
- Ohno, S., (1970). *Evolution by Gene Duplication*. Springer, New York.
- Ohta, T., (1996). The current significance and standing of neutral and nearly neutral theories. *Bioessays* **18**: 673–677.
- Ohta, T., (2000a). Evolution of gene families. *Gene* **259**: 45–52.
- Ohta, T., (2000b). Mechanisms of molecular evolution. *Philos. Trans. R. Soc. Lond. B* **355**: 1623–1626.
- Oldroyd, B.P., (1999). Coevolution while you wait: *Varroa jacobsoni*, a new parasite of western honeybees. *Trends Ecol. Evol.* **14**: 312–315.
- Page, R.D.M., (2011). Space, time, form: viewing the Tree of Life. *Trends Ecol. Evol.* **27**: 113–120.
- Pagel, M., (1999). Inferring the historical patterns of biological evolution. *Nature* **401**: 877–884.
- Palumbi, S.R., (1996). Nucleic acids II: The polymerase chain reaction. In: Hillis, D.M., Moritz, C., and Mable, B.K., Eds., *Molecular Systematics*, 2nd Ed., Sinauer, Sunderland, MA, pp. 205–248.
- Parfrey, L.W., Lahr, D.J.G., Knoll, A.H., and Katz, L.A., (2011). Estimating the timing of early eukaryotic diversification with multigene molecular clocks. *Proc. Natl. Acad. Sci. USA* **108**: 13624–13629.
- Pashley, D.P., and Ke, L.D., (1992). Sequence evolution in mitochondrial ribosomal and ND-1 genes in Lepidoptera: implications for phylogenetic analyses. *Mol. Biol. Evol.* **9**: 1061–1075.
- Pasteur, N., Pasteur, G., Bonhomme, F., Catalan, J., and Britton-Davidian, J., (1988). *Practical Isozyme Genetics*. Ellis Horwood, Chichester, UK.
- Pennisi, E., (2001). Linnaeus’s last stand? *Science* **291**: 2304–2307.
- Peterson, K.J., Dietrich, M.R., and McPeek, M.A., (2009). MicroRNAs and metazoan macroevolution: insights into canalization, complexity, and the Cambrian explosion. *Bioessays* **31**: 736–747.
- Poole, A.M., and Neumann, N., (2011). Reconciling an archaeal origin of eukaryotes with engulfment: a biologically plausible update of the Eocyte hypothesis. *Res. Microbiol.* **162**: 71–76.
- Post, R.J., Flook, P.K., and Wilson, M.D., (1992). DNA analysis in relation to insect taxonomy, evolution and identification. In: Crampton, J.M., and Eggleston, P., Eds., *Insect Molecular Science*, Academic Press, London, UK, pp. 21–34.

- Powell., J.R., (1991). Monophyly/paraphyly/polyphyly and gene/species trees: an example from *Drosophila*. *Mol. Biol. Evol.* **8**: 892–896.
- Powell, J.R., and Caccone, A., (1990). The TEACL method of DNA-DNA hybridization: technical considerations. *J. Mol. Evol.* **30**: 267–272.
- Quental, T.B., and Marshall, C.R., (2010). Diversity dynamics: molecular phylogenies need the fossil record. *Trends Ecol. Evol.* **25**: 434–441.
- Reeck, G.R., deHaen, C., Teller, D.C., Doolittle, R.F., Fitch, W.M., and Dickerson, R.E., et al. (1987). “Homology” in proteins and nucleic acids: a terminology muddle and a way out of it. *Cell* **50**: 667.
- Regier, J.C.J.W., Shultz, A., Zwick, A., Hussey, B., Ball, R., Wetzer, J.W., Martin, and Cunningham, C.W., (2010). Arthropod relationships revealed by phylogenomic analysis of nuclear protein-coding sequences. *Nature* **463**: 1079–1083.
- Reynaud, E.G., and Devos, D.P., (2011). Transitional forms between the three domains of life and evolutionary implications. *Proc. R. Soc. Lond. B* **278**: 3321–3328.
- Rice, W.R., and Hostert, E.E., (1993). Laboratory experiments on speciation: what have we learned in 40 years? *Evolution* **47**: 1637–1653.
- Rivera, M.C., Jain, R., Moore, J.E., and Lake, J.A., (1998). Genomic evidence for two functionally distinct gene classes. *Proc. Natl. Acad. Sci. USA* **95**: 6239–6244.
- Rodriguez-Trelles, F., Tarrio, R., and Ayala, F.J., (2001). Erratic over dispersion of three molecular clocks: GPDH, SOD, and XDH. *Proc. Natl. Acad. Sci. USA* **98**: 11405–11410.
- Roger, A.J., (1999). Reconstructing early events in eukaryotic evolution. *Am. Nat.* **154**: S146–S163.
- Rokas, A., and Holland, P.W.H., (2000). Rare genomic changes as a tool for phylogenetics. *Trends Ecol. Evol.* **15**: 454–459.
- Rokas, A., and Abbott, P., (2009). Harnessing genomics for evolutionary insights. *Trends Ecol. Evol.* **24**: 192–200.
- Ronquist, F., and Deans, A.R., (2010). Bayesian phylogenetics and its influence on insect systematics. *Annu. Rev. Entomol.* **55**: 189–206.
- Rota-Stabelli, O., Campbell, L., Brinkmann, H., Edgecombe, G.D., Longhorn, S.J., and Peterson, K.J., et al. (2011). A congruent solution to arthropod phylogeny: phylogenomics, microRNAs and morphology support monophyletic Mandibulata. *Proc. R. Soc. Lond. B* **278**: 198–306.
- Rowley, D.L., Coddington, J.A., Gates, M.W., Norrbom, A.L., Ochoa, R.A., Vandenberg, N.J., and Greenstone, M.H., (2007). Vouchering DNA-barcoded specimens: test of a nondestructive extraction protocol for terrestrial arthropods. *Mol. Ecol. Notes* **7**: 915–924.
- Sanderson, M.J., McMahon, M.M., and Steel, M., (2011). Terraces in phylogenetic tree space. *Science* **333**: 448–450.
- Sankoff, D., (2001). Gene and genome duplication. *Curr. Opin. Genet. Dev.* **11**: 681–684.
- Satta, Y., and Takahata, N., (1990). Evolution of *Drosophila* mitochondrial DNA and the history of the *melanogaster* subgroup. *Proc. Natl. Acad. Sci. USA* **87**: 9558–9562.
- Saville, B.J., Kohli, Y., and Anderson, J.B., (1998). mtDNA recombination in a natural population. *Proc. Natl. Acad. Sci. USA* **95**: 1331–1335.
- Sempere, L.F., Cole, C.N., McPeek, M.A., and Peterson, K.J., (2006). The phylogenetic distribution of metazoan microRNAs: insights into evolutionary complexity and constraint. *J. Exp. Zool. (Mol. Dev. Evol.)* **306B**: 575–588.
- Sessions, S.K., (1996). Chromosomes: molecular cytogenetics. In: Hillis, D.M., Moritz, C., and Mable, B.K., Eds., *Molecular Systematics*, 2nd Ed., Sinauer, Sunderland, MA, pp. 121–168.
- Schaefer, I., Norton, R.A., Scheu, S., and Maraun, M., (2010). Arthropod colonization of land—linking molecules and fossils in oribatid mites (Acari, Oribatida). *Mol. Phylogenetic Evol.* **57**: 113–121.
- Schlüter, D., and Conte, G.L., (2009). Genetics and ecological speciation. *Proc. Natl. Acad. Sci. USA* **106**: 9955–9962.
- Shoemaker, J.S., Painter, I.S., and Weir, B.S., (1999). Bayesian statistics in genetics. A guide for the uninitiated. *Trends Genet.* **15**: 354–358.

- Simon, C., Franke, A., and Martin, A., (1991). The polymerase chain reaction: DNA extraction and amplification. In: Hewitt, G.M., Johnston, W., and Young, J.P.W., Eds., *Molecular Techniques in Taxonomy*, Springer-Verlag, Berlin, Germany, pp. 329–355.
- Simpson, G.G., (1961). *Principles of Animal Taxonomy*. Columbia Univ. Press, NY.
- Smith, M.A., Woodley, N.E., Janzen, D.H., Hallwachs, W., and Hebert, P.D.N., (2006). DNA barcodes reveal cryptic host-specificity within the presumed polyphagous members of a genus of parasitoid flies (Diptera: Tachinidae). *Proc. Natl. Acad. Sci. USA* **103**: 3657–3662.
- Song, H., Buhay, J.E., Whiting, M.F., and Crandall, K.A., (2008). Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified. *Proc. Natl. Acad. Sci. USA* **105**: 13486–13491.
- Steel, M., and Penny, D., (2000). Parsimony, likelihood and the role of models in molecular phylogenetics. *Mol. Biol. Evol.* **17**: 839–850.
- Stewart, C.B., (1993). The powers and pitfalls of parsimony. *Nature* **361**: 603–607.
- Strain, D., (2011). 8.7 million: A new estimate for all the complex species on earth. *Science* **333**: 1083.
- Swofford, D.L., (2002). *PAUP\*. Phylogenetic Analysis Using Parsimony (\*and Other Methods)*. Sinauer, Sunderland, MA.
- Swofford, D.L., and Olsen, G.J., (1990). Phylogeny reconstruction. In: Hillis, D.M., and Moritz, C., Eds., *Molecular Systematics*, Sinauer, Sunderland, MA, pp. 411–501.
- Swofford, D.L., Olsen, G.J., Waddell, P.J., and Hillis, D.M., (1996). Phylogenetic inference. In: Hillis, D.M., Moritz, C., and Mable, B.K., Eds., *Molecular Systematics*, 2nd Ed., Sinauer, Sunderland, MA, pp. 407–514.
- Tabachnick, W.J., (1992). Genetic differentiation among populations of *Culicoides variipennis* (Diptera: Ceratopogonidae), the North American vector of bluetongue virus. *Ann. Entomol. Soc. Am.* **85**: 140–147.
- Tabachnick, W.J., (1996). The genetics of *Culicoides variipennis* and the epidemiology of bluetongue disease in North America. *Annu. Rev. Entomol.* **45**: 20–40.
- Tamura, K., (1992). The rate and pattern of nucleotide substitution in *Drosophila* mitochondrial DNA. *Mol. Biol. Evol.* **9**: 814–825.
- Taylor, H.R., and Harris, W.E., (2012). An emergent science on the brink of irrelevance: a review of the past 8 years of DNA barcoding. *Mol. Ecol. Resour.* **12**: 377–388.
- Telford, M.J., and Copley, R.R., (2011). Improving animal phylogenies with genomic data. *Trends Genet.* **27**: 186–195.
- Telford, M.J., Bourlat, S.J., Economou, A., Papillon, D., and Rota-Stabelli, O., (2008). The evolution of the Ecdysozoa. *Philos. Trans. R. Soc. Lond. B* **363**: 1529–1537.
- Thompson, J.D., et al. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**: 4673–4680.
- Ting, C.T., Tsaur, S.C., and Wu, C.I., (2000). The phylogeny of closely related species as revealed by the genealogy of a speciation gene. *Odysseus*. *Proc. Natl. Acad. Sci. USA* **97**: 5313–5316.
- Trautwein, M.D., Wiegmann, B.M., Beutel, R., Kjer, K.M., and Yeates, D.K., (2012). Advances in insect phylogeny at the dawn of the postgenomic era. *Annu. Rev. Entomol.* **57**: 449–468.
- Van de Peer, Y., Neefs, J.M., De Rijk, P., and De Wachter, R., (1993). Reconstructing evolution from eukaryotic small-ribosomal-subunit RNA sequences: calibration of the molecular clock. *J. Mol. Evol.* **37**: 221–232.
- Via, S., (2009). Natural selection in action during speciation. *Proc. Natl. Acad. Sci. USA* **106**: 9939–9946.
- Wagner, A., (1998). The fate of duplicated genes: loss or new function? *Bioessays* **20**: 785–788.
- Wake, D.B., Wake, M.H., and Specht, C.D., (2011). Homoplasy: from detecting pattern to determining process and mechanism of evolution. *Science* **331**: 1032–1035.
- Warnock, R.C.M., Yang, Z., and Donoghue, P.C.J., (2012). Exploring uncertainty in the calibration of the molecular clock. *Biol. Lett.* **8**: 156–159.

- Weir, B.S., (1990). *Genetic Data Analysis*. Sinauer, Sunderland, MA.
- Werman, S.D., Springer, M.S., and Briteen, R.J., (1996). Nucleic acids I; DNA-DNA hybridization. In: Hillis, D.M., Moritz, C., and Mable, B.K., Eds., *Molecular Systematics*, 2nd Ed., Sinauer, Sunderland, MA, pp. 169–204.
- Whelan, S., Lio, P., and Goldman, N., (2001). Molecular phylogenetics: state-of-the art methods for looking into the past. *Trends Genet.* **17**: 262–272.
- White, M.J.D., (1973). *Animal Cytology and Evolution*, 3rd Ed., Cambridge University Press, Cambridge, UK.
- White, M.J.D., (1978). *Modes of Speciation*. W.H. Freeman, San Francisco, CA.
- White, P.S., and Densmore III, L.D., (1992). Mitochondrial DNA isolation. In: Hoelzel, A.R., Ed., *Molecular Genetic Analysis of Populations. A Practical Approach*, IRL Press, Oxford, UK, pp. 29–58.
- Whitfield, J.B., and Purcell III, A.H., (2013). *Daly and Doyen's Introduction to Insect Biology and Diversity*, Third Ed., Oxford Univ. Press, NY, [Although the book became available in 2012, it has a 2013 copyright date.]
- Wiegmann, B.M., Trautwein, M.D., Kim, J.W., Cassel, B.K., Bertone, M.A., and Winterton, S.L., et al. (2009). Single-copy nuclear genes resolve the phylogeny of the holometabolous insects. *BMC Biol.* **7**: 34.
- Wiley, E.O., and Lieberman, B.S., (2011). *Phylogenetics Theory and Practice of Phylogenetic Systematics*, 2nd Edition, Wiley-Blackwell, Hoboken, NJ.
- Wilson, A.C., Ochman, H., and Prager, E.M., (1987). Molecular time scale for evolution. *Trends Genet.* **3**: 241–247.
- Wilson, I., Weale, M., and Balding, D., (2003). Inferences from DNA data: Population histories, evolutionary processes and forensic match probabilities. *J. R. Statist. Soc. Ser. A* **166**: 155–188.
- Winker, K., Rocque, D.A., Braile, T.M., and Pruett, C.L., (2007). Vainly beating the air: species-concept debates need not impede progress in science or conservation. *Ornithol. Monogr.* **63**: 30–44.
- Woese, C.R., (1987). Bacterial evolution. *Microbiol. Rev.* **51**: 227–271.
- Woese, C.R., (2000). Interpreting the universal phylogenetic tree. *Proc. Natl. Acad. Sci. USA* **97**: 8392–8396.
- Woese, C.R., Kandler, O., and Wheelis, M.L., (1990). Towards a natural system of organisms: proposal for domains Archaea, Bacteria and Eucarya. *Proc. Natl. Acad. Sci. USA* **87**: 4576–4579.
- Wu, C.I., Hollocher, H., Begun, D.J., Aquadro, C.F., Xu, Y., and Wu, M.L., (1995). Sexual isolation in *Drosophila melanogaster*: a possible case of incipient speciation. *Proc. Natl. Acad. Sci. USA* **92**: 2519–2523.
- Xia, X., and Yang, Q., (2011). A distance-based least-square method for dating speciation events. *Mol. Biol. Evol.* **59**: 342–353.
- Yang, Z., (2007). PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**: 1586–1591.
- Yang, Z., and Bielawski, J.P., (2000). Statistical methods for detecting molecular adaptation. *Trends Ecol. Evol.* **15**: 496–503.
- Yassin, A., Markow, T.A., Parechania, A., O'Grady, P.M., and DeSalle, R., (2010). The genus *Drosophila* as a model for testing tree- and character-based methods of species identification using DNA barcoding. *Mol. Phylogenetic Evol.* **57**: 509–517.
- Zahariev, M., Dahl., V., Chen, W., and Levesque, C.A., (2009). Efficient algorithms for the discovery of DNA oligonucleotide barcodes from sequence databases. *Mol. Ecol. Res.* **9** (Suppl. 1): 58–64.
- Zuckerkandl, E., and Pauling, L., (1965). Evolutionary divergence and convergence in proteins. In: Bryson, V., and Vogel, H.J., Eds., *Evolving Genes and Proteins*, Academic Press, New York, pp. 97–166.

# Insect Population Ecology and Molecular Genetics

## Chapter Outline

13.1 Overview 592

13.2 Introduction 593

13.3 What is Molecular Ecology? 593

13.4 Collecting Arthropods in the Field for Analysis 594

13.5 Molecular Ecological Methods 595

    13.5.1 Allele-Specific PCR 596

    13.5.2 Allozymes (Protein Electrophoresis) 596

    13.5.3 Amplified Fragment Length Polymorphisms (AFLP-PCR) 599

    13.5.4 Double-Strand Conformation Polymorphism (DSCP) 599

    13.5.5 Heteroduplex Analysis (HDA) 600

    13.5.6 Microarrays 601

    13.5.7 Microsatellites 602

    13.5.8 RFLP Analysis 604

    13.5.9 PCR-RFLP 604

    13.5.10 RAPD-PCR 605

    13.5.11 Sequencing 606

        13.5.11.1 *DNA Barcoding* 607

    13.5.12 Single Nucleotide Polymorphism (SNP) Markers 608

13.6 Analysis of Molecular Data 608

    13.6.1 Allozymes 609

    13.6.2 Microsatellites 612

    13.6.3 RAPD-PCR 613

    13.6.4 RFLPs 613

    13.6.5 Sequencing 614

13.7 Case Studies in Molecular Ecology and Population Biology 614

    13.7.1 Genetic Variability in the Fall Armyworm: Incipient Species or Multiple Species? 614

    13.7.2 Analyses of Natural Enemies 617

        13.7.2.1 *Analyses of Adult Parasitoid Wasps to Reveal Larval Hosts (MAPL)* 617

        13.7.2.2 *Analyses of the Establishment of a Parasitoid Released in a Classical Biological Control Program* 618

        13.7.2.3 *Gut Analyses of Predatory Insects to Reveal Prey Species* 618

    13.7.3 Population Isolation and Introgression in Periodical Cicadas 620

13.7.4	Eradicating Medflies in California?	625
13.7.4.1	<i>The Controversy</i>	625
13.7.4.2	<i>Species-Specific Diagnostics</i>	628
13.7.4.3	<i>Geographic Origin of Medfly Populations</i>	630
13.7.4.4	<i>Is the Medfly Established in California?</i>	631
13.7.4.5	<i>The End?</i>	637
13.7.4.6	<i>Some Lessons Learned</i>	639
13.7.5	Plant Defenses to Insect Herbivory	639
13.7.6	Origins of Insect Populations	641
13.7.6.1	<i>Emerald Ash Borer</i>	641
13.7.6.2	<i>Domesticated Populations of Aedes aegypti</i>	641
13.7.6.3	<i>Bumblebees from Greenhouses Invade Wild Conspecific Populations</i>	642
13.7.6.4	<i>The Tobacco Aphid in the New World</i>	642
13.7.6.5	<i>Populations of Aedes taeniorhynchus in the Galapagos Islands: Rapid Evolution of an Invasive Species</i>	643
13.7.6.6	<i>Industrial Melanism in Peppered Moths</i>	643

**13.8 Applied Pest Management 643**

13.8.1	Monitoring Biotypes, Species, and Cryptic Species	644
13.8.2	Monitoring Vectors of Disease	645
13.8.3	Pesticide Resistances and Pest Management	645
13.8.4	Monitoring Pest-Population Biology	646
13.8.5	The “So What?” Test	646

**Relevant Journals 647****References Cited 647**

## 13.1 Overview

Molecular-genetic techniques provide powerful tools for the study of insect biology, ecology, and population genetics in both natural and laboratory populations. Analysis of proteins, nuclear or mitochondrial DNA and messenger RNA can be used to answer ecological questions at the individual, population, or ecosystem level. Analysis of proteins by electrophoresis has been useful with many insects, but some taxa with low levels of detectable genetic variation cannot be studied unless more-sensitive DNA markers are used. DNA analyses can identify biotypes, sibling species, determine paternity or whether hybridization or introgression occurs, and provide information on founder effects, population genetic structure, gene flow, inbreeding, genetic bottlenecks, dispersal, predation, and selection intensity. Large amounts of genetic variation can be sampled rapidly and inexpensively in large numbers of individuals by the RAPD-PCR method or by restriction-enzyme digests of DNA amplified by the PCR (RFLP-PCR) or by the AFLP-PCR method. Although technically more challenging and expensive, DNA fingerprinting using microsatellite DNA, heteroduplex analysis, or double-strand conformation polymorphism (DSCP) provide information on genetic variation at the individual and population level. The use of DNA microarray (DNA chip)

analysis has allowed researchers to evaluate the responses of plants to insect attack; microarray analysis could answer other important ecological questions.

Improvements have been made in the statistical methods and population genetics models used to analyze molecular data. The continued improvement in molecular techniques and analysis methods in population biology and ecology will provide opportunities to resolve both fundamental and applied questions in insect population ecology, population genetics, and pest management. The field has advanced to the point where we are no longer asking, "can we do that?" However, it also is relevant to ask the "so what?" question: do molecular methods solve real problems that are "not already solvable by simpler and cheaper methods" (Curtis 2002)?

## 13.2 Introduction

The fields of ecology and population genetics generally use synthetic, rather than reductionist, research approaches. It is noteworthy that a reductionist approach to biology (molecular genetics) is providing new tools for resolving problems in population genetics and ecology, as well as applied pest-management programs.

Insect ecology is the study of insects in their environment (Speight et al. 2008, Freeland et al. 2011, Price et al. 2011, Schowalter 2011). Insect ecologists are concerned with the biology of groups of insects and with the pattern of relationships between insects and their environment. Ecology thus is concerned with individuals, populations, and communities in ecosystems. Insect ecology is an important component of applied pest-management programs and conservation, as well as of more fundamental value in elucidating ecological principles. Ecology can be divided into autecology and synecology.

**Autecology** deals with the study of the individual or an individual species, its life history, behavior, and adaptations to the environment. **Synecology** investigates groups of organisms associated together as a unit (Odum 1971, Southwood 1978, Price 1997, Huffaker and Gutierrez 1999, Southwood and Henderson 2000). At present, most ecological research using molecular techniques is autecological.

## 13.3 What is Molecular Ecology?

**Molecular ecology** was defined by Weiss (1950) as, "the entire continuum of biological interactions between the molecular, cellular, organismal levels to the environment" (Lambert 1995). Molecular tools provide ecologists with diverse methods for evaluating these interactions and allow ecologists to answer questions that have been difficult to resolve using traditional methods.

Population geneticists study how genetic principles apply to entire populations (Hartl 1981, Real 1994, Hartl and Clark 1997, Rowntree et al. 2011). One of the most striking features of insect populations is their phenotypic diversity. An underlying assumption that population geneticists make is that this phenotypic diversity is matched by genetic diversity and they attempt to deal with the phenotypic and genotypic differences among individuals.

Population genetics and population ecology have been distinct disciplines, but they have become less distinct recently (Slatkin 1987, Kellenberger 1994, Mitton 1994, Real 1994, Hoffman et al. 1995, Loxdale and Luschai 1998, Sunnucks 2000, Black et al. 2001). The molecular analysis of genes and genetic systems may provide insights for both autecological and synecological studies because an insect's heredity determines its behavior and ability to survive in specific environments and communities. Changes in genes and gene frequencies in populations over evolutionary time are important for understanding both speciation and community structure (Hoffman et al. 1995, Behura 2006).

The application of molecular-genetic techniques to the study of insect population ecology will play an ever more significant role as insect ecologists explore the power, and limitations, of these new tools (Sunnucks 2000, Tittiger 2004, Behura 2006, Stinchcombe and Hoekstra 2008, Gilad et al. 2009, Nadeau and Jiggins 2010). Analyses could provide better understanding of biodiversity, biosafety issues, biotype ecology and evolution, colonization processes, conservation biology, diet analysis, dispersal, gene flow, geographical distribution, host-parasite interactions, hybridization or introgression, insect-plant interactions, kinship, paternity, pesticide resistance, population structure, species identity, sperm precedence, and vector biology (Mitton 1994, Hoffmann et al. 1995, Tabachnick and Black 1995, Roderick 1996, Cavalli-Sforza 1998, Howard and Berlocher 1998, Schwartz et al. 1998, Rieseberg 1998, Bohonak 1999, Davies et al. 1999b, Wang and Caballero 1999, Berticat et al. 2000, ffrench-Constant et al. 2000, Sunnucks 2000, Baldwin et al. 2001, Black et al. 2001, Hewitt 2001).

This chapter surveys methods that can be used and provides examples in which molecular methods have been used to resolve population biology and ecology problems, as well as solutions for pest-management programs.

### 13.4 Collecting Arthropods in the Field for Analysis

The ability to collect insects from the field may be regulated by state, federal, or international regulations (Dick et al. 1993). Permits are required to collect organisms on certain federal lands (wildlife refuges, national parks, and national monuments) and many state lands in the United States and elsewhere. Endangered

and threatened species in the United States are regulated by the Fish and Wildlife Service of the Department of Interior and permits must be obtained to collect, possess, or transport any species on the List of Endangered and Threatened Wildlife. Furthermore, the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) requires collecting permits and restricts collection or importation of any species on the international list. For example, sampling DNA from the hind wing of the endangered butterfly *Neonympha mitchellii mitchellii* was permitted only after Hamm et al. (2010) showed that they could remove small amounts of hind wing (2–3 mm<sup>2</sup>) from two other butterfly species without having a “significant impact on the behavior or survival.”

Most countries limit importation of live organisms that could be harmful to crops, livestock or humans. In the United States, for example, the Plant Protection and Quarantine (PPQ) branch of the U.S. Department of Agriculture, Animal and Plant Health Inspection Service (USDA–APHIS) must be contacted before importing and transporting live plants, arthropod natural enemies, arthropod plant pests, plant pathogens, and vectors of plant or animal disease. The Office of Biosafety of the Centers for Disease Control regulates the importation of agents of human disease or vectors that could harbor these agents. Shipping live insects by mail or by shipping services also is restricted; many will not accept live insects. Shipment of insects in alcohol or in dry ice also is limited due to concerns about airline safety.

### 13.5 Molecular Ecological Methods

Proteins, nuclear DNA, mitochondrial DNA, and DNA from symbionts such as *Wolbachia* can be analyzed to resolve ecological problems; each has advantages and disadvantages. The ease of study, amount of variation that can be detected, and cost differs with each target. Analysis of genetic variation at the individual, population or species level can be achieved with both single- and multiple-copy DNA sequences. Mitochondrial (mt) DNA analysis provides sufficient variation for studies of individuals, populations, or species, depending upon the region of the mitochondrion analyzed. Because the genomes of many arthropod species have been, or soon will be, completely sequenced, many more genes soon will become available for analysis in ecological studies.

Potential techniques include allozyme electrophoresis, restriction-fragment-length polymorphisms (RFLPs) of mtDNA or nuclear DNA, DNA fingerprinting by analysis of microsatellites, RAPD-PCR, heteroduplex analysis (HDA), amplified-fragment-length polymorphisms (AFLP-PCR), sequencing of both nuclear and mtDNA, allele-specific PCR by using standard or high-fidelity PCR protocols or quantitative PCR, and microarray analysis. These methods were described in

Chapters 6–8. Each varies in the time required, ease of execution, cost, and level of genetic variability that can be detected ([Table 13.1](#)).

DNA-based methods provide a way to examine DNA directly. Which DNA-based technique should be used in a particular project depends on the goals and the level of DNA variation in the species of interest ([Moritz et al. 1987](#), [Mitton 1994](#), [Roderick 1996](#), [Loxdale and Luschai 1998](#)). When attempting to resolve problems involving species not analyzed previously by that specific molecular method, it is difficult to predict whether a particular DNA sequence or technique will be informative. This is because different insect groups differ in the amount and type of genetic variability they contain. Thus, molecular ecology remains a developing field that is still refining the molecular tools used and the statistical methods utilized for data analysis.

### **13.5.1 Allele-Specific PCR**

Allele-specific PCR is rapid, easy, and appropriate for many population biology or ecology studies ([Erlich 1989](#), [Arnheim et al. 1990](#), see Chapter 8). Allele-specific PCR requires DNA sequence data so that primers can be developed and synthesized (see Chapter 8), although some “universal” primers for ribosomal and mt DNA can be used on many species ([Table 13.1](#)).

Ecologists have long wanted to know who is eating whom in the environment. Allele-specific PCR has been used to evaluate the diets of predators ([Agusti et al. 1999](#), [Zaidi et al. 1999](#), [Hoogendorn and Heimpel 2001](#)). Persistence time of prey DNA in the gut, size of target DNA sequences, as well as the abundance (single copy vs. multiple copy genes) of prey DNA in predator guts vary by species and by temperature (which affects the digestion rate). This application of allele-specific PCR must be developed for each predator-prey system, making validation of gut analyses time consuming (for more on this topic, see Section 13.7.2.3).

A variation of allele-specific PCR, Long or high-fidelity PCR, was described in Chapter 8. Because two DNA polymerases are used (one of which has the ability to proofread and correct errors in incorporation), high-fidelity PCR allows microbial DNA to be detected even when it is mixed with insect or plant DNA. High-fidelity PCR is especially useful when monitoring arthropod vectors of disease agents. High-fidelity PCR can detect very low titers of microbial symbionts such as *Wolbachia* or of plant pathogens within insects ([Jeyaprakash and Hoy 2000](#), [Hoy et al. 2001](#)).

### **13.5.2 Allozymes (Protein Electrophoresis)**

Allozymes have been used to analyze mating systems (random versus assortative mating), inbreeding, genetic drift, hybridization, effective population size,

**Table 13.1: Molecular Methods that Can be Used to Evaluate Insects in Ecological Studies.**

Technique	Level of discrimination	Advantages (+) and disadvantages (-)	Selected references
	<ul style="list-style-type: none"> <li>Type of data obtained (gene frequencies or base-pair changes)</li> </ul>		
AFLP-PCR	Detect differences in individuals and populations	<ul style="list-style-type: none"> <li>+ More reliable than RAPD-PCR, more user-friendly than RFLPs and microsatellites, samples large amounts of genome, sequence information not needed</li> <li>- May require relatively large amounts of clean DNA, requires multiple operations</li> </ul>	Mueller and Wolfenbarger 1999, see Chapter 8
Allele-specific PCR	<ul style="list-style-type: none"> <li>Detect single nucleotide differences in individuals and populations</li> <li>• Gene frequency and base pair changes</li> </ul>	<ul style="list-style-type: none"> <li>+ Small amounts of DNA required, relatively rapid and inexpensive, results can be visualized by staining with ethidium bromide/other labels</li> <li>- DNA sequence information needed for primers, or consensus primers, each reaction provides information for only one locus</li> </ul>	See Chapter 8, Arnheim et al. 1990, Innis et al. 1990, Erlich 1989
Long PCR variant	<ul style="list-style-type: none"> <li>• Same as above</li> </ul>	<ul style="list-style-type: none"> <li>+ Detects microbial DNA mixed with insect or plant DNA, is 6–8 orders of magnitude more sensitive than standard PCR</li> <li>- More expensive because it uses two polymerases, including one that corrects errors, care must be taken to avoid contamination because of increased sensitivity</li> </ul>	Jeyaprakash and Hoy 2000, Hoy et al. 2001
Double-strand conformation polymorphism (DSCP)	<ul style="list-style-type: none"> <li>• Detect changes in mobility of double-stranded DNA molecules on polyacrylamide gels</li> </ul>	<ul style="list-style-type: none"> <li>+ Can use PCR products for analysis, rapid and inexpensive, can identify new haplotypes for additional analysis</li> <li>- Some mutations don't produce changes in mobility, thus won't work with all PCR products, sequence differences can't be estimated, sequencing may be required</li> </ul>	Hagerman 1990, Saad et al. 1994, Atkinson and Adams 1997
Microsatellites	Detect differences in individuals and populations in tandemly repeated units in nuclear DNA	<ul style="list-style-type: none"> <li>+ High levels of variation present in most insects</li> <li>- Comigrating bands may not be identical alleles at a locus, relatively large amounts of clean, high-molecular weight DNA required, labeled probes required, relatively expensive and labor intensive, time and effort are required to identify repeated units</li> </ul>	Bruford et al. 1992, Kirby 1990, Zane et al. 2002
RAPD-PCR	Differences in single nucleotides in nuclear DNA	<ul style="list-style-type: none"> <li>+ Useful for species with limited genetic information, efficient, relatively inexpensive, requires little DNA</li> </ul>	Hadrus et al. 1992, Haymer 1994

(Continued)

**Table 13.1: (Continued)**

Technique	Level of discrimination	Advantages (+) and disadvantages (-)	Selected references
			<ul style="list-style-type: none"> <li>Type of data obtained (gene frequencies or base-pair changes)</li> </ul>
RFLPs	<ul style="list-style-type: none"> <li>Gene frequency data</li> </ul> <p>Differences in single nucleotides detected by sequences recognized by restriction endonucleases in nuclear and mtDNA</p> <ul style="list-style-type: none"> <li>Gene frequency and changes in base pairs</li> </ul>	<ul style="list-style-type: none"> <li>Sensitive to DNA concentration, no genetic information on PCR products, can yield nonreproducible products, markers are dominant and heterozygotes may be difficult to identify, Incorrect scoring can occur if two different fragments comigrate</li> <li>+ mt DNA most often analyzed, standard probes are available</li> <li>- Requires large amounts of DNA; Usually requires radiolabeled probes, Single locus or several loci only analyzed, Relatively expensive and technically demanding</li> </ul>	<p>Edwards and Hoy 1993, MacPherson et al. 1993, Landry et al. 1993</p> <p>Aquadro et al. 1992, Dowling et al. 1990, Tegelstrom 1992, White and Densmore 1992</p>
PCR-RFLPs	<p>Differences in single nucleotide sequences in nuclear and mtDNA recognized by the specific restriction enzyme used</p> <ul style="list-style-type: none"> <li>Gene frequency data</li> </ul>	<ul style="list-style-type: none"> <li>+ Requires only a small amount of DNA, can be visualized with EtBr, less expensive and more sensitive than standard RFLPs</li> <li>- Specific primers required, two separate procedures are required, making it more time consuming and expensive than allele-specific PCR</li> </ul>	Karl and Avise 1993
Protein electrophoresis	<p>Detect changes in charged amino acids</p> <ul style="list-style-type: none"> <li>Gene-frequency data</li> </ul>	<ul style="list-style-type: none"> <li>+ Inexpensive, many protocols available, produces codominant Mendelian characters of enzymes important in physiology</li> <li>- Less sensitive than DNA tests, number of tests that can be performed may be limited in small insects, proteins subject to environmental influences</li> </ul>	<p>May 1992, Pasteur et al. 1988, Murphy et al. 1990</p>
Sequencing PCR-amplified DNA	<p>Differences in single nucleotides of nuclear and mtDNA including coding and noncoding regions</p> <ul style="list-style-type: none"> <li>Gene frequency and changes in base pairs</li> </ul>	<ul style="list-style-type: none"> <li>+ Relatively small amounts of DNA needed, high resolution possible, some universal PCR primers available</li> <li>- Time-consuming and expensive, relatively small portion of genome can be sampled, technically more demanding than other methods, not often used when large numbers of insects must be screened due to cost</li> </ul>	Hoelzel and Green 1992

degree of genetic differentiation among populations, and migration. Extensive protein (allozyme) variation has been found in some natural insect populations. Exceptions often include haplo-diploid Hymenoptera and clonal organisms such as aphids (Crozier 1977, Lester and Selander 1979, Murphy et al. 1990). Even for other insects, however, allozyme studies may underestimate the amount of variation, detecting only  $\approx 30\%$  of the actual genetic diversity as determined by DNA-based methods.

Protein electrophoresis is a cost-effective technique and is relatively easy to perform. For example, allozyme variability was used to identify Japan as the likely origin of the mosquito *Aedes albopictus* that recently colonized the United States and Brazil (Kambhampati et al. 1991). Allozymes were used to demonstrate genetic differentiation between sympatrically occurring hawthorn and apple populations of *Rhagoletis pomonella* (Feder et al. 1988). Unfortunately, protein electrophoresis may not detect sufficient variation to answer some questions, and the number of analyses that can be performed with very small insects may be limited because of inadequate amounts of proteins (Table 13.1). Proteins are less stable than DNA and thus may be more sensitive to handling and storage problems.

### 13.5.3 Amplified Fragment Length Polymorphisms (AFLP-PCR)

AFLP is a PCR-based method to develop large numbers of markers for population analyses (Mueller and Wolfenbarger 1999, see Chapter 8). AFLP-PCR is a relatively inexpensive and reliable method of identifying many genetic markers without requiring DNA sequence information. As with RAPD-PCR and microsatellite analyses, AFLP-PCR screens multiple different regions of the genome. AFLP markers have been useful for assessing genetic differences among individuals, populations, and species (Mueller and Wolfenbarger 1999). AFLP markers may be more easily replicated than RAPD-PCR, although AFLP-PCR is more difficult to use and develop (Table 13.1).

### 13.5.4 Double-Strand Conformation Polymorphism (DSCP)

DSCP is used to detect single-base changes in DNA. DSCP is detected by differences in electrophoretic mobility in nondenaturing acrylamide gels of double-stranded DNA (Hagerman 1990, Saad et al. 1994, Atkinson and Adams 1997). Single-base changes in the DNA may alter the curvature of the helical axis of ds DNA, that could lead to changes in electrophoretic mobility. Not all mutations affect DNA curving, so some are undetected by this approach (Table 13.1).

DSCP markers could provide markers for species determination, kinship and paternity analysis, as well as other aspects of population genetics that require variation from a rapidly evolving region of DNA. For example, Atkinson and Adams (1997) analyzed the mitochondrial control region of the termite *Nasutitermes corniger* by DSCP using PCR products and discovered highly variable markers for population studies. Higher levels of polymorphism were found by DSCP than by RFLP analysis. The DSCP data suggested that some termite colonies contained unrelated queens, each of which produced workers.

### 13.5.5 Heteroduplex Analysis (HDA)

Heteroduplex analysis combines some of the advantages of allele-specific PCR and RFLP methods with the advantage of direct sequence analysis to detect new alleles (Tang and Unnasch 1997). This method originally was used to identify virus isolates or detect immune genotypes in humans. HDA detects changes in mobility on an electrophoresis gel of heteroduplex products formed between the strands of a probe DNA and a test DNA molecule. The number and type of mismatched bases within a given heteroduplex product determines the conformation and mobility of the DNA duplex during electrophoresis. HDA is sufficiently sensitive to detect single-base changes in fragments up to 500 bp.

HDA involves obtaining PCR products from the probe DNA and test DNA. The DNAs are mixed, and then denatured by heating; heteroduplex and homoduplex products are formed during the cooling of the sample. This results in four products: the probe and sample DNA because the probe and sample strands reanneal to themselves (homoduplexes); and two heteroduplex products, each comprising one strand from the probe and one from the test DNA. These homo- and heteroduplex DNAs are separated by polyacrylamide gel electrophoresis and the separated products are detected by ethidium bromide staining. The relative amount of retardation of the heteroduplex products compared to the homoduplexes reflects the number and type of mismatched nucleotides between the probe and test DNA. This allows new alleles to be detected, which can be further analyzed by sequencing to determine their relationship to previously identified sequences. HDA is sufficiently rapid that multiple individuals can be screened, allowing one to determine the allele frequency in the population.

Tang and Unnasch (1997) suggested HDA be applied to medically important vectors of disease. They argued that HDA is simple, rapid, inexpensive, and capable of detecting small differences among DNA sequences. Disadvantages to HDA include the difficulty of measuring differences in the mobility of the sequences on the gels; such differences are only a rough estimate of the relative genetic

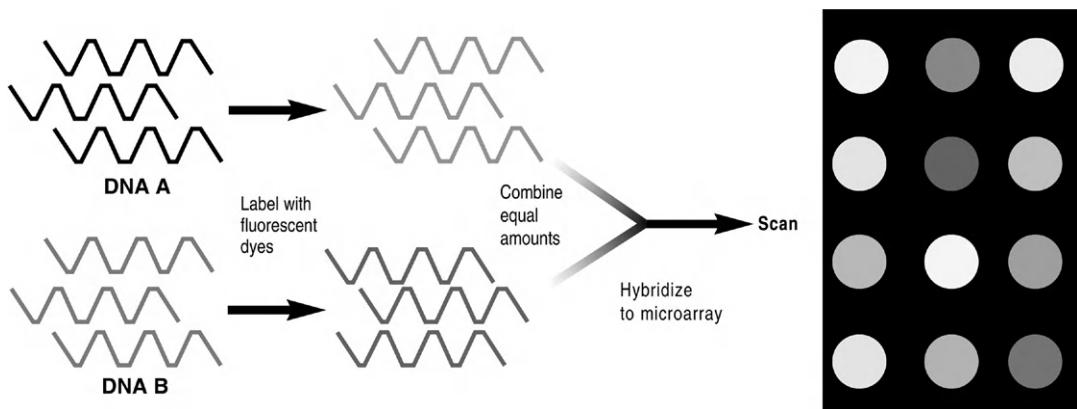
distance between two sequences. Furthermore, some changes may yield a bend in the DNA molecule, resulting in a disproportionate change in gel mobility. Further analysis of this method will resolve its value for molecular ecology.

### 13.5.6 Microarrays

DNA microarrays (also known as gene chips, genome chips, and gene arrays) are a technology that began to be used in the mid-1990s to analyze genome-wide patterns of gene expression within and among species (Gibson 2002, see Chapter 6). Initial applications of microarrays involved gene discovery, disease diagnosis, drug discovery, and toxicological analyses comparing tissues or cells. However, this technology provides a tool that allows entomologists to move beyond the one-gene-one experiment paradigm.

Microarrays allow the researcher to measure the relative quantities of specific mRNAs in two or more samples for hundreds or thousands of genes simultaneously. When a sample containing bits of fluorescently labeled cDNA is added to the chip containing spotted samples of DNA or cDNA, the labeled samples (e.g., one may be labeled green and one red) will anneal to the DNA on the chip that has complementary sequences. The chip contains known DNA sequences in a specific array. After complementary base pairing, the chip can be scanned and the colors will tell you which genes have hybridized with the sample DNA by complementary base pairing. Intermediate colors indicate those genes were active in both samples. When one sample contains cDNA but the other sample does not, the sample will fluoresce either red or green. If the same genes are turned on in both insects, both red and green dyes will be present on the microarray “dot” and appear as yellow (Figure 13.1). Thus, DNA microarrays can provide information on the transcriptional changes (transcriptomics) that occur in specific cells, tissues, or whole organisms under specific conditions.

DNA microarrays provide a way to analyze ecological interactions between plants and arthropod pests. For example, microarrays were used to assess the multiple transcriptional changes that occur after plants are attacked by insects (see case study below on insect–plant interactions). DNA microarrays should be useful for other ecological analyses. Scientists could use microarrays to evaluate the major events that occur during parasitism of insects by pathogens or nematodes or to evaluate the role of symbionts or what genes are involved in diapause induction and termination. DNA microarrays might be applied to DNA profiling to characterize genetic differences among biotypes or closely related species (Gibson 2002). Limitations to the application of microarrays to ecological problems currently include their perceived expense and lack of availability for nonmodel insect species (Gibson 2002). As with DNA sequencing, however, the



**Figure 13.1** In some microarray experiments, cDNAs copied from the messenger RNAs of two different insects (A and B) are each labeled with a different colored fluorescent dye. The labeled cDNAs are then combined in equal proportions and used as probes to hybridize with known DNA or RNA sequences on the microarray. After hybridization, the microarray is scanned, and the colors and their relative intensities are recorded. If, for example, cDNA from insect A is labeled with red and cDNA from insect B is labeled with green, when equal amounts of A and B cDNAs are applied to the array the resulting color patterns indicate which genes are active in insect A and B. A red dot will indicate high levels of expression of the corresponding gene on the microarray in insect A (and no expression in insect B). A green dot in that area indicates high levels of expression of that gene in insect B (and no expression in A). If the same gene is equally active in both insects, then the color will be a blend of red plus green, or yellow. [The dots on the right side of the illustration are shown in shades of gray here, but normally would be in color.]

cost of microarrays is declining rapidly as more core facilities are devoted to this technology and as more arthropod genomes are sequenced.

Analysis of microarray data must deal with the challenge of comprehending and interpreting the resulting massive amounts of data. As with any new technology, quality control and adequate statistical methods must be developed and used (Kerr and Churchill 2001a,b, Quackenbush 2001).

### 13.5.7 Microsatellites

DNA fingerprinting may involve the use of “microsatellite” or “minisatellite” sequences, which consist of arrays of up to several hundred simple sequence repeats (SSRs). In arthropods, these repeats most often consist of repeats of dinucleotides (AC, AT, AG), trinucleotides (AGC, AAC, AAT) or tetranucleotides (ACAT, AAAT, AAAC) (Toth et al. 2000). Microsatellites typically are scattered throughout the chromosomes of most organisms in both protein-coding and noncoding regions (Bruford et al. 1992, Estoup et al. 1993, Ashley and Dow

1994, Toth et al. 2000). Microsatellites mutate at a high rate and are thought to play a significant role in genome evolution by creating and maintaining quantitative genetic variation (Kashi et al. 1997). For example, when microsatellites are found in promoter regions, they may influence transcriptional activity. The length of microsatellites could also affect protein–protein interactions during transcription.

Two models were proposed to account for the high mutation rate in microsatellites. The first model, DNA polymerase slippage, assumes that replicating DNA strands transiently disassociate and then reassociate in a misaligned form, which will result in length and sequence variations. The second model involves unequal recombination to produce mutations in microsatellites. Understanding the evolution of microsatellites is considered important in using them for ecological analyses.

Microsatellite analysis results in a pattern of DNA bands on gels that resemble “bar codes” used to identify items in stores. DNA fingerprinting can be used to evaluate DNA variability at the individual and population level and was first conducted on humans and other vertebrates (Jeffreys 1987). The banding patterns produced often are specific to a particular individual (except for monozygotic twins), are inherited in a Mendelian manner, and are generally stable within an individual’s germ-line and somatic tissues. Polymorphisms are visualized by hybridizing a labeled probe to genomic DNA that has been cut with a restriction enzyme and separated into bands on a gel by electrophoresis.

Microsatellite markers can identify individual insects, or their progeny, evaluate kinship, resolve whether a mating has been successful, and reveal differentiation among closely related populations in the field (Burke 1989, Wang et al. 1999). Microsatellites could be useful in monitoring establishment and dispersal of specific biotypes, including those with low levels of protein variation such as parthenogenetic aphids or hymenopteran parasitoids (Table 13.1). Analysis of microsatellites has become a popular method for identifying high levels of genetic variability.

Unfortunately, microsatellite sequences usually differ in different species, even closely related species. This means that microsatellite sequence data usually must be obtained for each species under study, making microsatellites relatively time consuming and expensive to develop. Furthermore, different taxonomic groups may exhibit differences in the ease with which microsatellites can be isolated (Neve and Meglecz 2000). For example there were only five microsatellite studies published on Lepidoptera between 1997 and 1999, but 47 were published on Hymenoptera (Neve and Meglecz 2000); it is not clear whether equal

efforts were expended or whether the disparity represents true differences in ease of isolation.

A variety of methods have been developed for isolating microsatellites (Tables 13.1, 13.3). [Zane et al. \(2002\)](#) reviewed the methods and provide a “fast and easy protocol that is a combination of different published methods.” Their goal was to “provide a well established universal protocol” but they recognized that “completely new approaches [may] become available due to a better knowledge of microsatellite evolution combined with new technical advances.” [Techen et al. \(2010\)](#) improved on methods to isolate microsatellite sequences, and an outline of their methods is presented in [Table 13.3](#). At present, a careful evaluation of the experimental strategy has to be carried out for each experiment ([Zane et al. 2002](#)). [Goodman \(1997\)](#), [Cornuet et al. \(1999\)](#), [Luikart and England \(1999\)](#), [Balloux and Lugon-Moulin \(2002\)](#), and [Softlinks \(2007\)](#) reviewed statistical issues associated with the analysis of microsatellite markers.

### **13.5.8 RFLP Analysis**

RFLP analysis can be used to analyze variation in both mtDNA and nuclear DNA ([Table 13.1](#)). Depending on which restriction enzymes are used and target sequences analyzed, extensive variation can be discerned. However, RFLP analyses require relatively large amounts of very clean DNA, which may not be obtainable from single individuals of small insects. The DNA must be digested, electrophoresed, blotted, and probed to detect the variation. Probes must be developed, either as consensus sequences from other species, or after cloning and sequencing of species-specific DNA. Working with large numbers of individual insects is relatively time-consuming and expensive ([Table 13.1](#)).

### **13.5.9 PCR-RFLP**

A modification of RFLP analysis, called **PCR-RFLP**, eliminates many of the disadvantages of traditional RFLP analysis ([Karl and Avise 1993](#), [Table 13.1](#)). If no probe is available, a genomic DNA library can be constructed and clones isolated and sequenced. Alternatively, degenerate primers can be designed and the PCR products cloned and sequenced. Once species-specific sequences are available, allele-specific PCR primers can be designed. Subsequently, nuclear DNA can be amplified by the PCR using these primers and the PCR product can be digested with appropriate restriction enzymes. The cut DNA is visualized after electrophoresis by staining with ethidium bromide.

The advantage of PCR-RFLP is that DNA extracted from a single individual is sufficient (after PCR amplification) to provide bands that can be visualized.

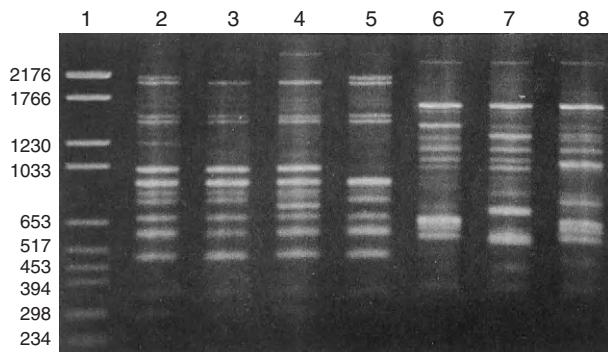
PCR-RFLP makes RFLP analysis suitable for studying individual specimens of very small species, requires no labeled probes, and is faster and less expensive than standard RFLP analysis. If consensus primers are available, then cloning is not required. For example, Simon et al. (1993) analyzed mtDNA in 13- and 17-year periodical cicadas using standard primers for the COII-A6-A8-COIII segment (Simon et al. 1991). A disadvantage to PCR-RFLP is that the method requires two procedures: allele-specific PCR followed by restriction digest.

### 13.5.10 RAPD-PCR

Hadrys et al. (1992) noted that RAPD-PCR is a versatile method for molecular ecology because it can be used to determine biotypes or species identity, assess kinship, and analyze paternity. It can estimate genetic variation within populations and can be used to monitor colonization. RAPD-PCR is suitable for studying insects for which very little genetic information is available, requires very small amounts of DNA, and can be used with very small insects. It is rapid and relatively inexpensive when compared to RFLP analysis, DNA sequencing, PCR-RFLP, DSCP, microsatellite, or microarray analysis (Table 13.1, see Chapter 8).

Because RAPD-PCR uses short primers of arbitrary sequence (10-mers), it does not require the investigator to know the sequences of specific genes in order to develop primers for the PCR. Haymer (1994) evaluated the sequences of various RAPD primers used on insects and listed 55 that had been particularly informative. RAPD-PCR primers sample both single-copy and repetitive DNA. Although the repeatability and reliability of RAPD-PCR can be problematic *if care is not exercised*, RAPD-PCR can provide inexpensive, repeatable and useful data (Penner et al. 1993, Edwards and Hoy 1993, MacPherson et al. 1993).

RAPD-PCR was used to analyze the amount of genetic variation in the parasitoids *Trioxys pallidus* and *Diglyphus begini* (Figure 13.2). DNA from individual *T. pallidus* males was amplified using 120 different primers (Edwards and Hoy 1993). Of the 120 primers, 92 produced a total of 342 scorable bands, of which 118 exhibited presence/absence polymorphisms. *Diglyphus begini* was evaluated with 25 primers, and 17 produced a total of 51 scorable bands in haploid males. The level of genetic variation detected was greater than any found in Hymenoptera by using allozymes (Menken 1991, Packer and Owen 1992) and comparable to the variation detectable with RFLPs. The bands considered reliable were inherited as dominant Mendelian traits in diploid females. Because only small amounts of DNA are needed for each RAPD reaction, multiple reactions could be conducted using DNA from a single individual (Edwards and Hoy 1993, 1995). RAPD-PCR can be used to analyze population structure and gene flow and to monitor establishment and dispersal of particular biotypes



**Figure 13.2** RAPD-PCR banding patterns obtained from individual *Trioxys pallidus* male wasps (lanes 2–5) and individual *Diglyphus begini* males (lanes 6–8) provide clear species differences. Size standards are in lane 1 for reference. (Photograph provided by O. R. Edwards.)

(Black et al. 1992, Kambhampati et al. 1992, Hadrys et al. 1993, Edwards and Hoy 1995).

RAPD-PCR has drawbacks. RAPD bands are inherited as dominant traits in diplo-diploid species, thus heterozygotes normally cannot be identified. In the haplo-diploid Hymenoptera, this difficulty is overcome by performing the analysis on haploid males only, or by testing the genotype of females by testing her male progeny (Edwards and Hoy 1993). RAPD-PCR conditions should be optimized for each species (Hadrys et al. 1992). Reactions must be repeated to determine which bands are “reliable” and consistent. RAPD-PCR is sensitive to the concentration of template DNA in the reaction, so reaction conditions must be optimized carefully and DNA extraction techniques must be consistent (Edwards and Hoy 1993). Primer quality is critical; fresh, undegraded primers should be used.

### 13.5.11 Sequencing

Sequencing provides large amounts of information about mtDNA and nuclear DNA (Thomas and Klaper 2004). However, sequencing, because of time and economic constraints, can sample only a tiny fraction of the total genome. Also, despite reductions in the cost of sequencing, it remains more expensive and time consuming than PCR-based methods. Sequencing of DNA amplified by allele-specific PCR requires information about sequences to develop specific primers. Sequencing has been used infrequently for large-scale population studies because of cost (Table 13.1), but it is a useful tool for other techniques, such as developing allele-specific primers for the PCR. See Chapter 12 for methods of

sequence analysis. A variety of computer programs for data analysis are available at the J. Felsenstein website ([www.evolution.genetics.washington.edu](http://www.evolution.genetics.washington.edu)) and Mount (2004) and Gibson (2002) provide detailed information for novices.

### 13.5.11.1 DNA Barcoding

DNA barcoding, the use of a short, standardized genetic marker (usually cytochrome oxidase subunit I [COI] sequences, see Chapter 12) to provide rapid DNA-based identifications of organisms, can be used in molecular ecology studies (Smith et al. 2009, Jinbo et al. 2011). Although barcoding may not always reliably identify some insect species (Whitworth et al. 2007, Song 2008), it can be useful for surveying biodiversity, may allow rapid identification of species and cryptic species, and may be useful for associating dimorphic sexes in insects (Whitworth et al. 2007). Whitworth et al. (2007) used barcoding to confirm that cryptic species of bees in the genera *Andrena* and *Ceratina* are present in Nova Scotia, Canada (Sheffield et al. 2009).

Casquet et al. (2012) tested Chelex in a modified protocol as a method of extracting DNA for high-throughput sequencing to develop the data necessary for large-scale biodiversity studies by barcoding. By eliminating the boiling step, Casquet et al. (2012) were able to reduce handling steps, and the procedure could be performed in a 96-well plate, thereby allowing the extractions to be automated. The method was described as “quick and cheap” and produced ds DNA from spiders because boiling was eliminated. The DNA produced could be stored for 1.5 years, perhaps because the DNA was double-stranded.

Song et al. (2008) found that barcoding may overestimate the number of species of grasshoppers. Nuclear mitochondrial pseudogenes (numts) are non-functional copies of mtDNA in the nucleus that can be found in a number of eukaryotes and can be coamplified with mtDNA using conserved universal primers. Numts can be highly divergent from the mtDNA sequences in the individual and sequencing of the numts, using the standard of 3% sequence divergence, can result in erroneous estimates of species diversity. More than 82 eukaryotes are now known to have numts, and it appears that numts may be more common than realized, providing a challenge to DNA barcoding as an accurate method for species identification (Song et al. 2008). Pamilo et al. (2007) searched the complete genomes of *D. melanogaster*, *Anopheles gambiae*, *Apis mellifera*, and *Tribolium castaneum* for numts and found that *Tribolium* and *Apis* had numerous numts, whereas numts in the other two insects were rare.

Moulton et al. (2010) tested the hypothesis that increasing primer specificity could eliminate numt coamplification in DNA barcoding. They studied 11 lineages in the Orthoptera and found that numts were coamplified in all 11 lineages when

using universal primers, "suggesting that numts may be widespread in other taxonomic groups." Increased primer specificity reduced numt coamplification in some species, but only eliminated it in one species tested. Moulton et al. (2010) concluded, "numt coamplification is a serious problem for DNA barcoding and more quality control measures should be implemented to identify and eliminate numts before using mitochondrial barcodes for species diagnoses." Blacket et al. (2012) found that the Queensland fruit fly *Bactrocera tryoni* also had a numt that could be coamplified. However, they designed an alternative primer that allowed them to amplify the COI gene rather than the numt.

#### **13.5.12 Single Nucleotide Polymorphism (SNP) Markers**

cDNAs from species can be sequenced using next-generation sequencing methods, and large numbers of transcripts can be sampled (Helyar et al. 2011). Genetic variation within these expressed sequence tag (EST) libraries can be detected and used to study population genetics in insects or for breeding programs. SNPs have been used in assays for pesticide resistance genes (Brun-Barale et al. 2005). Selection of honey bees for resistance to Varroa mites could involve breeding for hygienic behavior; Spotter et al. (2012) obtained 70,000 SNPs obtained from the honey bee genome project and validated them by using next-generation sequencing of pooled DNA samples of bees resistant to Varroa and of DNA from bees that are susceptible. Approximately 44,000 SNPs were validated that could be used to select for multiple desirable traits in honey bees.

The traditional methods for identifying SNP markers is time-consuming and expensive, but Black and Vontas (2007) describe methods for obtaining SNP markers from insects that do not involve genome sequencing.

### **13.6 Analysis of Molecular Data**

Molecular ecology is a rapidly changing field of study (Ferraris and Palumbi 1996, Symondson and Liddell 1996). Methods of analyzing molecular population data are being developed and improved. For reviews and overviews of methods, see Weir and Cockerham (1984), Slatkin and Barton (1989), Doolittle (1990), Lynch and Crease (1990), Weir (1990), Hoelzel and Dover (1991), Hoelzel and Bancroft (1992), Ferraris and Palumbi (1996), Bossart and Pashley-Powell (1998), Estoup and Angers (1998), Howard and Berlocher (1998), Rieseberg (1998), Schnabel et al. (1998), Templeton (1998), Davies et al. (1999a,b), Bohonak (1999), Cornuet et al. (1999), Goodnight and Queller (1999), Schwartz et al. (1998), Black et al. (2001), and Hewitt (2001). Computer software packages for molecular population genetic analyses are available from several sources (e.g., Rozas and Rozas [1997] and the Felsenstein website: [www.evolution.genetics.washington.edu/pub](http://www.evolution.genetics.washington.edu/pub)).

A variety of analyses can be conducted on molecular data to estimate parameters such as genetic diversity (heterozygosity and proportion of polymorphic loci), interpopulation diversity, genetic distance, effective population size, kinship, paternity, and the effect of migration on population diversity, as described below. Fundamentally, the techniques described above produce one of two types of data: sequence data or allele-frequency data (Table 13.1). Details of the statistical methods used are beyond the scope of this chapter. However, the references cited above can provide an entry to the extensive and growing literature (Table 13.2).

### 13.6.1 Allozymes

The visualization and interpretation of allozyme data was reviewed by [Pasteur et al. \(1988\)](#) and [May \(1992\)](#). Allozyme data can be used to obtain gene frequencies ([Hoelzel and Bancroft 1992](#)).

$$p = (2N_{AA} + N_{Aa}) / 2N \text{ and } q = (2N_{aa} + N_{Aa}) / 2N$$

where  $p$  is the frequency of the A allele and  $q$  is the frequency of the a allele;  $N$  is the total number of individuals in the sample; and  $N_{AA}$ ,  $N_{aa}$  and  $N_{Aa}$  are the number of individuals with AA, aa, and Aa genotypes, respectively. According to the Hardy–Weinberg rule, the proportion of AA individuals should be  $p^2$ ; the proportion of aa individuals should be  $q^2$ ; and the proportion of heterozygotes should be  $2pq$  in an ideal population (infinitely large random-mating population) in which there is no selection, migration, or mutation. Such a population is in **Hardy–Weinberg equilibrium**.

**Polymorphism (P)** and **heterozygosity (H)** can be calculated for allozyme data ([Hoelzel and Bancroft 1992](#)).  $P$  is the proportion of polymorphic loci and  $H$  is the proportion of heterozygous loci. When the population is in Hardy–Weinberg equilibrium, heterozygosity can be calculated from allele frequencies at a given locus by the following equation:

$$h = 1 - \sum X_i^2$$

where  $X_i^2$  is the frequency of the  $i$ th allele at a given locus. The proportion of heterozygous individuals ( $H$ ) is the average heterozygosity for all loci studied, so is calculated as the mean of  $h$  over all loci.

**Genetic distance** between populations can be calculated using allozyme data. Most analyses of genetic distance assume that molecular genetic changes are accumulating gradually at a constant rate and that most changes are selectively neutral. This suggests that the genetic changes can be used to estimate

**Table 13.2: Selected Sequences and Databases Relevant to Insect Molecular Ecology.**

Primer or database type		Reference(s) <sup>a</sup>
<b>Insect genomes</b>		
<i>Bombyx mori</i>	SilkBase	<a href="http://silkbbase.ab.a.4-tokyo-ac.jp/cgi">http://silkbbase.ab.a.4-tokyo-ac.jp/cgi</a> , Mita et al. 2004
<i>Drosophila melanogaster</i>	FlyBase	<a href="http://flybase.org/">http://flybase.org/</a> , Gilbert 2002, Wilson et al. 2008
<i>Anopheles gambiae</i>	<i>AnoBase</i> , The <i>Anopheles</i> database VectorBase: <i>Anopheles</i>	<a href="http://www.anobase.org/">http://www.anobase.org/</a> <a href="http://agambiae.vectorbase.org/">http://agambiae.vectorbase.org/</a>
<i>Apis mellifera</i>	Expressed sequence tags BeeBase	Whitfield et al. 2002
<b>Hymenoptera genomes</b>		<a href="http://hymenopteragenome.org/beebase/">http://hymenopteragenome.org/beebase/</a> , Munoz-Torres et al. 2011
<b>Microarrays</b>	Overview of methods and data analysis	Quackenbush 2001; Gibson 2002; Kerr 2001a,b; <a href="http://www.gene-chips.com">www.gene-chips.com</a>
<b>Microsatellites</b>	Overview of methods and analysis  InSatDb Unigene Microsatellite DB Microsatellite DB of fully sequenced insect genomes	Estoup and Angers 1998, Toth et al. 2000, Zane et al. 2002 Archak et al. 2007 veebyasg.ubfi. <i>Nucleic Acids Res.</i> annual updates
<b>Mitochondrial DNA</b>	MitBASE  AMmtDB Variable number of tandem repeats in mt DNA PCR of long sections in 14 orders 12S rRNA PCR primers Alignments of mitochondria from 13 insect orders Analysis of 12 species in four orders Long-PCR primers for entire mt genomes IMGD: comparative database of 25,747 insect mt genomes	Attimonelli et al. 2000 Lanave et al. 2002 Lunt et al. 1998 Roehrdanz and DeGrugillier 1998 Hickson et al. 1996 Simon et al. 1994 Buckley et al. 2000 Chandra et al. 2006 Grandebau et al. 2005 Lee et al. 2009
<b>RAPD-PCR</b>	Effective 10-mer primers in insects reviewed	Hadrys et al. 1992, Haymer 1994
<b>Ribosomal</b>		
5S rRNA	5S rRNA	Cullings and Vogler 1998, Szymanski et al. 2002
rRNA	Small subunit rRNA	Wuyts et al. 2002
<b>Single nucleotide polymorphisms (SNPs)</b>	Variation in protein-coding and RNA-coding genes	Black and Vontas 2007, Coates et al. 2011

<sup>a</sup>A review of databases is published each January by the journal *Nucleic Acids Research*; search the most recent issue for updated locations and new databases. Also search the National Center for Biotechnology Information (NCBI) website for a variety of databases and data analysis methods ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)).

the time of genetic differentiation within and between populations. When DNA variation is measured directly, the statistical analyses assume 1) nucleotides are randomly distributed in the genome, 2) variation arises by base substitution, 3) substitution rates are the same for all nucleotides, and 4) all relevant bands or fragments can be detected and bands that comigrate but are different are not scored as identical (Hoelzel and Bancroft 1992). Although the first three assumptions usually are not valid, it is thought that small deviations from them will not alter the conclusions significantly.

The most commonly used method for analyzing genetic distance in populations by protein polymorphisms is that of Nei (1972). In two populations, X and Y, the probability that two randomly chosen genes at a single locus ( $j_k$ ) are identical is determined as follows:

$$j_x = \sum X_i^2 \text{ and } j_y = \sum Y_i^2$$

where  $x_i$  and  $y_i$  are the frequencies of the  $i$ th alleles at a given locus in populations X and Y, respectively. If there are two alleles at this locus with frequencies  $p$  and  $q$ , then

$$j = p^2 + q^2$$

The probability that a gene is identical at the same locus in populations X and Y is

$$j_{xy} = \sum X_i Y_i$$

The normalized identity ( $I$ ) between populations X and Y for all loci is as follows:

$$I = J_{xy} / (J_x J_y)^{1/2}$$

where  $J_{xy}$ ,  $J_x$ , and  $J_y$  are the arithmetic means of  $j_{xy}$ ,  $j_x$ , and  $j_y$ , respectively, over all loci. **Nei's standard genetic distance ( $D$ )** between populations X and Y is then as follows:

$$D = -\ln(I)$$

where I is multiplied by the natural logarithm ( $\ln$ ) to give a value that is 0.0 for genotypes that are completely dissimilar. The relationship between D and time ( $t$ ) is as follows:

$$t = 0.5aD$$

where  $a$  is the average rate of detectable change per locus per year.

Interpopulation diversity using allozyme data are usually measured using the **coefficient of gene differentiation ( $G_{ST}$ )**.  $G_{ST}$  is derived by estimating the average similarity within and between populations.  $G_{ST}$  is an extension of Wright's correlation ( $F_{ST}$ ) between two gametes drawn at random from each subpopulation. The coefficient of differentiation is determined as follows:

$$G_{ST} = (H_T - H_S) / H_T$$

where  $H_S$  is the average gene diversity within populations, and  $H_T$  is the inter-population gene diversity.

### 13.6.2 Microsatellites

Microsatellites are the most widely applied molecular markers in population genetic studies, conservation genetics, and paternity tests because the markers are codominant, highly reproducible, and can be obtained by the PCR once primers can be developed (Ellegren 2004). Microsatellites can be generated by analysis of sequences available in GenBank, especially for species that have had their genomes sequenced. Alternatively, microsatellite-enriched libraries can be generated, although this can be time-consuming, costly, and technically complex. Techén et al. (2010) describe an optimized method for constructing microsatellite-enriched libraries (Table 13.3). Microsatellites can identify multiple loci or single loci in individuals (Zane et al. 2002, Ellegren 2004). Satellite sequences scattered throughout the chromosomes can produce a series of bands that are often specific to an individual insect. However, microsatellites may detect so much variation within populations that it is difficult to analyze them unless inbreeding has occurred in the population under study so that some of the variability has been lost. DNA fingerprinting using microsatellites can be done with the PCR using specific or consensus primers (Kirby 1990).

Population estimates of allele and genotype frequencies can be tested for correspondence to Hardy-Weinberg equilibrium conditions (Bruford et al. 1992), and genetic differentiation can be calculated from microsatellite data (Goodman 1997, Goodnight and Quellar 1999). The high level of variation detected by satellite data make it feasible to test for paternity and to conduct studies of variability within both sexual and clonal populations (Brookfield 1992). Variation and genetic distance also can be calculated (Hoelzel and Bancroft 1992). Single-locus DNA fingerprinting is easier to analyze because there are fewer bands but advances are being made in analysis of multiple microsatellite loci (Estoup and Angers 1998).

**Table 13.3: A Simplified Method for Constructing Microsatellite-Enriched Libraries.**

- Genomic DNA is digested in four separate reactions using two endonucleases that produce blunt ends to produce DNA fragments 200–800 bp in length.
- Digested DNA is A-tailed to facilitate ligation to adapters.
- Two random adapters 20–23 nt in length are designed with a 3' T overhang to facilitate ligation to genomic DNA fragments having an A overhang at the 3' end.
- Genomic DNA and the adapter sequences are ligated, then cleaned.
- DNA (with the adapters at each end) from the ligation step is amplified by the PCR using the adapter oligo as a primer.
- Cleaned DNA from the PCR is hybridized to biotinylated oligo repeats.
- Magnetic beads are prepared to bind the biotinylated DNA.
- Beads are washed, releasing single-stranded DNA.
- Single-stranded DNA is amplified by the PCR.
- The PCR products are cloned into a T-A vector and transformed into cells. Plasmids are extracted and sequenced.
- Sequenced clones are analyzed to identify sequence repeats with 1–8-bp motifs.
- Specific PCR primers are designed based on the motifs and tested to determine whether they produce useful PCR products for analysis.

Adapted from Techén et al. (2010).

### 13.6.3 RAPD-PCR

RAPD-PCR bands are considered as dominant loci in diplo-diploid organisms, and scored as present or absent (Hadrys et al. 1992). Kambhampati et al. (1992) discussed appropriate statistical methods for analysis of data. It appears that RAPD-PCR loci can be used to determine paternity, kinship, and hybridization, as well as to estimate population heterozygosity, effective population size, identify biotypes and cryptic species, genetic distance between populations, and interpopulation diversity (Table 13.1).

### 13.6.4 RFLPs

Visualization and interpretation of RFLP data was described by Aquadro et al. (1992) and Dowling et al. (1990). Restriction patterns can be compared either by the lengths of the fragments or by comparing actual restriction sites. Restriction patterns can be classified as haplotypes and a measure of diversity can be derived as a function of the frequency of the different haplotypes (Hoelzel and Bancroft 1992). The term *haplotype* is a contraction of haploid and genotype and describes the combination of linked alleles in a cluster of related genes. Likewise, genetic distance is measured as an estimate of the number of base substitutions per nucleotide separating the two populations. Interpopulation diversity ( $G_{ST}$ ) can be estimated in a manner similar to that for allozyme data,

but gene identities must be estimated from RFLP patterns. RFLP data also can be analyzed as changes in base pairs if the assumption is made that each change in restriction pattern is caused by a change in a single base pair.

### 13.6.5 Sequencing

Computer programs are used to analyze DNA sequence data to determine the best alignment (Doolittle 1990, Gribskov and Devereux 1991, Gibson and Muse 2002, Mount 2004). The identity of two sequences is compared on the percentage of shared bases. Deletions and insertions (indels) are usually scored as a single change regardless of length (Hoelzel and Bancroft 1992). As with proteins or RFLP data, nucleotide diversity, genetic distance, and interpopulation diversity can be estimated (Hoelzel and Bancroft 1992). Also see Chapter 12 for details on DNA sequence analysis methods.

## 13.7 Case Studies in Molecular Ecology and Population Biology

An extensive and growing literature published in a variety of journals and books makes it impossible to provide a comprehensive overview of the effect that molecular methods are having on theoretical and applied insect ecology. Thus, some “case studies” are presented to illustrate applications of different molecular tools and their statistical methods.

### 13.7.1 Genetic Variability in the Fall Armyworm: Incipient Species or Multiple Species?

The fall armyworm, *Spodoptera frugiperda*, is a polyphagous lepidopteran “species” that attacks over 60 plants, particularly corn and rice (but also Bermuda grass, soybean, peanut, cotton, and alfalfa). It is found throughout the Western Hemisphere but, because it does not diapause, must colonize northern states in the United States and in Canada by migrating from overwintering sites in Florida and Texas (Nagoshi and Meagher 2008, Nagoshi et al. 2009). Its range also extends south into Argentina and Brazil (Nagoshi et al. 2007). The “fall armyworm” has exhibited anomalous differences in tolerance to pesticides and other life-history traits that are important in pest-management practices. It appears that it is, in fact, undergoing “incipient speciation” (Pashley 1986). However, the genetics of the fall armyworm may be complex because Nagoshi et al. (2007) reported that the “corn strain” might consist of subgroups.

The undetected presence of cryptic species has practical significance for pest-management programs. Pashley (1986) showed there are two “species,” each associated with different host plants in the southern United States; one is associated with rice (and Bermuda grass) and the other with corn (and cotton and

sorghum). The two occur sympatrically in the United States, exhibit a high level of reproductive isolation, are physiologically adapted to their different host plants, and the physiological differences are genetically based (Pashley 1988). These populations have undergone a series of molecular analyses to investigate their status. As you will note, the sophistication of the molecular methods and the statistical analysis methodology used increases through time.

Allozyme and RFLP analysis of mitochondrial (mt) DNA indicated the two types could be distinguished and that there is a near absence of gene flow (Pashley 1989). Lu et al. (1992) reported there were RFLP differences in genomic DNA in the two types after analyzing six colonies with 22 different markers. Lu et al. (1994) also found that repeated DNA sequences (microsatellites) in the two populations differed.

The extent to which the sympatric populations of rice and corn strains of *S. frugiperda* interbreed is problematic (Pashley 1986, Pashley et al. 2004). Some data suggest there is a unidirectional behavioral barrier to interstrain mating, but other data do not support this (see review by McMichael and Prowell 1999). To answer whether these populations interbreed, several molecular markers were used because allozymes, mtDNA and nuclear DNA markers did not provide sufficient resolving power to discriminate between the alternative hypotheses of low frequencies of hybridization vs. expected genetic overlap between two closely related populations. AFLP-PCR was evaluated to determine whether this tool might uncover unique genetic markers in each of the two populations, which would allow hybridization to be detected. AFLP genotyping or fingerprinting is thought to be a useful tool for assessing genetic diversity, relatedness, population structure and phylogenetic relationships (Mueller and Wolfenbarger 1999). Mueller and Wolfenbarger (1999) noted that AFLPs can be more reliable than RAPD markers, and can be easier to use than RFLPs and microsatellites.

McMichael and Prowell (1999) used ten AFLP markers to compare the two populations of *S. frugiperda*. The AFLP data identified two populations that matched up with the majority of individuals from one or the other of the host-associated strains, as defined by habitat and mtDNA. Unfortunately, not all individuals could be assigned to the "rice" or "corn" populations. To date, "no pair of markers shows complete congruence with each other or host of origin. In other words, allozyme or mtDNA genotypes characterizing 1 strain can occur in individuals collected on the other strain's host. Individuals on a single host can contain an allozyme genotype characteristic of 1 strain but a mtDNA genotype of the other" (McMichael and Prowell 1999). These results cannot discriminate between the alternative hypotheses: sharing of alleles because the variability in

the common ancestor of the two strains was retained during their divergence, or interstrain hybridization. The goal to identify diagnostic, and unique, AFLP markers failed and [McMichael and Prowell \(1999\)](#) concluded that future studies, in which AFLP data are combined with mt markers and allozymes, might resolve the hybridization question. Later, [Pashley et al. \(2004\)](#) compiled mtDNA data, an esterase locus and eight AFLP loci in moths collected in Louisiana, Florida, Puerto Rico, Guadeloupe and French Guiana and concluded that 16% of individuals were potentially hybrids with a minority being F1 progeny. The hybrids were found primarily in corn, a habitat used by both strains. [Pashley et al. \(2004\)](#) stated, "these data support introgressive hybridization between recently evolved species that are not completely reproductively isolated." The molecular data, in combination with other data, have shown the two populations are different and explain a long-standing concern of practical pest-management importance. Whether these populations are called species, incipient species, or host races is a judgment call that is based on whether the scientist is a "splitter" or a "lumper."

Another issue of relevance to pest-management specialists is where the fall armyworm overwinters and how it colonizes crops in the northern United States; efforts to understand their migrations previously were based on trapping fall armyworm adults ([Nagoshi and Meagher 2008](#)). [Nagoshi et al. \(2009\)](#) used COI mt genes to determine that most "corn" populations in the United States overwinter in southern Texas and Florida and adults from central Pennsylvania originated from Texas.

Studies of the fall armyworm in South America indicated two biotypes are present that have different host-plant preferences and other biological and ecological differences ([Nagoshi et al. 2007](#)). PCR-RFLP analysis of populations from Brazil indicated the mt haplotypes found in corn in Brazil were genetically distinct from corn populations in Florida and suggested future studies could resolve whether corn populations in South and Central America were of the Brazil or Florida haplotype, which could resolve questions about the long-range movements of these populations. [Clark et al. \(2007\)](#) conducted AFLP analyses of 23 populations from Mexico, the United States, Puerto Rico, Brazil, and Argentina and concluded that the majority of genetic variability is within populations and not between populations, suggesting there is minor gene flow and that the corn moths in the Western Hemisphere "are an interbreeding population." They stated, "*S. frugiperda* is highly genetically variable."

[Arias et al. \(2011\)](#) evaluated 174 microsatellite markers using the method developed by [Techen et al. \(2010\)](#) and screened 15 moths from eight families

each from Puerto Rico, Texas, and Mississippi in an effort to understand migration of the moth. This is of increasing importance to pest managers because populations in Puerto Rico are resistant to Cry toxins from *Bacillus thuringiensis* in transgenic corn. If the fall armyworm is not controlled, it can reduce corn yields by up to 73% (Arias et al. 2011). Currently, control relies on applications of pesticides timed to kill young larvae. Because Puerto Rico is only 1500 km from the United States, movement of resistant moths into the United States is possible. Arias et al. (2011) evaluated species-specific microsatellites as more-sensitive markers for population genetic analyses of the fall armyworm and conducted cluster analysis using 120 markers to calculate genetic distances among the populations sampled. These markers could become “an effective tool for population studies by using multiple loci to help better understand migrations and possible crosses of this insect.”

Lewter and Szalanski (2007) developed a PCR-RFLP method for analyzing a 611-bp region of the COI-COII genes of the mitochondrion to discriminate between seven noctuid species, including the fall armyworm, collected in pheromone traps in Arkansas. Moths found in the traps often are difficult to diagnose to species but the PCR-RFLP method allowed species diagnoses to be obtained within 1 day.

### 13.7.2 Analyses of Natural Enemies

Understanding the biology, behavior, and ecology of natural enemies should allow pest managers to improve their use in pest-management programs. Information on the hosts or prey of many natural enemies is limited. Monitoring natural enemies in the field after releases in classical biological control programs can be expensive and difficult. Molecular analyses are being developed to learn more about host or prey ranges, as described in the upcoming sections, and for monitoring establishment and spread after release.

#### 13.7.2.1 Analyses of Adult Parasitoid Wasps to Reveal Larval Hosts (MAPL)

Females of a hymenopteran or dipteran parasitoid species deposit eggs or larvae in or near a host. The immature parasitoid then devours the host, leaving it as an adult to locate another host. Many parasitoids have unknown life histories, but many adult parasitoids can be collected even if their larval hosts are unknown (Rougerie et al. 2011). Rougerie et al. (2011) were able to sequence a diagnostic DNA marker (a COI DNA barcode sequence) from 24% of 297 adult parasitoids that allowed them to identify the host used by the wasps during their larval stages. They called this method **Molecular Analysis of Parasitoid Linkages** or **MAPL**. Their results indicated that larval-host DNA can persist

through the molt to the adult stage of some wasp larvae and suggest MAPL may allow host-parasitoid associations to be discovered with precision if DNA barcode libraries become available for a majority of potential hosts in a specific geographic area. The use of MAPL also could reveal the use of nontarget hosts by parasitoids and allow extensive host-parasitoid associations. Note that DNA from host insects can persist for some time and sensitive PCR analysis methods can detect it in their parasitoid adults.

#### *13.7.2.2 Analyses of the Establishment of a Parasitoid Released in a Classical Biological Control Program*

*Lipolexis oregmae* was imported into Florida and released in a classical biological control project as a natural enemy of the brown citrus aphid ([Walker and Hoy 2003](#)). Another parasitoid, *Lysiphlebus testaceipes*, was known to parasitize this aphid in Florida, which complicated monitoring for establishment and spread of *L. oregmae*. Traditional methods for monitoring populations of aphids for parasitoids involve collecting aphids and holding them in the laboratory until adult parasitoids emerge. This is labor-intensive and many aphids die from fungal infections without producing a parasitoid adult. [Persad et al. \(2004\)](#) developed a high-fidelity PCR assay that allowed them to discriminate between immatures of *L. oregmae* and *L. testaceipes* without having to hold the aphids until the adult parasitoids emerged. The method was efficient and sensitive; a sensitivity analysis of the high-fidelity PCR protocol indicated that adding DNA extracted from a single parasitoid to the pooled DNA from 300 specimens of the brown citrus aphid allowed consistent detection of the parasitoid. This method of analysis was used to indicate establishment had occurred and subsequent analyses, using single aphids, could be conducted to indicate the rates of parasitism. This PCR protocol was used to document that *L. oregmae* had colonized brown citrus aphids in Jamaica, even before a proposed classical biological control program was conducted ([Hoy et al. 2007](#)). As a result, the time-consuming and expensive process of rearing and releasing *L. oregmae* in Jamaica was avoided.

#### *13.7.2.3 Gut Analyses of Predatory Insects to Reveal Prey Species*

Generalist insect predators can be difficult to study under field conditions, especially when they are active in the soil or at night. No-choice studies in the laboratory may not reveal the true importance of predation on specific prey species. PCR analyses of DNA present in predator guts can be used to determine what prey species are being eaten if experimental methods are validated ([Foltan et al. 2005](#), [Juen and Traugott 2005](#), [Sheppard et al. 2005](#), [King et al. 2008](#), [Weber and Lundgren 2009](#)). Sequences commonly used are the COI and COII fragments of mitochondria and the internal transcribed sequence 1 (ITS1) region of ribosomal RNA genes in the mitochondria. RAPD-PCR markers have been tested, but are less often used.

Molecular analyses of gut DNA require that a series of steps be conducted to validate the tests. These steps include sampling each potential prey species found in the appropriate environment; sequencing appropriate genes from each prey (choosing the appropriate number of each species to sample to obtain an accurate estimate of the variability within each prey species); designing PCR primers that are species-specific; amplification of a diagnostic, but relatively small, fragment because the DNA is degraded in the gut to a greater or lesser degree; and feeding each prey species to hungry predators and conducting PCR analyses to determine the time it takes after feeding at different temperatures for the PCR signal to be undetectable.

Variables that can affect the decay rate of the DNA in the predator's gut include the number of prey consumed, the size of the prey (which affects the quantity of the DNA present initially), the number of different prey species consumed at nearly the same time, the preservation methods for the predators, the DNA extraction methods, and the temperature at which the predators were held. It can be difficult to determine whether the predators fed on live prey or on dead prey (Foltan et al. 2005) or whether a generalist predator fed on another predator that had fed on a particular prey species (secondary predation) (Sheppard et al. 2005). Sensitivity tests for the PCR should be conducted using prey DNA mixed with predator DNA in serial dilutions in order to determine how sensitive the tests are, and both positive and negative controls should always be used to reveal problems with the reaction or with contamination, respectively. Furthermore, the primers should be tested on predator DNA to confirm there is no cross amplification.

King et al. (2008) reviewed the methods used in gut analyses of predators and recommended specific practices to improve procedures. Primer design and testing and assay optimization are described, and the need for negative controls was emphasized. Multiplex PCR can sometimes be used to screen for different prey species simultaneously, and quantitative PCR can potentially quantify predation rates (Weber and Lundgren 2009). However, if hundreds of predators need to be studied for predation on many prey species, the cost of such studies is high, and the evaluations become tedious and lengthy. King et al. (2008) speculated that microarrays or pyrosequencing of DNA in predator guts might be developed in the future for these assays.

Methods also are needed to ensure that PCR analyses of gut contents are not reporting DNA contamination on the external insect (Greenstone et al. 2012). Treatment with 2.5% commercial bleach reliably eliminated external DNA contamination but 80% ethanol (EtOH) did not. The bleach treatment did not affect the ability to detect the true prey species in the gut by the PCR.

Another issue of interest to molecular ecologists studying the diet of predators is the fact that generalist predators may be preying on species that are not yet characterized and O'Rorke et al. (2012) reviewed several methods to enrich prey DNA to improve detection, especially of short DNA fragments that have been partially digested in the predator's gut.

### 13.7.3 Population Isolation and Introgression in Periodical Cicadas

Molecular markers have been used to resolve the evolutionary origins of species of periodical cicadas (*Magicicada*). The biology, ecology, and evolution of periodical cicadas are complex and unusual (for a review, see Marshall 2001). Periodical cicadas feed underground on roots in the deciduous forests of the eastern United States for either 13 or 17 years and emerge in very large numbers as adults to mate and deposit eggs nearly every year in some part of the range. The immense populations, sometimes as large as 1.5 million individuals per acre that emerge in the same year, are called broods. This synchronized emergence may have evolved because the large numbers allow most of the individuals to escape predation at a particular location and the long life cycles may prevent predator populations from synchronizing with the local emergences.

In the Mississippi Valley and southern United States, the life cycle of three *Magicicada* species is 13 years, whereas it is 17 years for three species in the north and west. Each species appears most closely related to another with the alternative life cycle, so that there are "species pairs" (13 paired with 17). This pattern suggests that speciation in *Magicicada* may involve a combination of geographic isolation and life-cycle changes that create reproductive isolation by changes in emergence patterns.

Thirteen of the possible 17 broods of a 17-year cicada *M. septendecim* and three of the possible 13 broods of a 13-year cicada *M. tredecim* have been identified and their emergence patterns have been determined and monitored (Marshall 2001). Although most broods emerge as scheduled, small numbers of a brood may emerge "out of step" with their cohort, which has created problems in understanding the species status of some broods. Some portions of 17-year broods appear to have accelerated their emergence by 4 years in certain sites, and both 13- and 17-year broods may emerge in the same geographic region. The reason for the 13- and 17-year cycles may be because the life cycle of the 17-year cicada includes a 4-year inhibition (diapause) of early nymphal growth. It was suggested that if this inhibition were eliminated, the 17-year brood could emerge after only 13 years. Because it appears that a single gene controls this aspect of the life cycle, a relatively simple genetic change could have a large effect. An alternative hypothesis for the change in brood duration is that a

17- and 13-year brood emerged together in 1868 in Illinois and "hybridized," which resulted in a population that subsequently emerged every 13 years.

Were the cicadas newly emerging after 13 years derived by hybridization or by loss of a 4-year diapause? To resolve this intriguing evolutionary and ecological question, [Martin and Simon \(1988\)](#) analyzed the abdominal sternite color, the frequency of allozyme polymorphisms, and mt DNA of the "hybrid" Illinois population. The data indicated that mt DNA in the 13-year Illinois brood is like that in the adjacent 17-year brood and distinct from that of the neighboring 13-year brood. The new brood is like the 17-year brood in abdominal color and frequency of PGM (protein) polymorphism. The results are consistent with the hypothesis that the two populations hybridized and that the 13-year life-cycle trait is dominant.

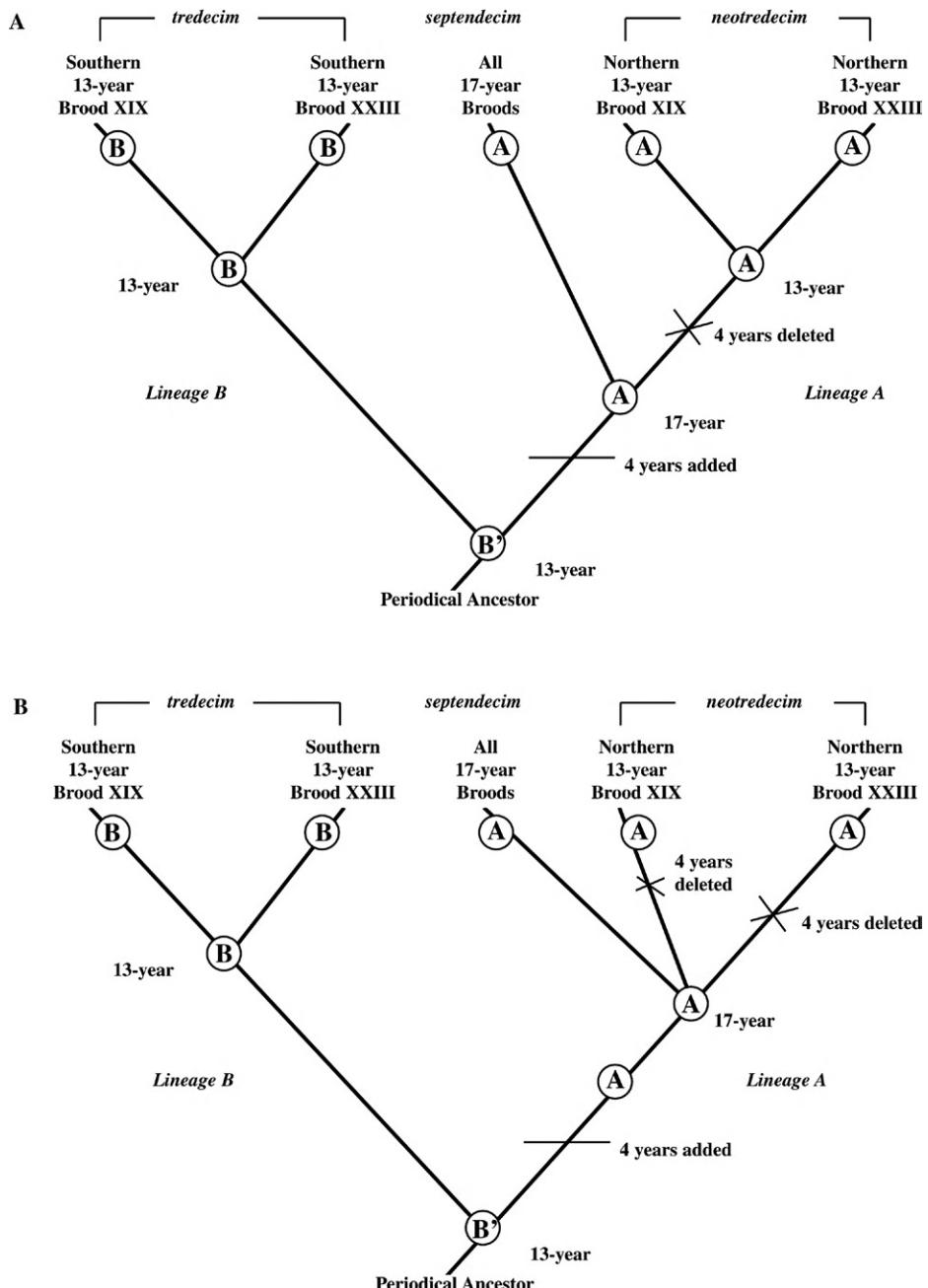
Nevertheless, [Martin and Simon \(1988\)](#) rejected the hybridization hypothesis. They pointed out that most cases of hybridization involve narrow zones in which the species come into contact, yet this new periodical cicada population occurs over a large area. They also noted that the complete elimination of one of the mt DNA genotypes throughout the entire region would require extremely strong selection because there have been only eight generations since 1868 upon which selection could have occurred. Furthermore, if hybridization occurred, intermediate phenotypes should have occurred, but did not. As a result, the findings were interpreted as evidence for a widespread life-cycle switch, *without hybridization*, in which a large number of 17-year cicadas (*M. septemdecim*) underwent a 4-year acceleration in development to become 13-year cicadas in northern Arkansas, Missouri, Illinois, and southeastern Iowa ([Martin and Simon 1988](#)). After 1868, a portion of the progeny of brood X in these areas had a permanent 4-year acceleration in development rate. Some cicadas in the region did not switch life-cycle length after 1868 and continued to emerge after 17 years. [Martin and Simon \(1988\)](#) suggest the switch in life-cycle length could have been triggered by environmental causes, perhaps by high-density populations. This change in life-cycle length would disrupt gene flow and initiate genetic divergence in the formerly unified 17-year cicada population. Furthermore, the individuals that switched their life cycle from 17 to 13 years now emerge synchronously with a previously isolated brood that emerges every 13 years (a *M. tredecim* brood). Gene flow between these previously isolated broods could occur if they can mate and produce viable progeny.

Further analyses of the populations by [Marshall and Cooley \(2000\)](#) found that the two 13-year sympatric cicada populations in the midwestern United States exhibit song differences and are unlikely to interbreed in the field. As a result,

they described the “new” 13-year population as a previously undescribed species called *Magicicada neotredecim*. *Magicicada neotredecim* and *M. tredecim*, both 13-year cicadas, overlap geographically and, since 1868, their broods overlap chronologically. The *tredecim* and *neotredecim* populations differ in abdomen coloration and mtDNA. Such traits in *M. neotredecim* are not consistently different from the sympatric populations of the 17-year *M. septemdecim* and thus *neotredecim* appears most closely related to this geographically adjacent population with a 17-year life cycle. [Marshall and Cooley \(2000\)](#) suggest this is evidence that speciation in *Magicicada* involves temporal isolation.

[Simon et al. \(2000\)](#), in a companion article, supported the conclusion that *M. neotredecim* is a new cryptic species, and provided mt DNA data to support the conclusion that a life-cycle switch occurred to produce two overlapping 13-year cicada lineages. Furthermore, the genetic evidence suggests that assortative mating is taking place in the area where the two populations of 13-year cicadas overlap. Thus, [Simon et al. \(2000\)](#) propose two possible scenarios for the evolution of the *septemdecim*–*tredecim*–*neotredecim* species ([Figure 13.3](#)). In both models, the ancestor is a cicada with a 13-year life cycle, which gave rise to a 13-year lineage (*tredecim* lineage) and to a new lineage that had a 4-year extension (17-year *septemdecim* lineage). In one scenario, *neotredecim* evolved from *septemdecim* via a single 4-year life cycle reversion (instantaneous speciation) ([Figure 13.3A](#)). In scenario two, the *septemdecim* lineage gave rise to two independent 17- to 13-year life-cycle reversions, resulting in two *neotredecim* broods ([Figure 13.3B](#)). There were no genetic or behavioral data to distinguish between these two hypotheses.

[Simon et al. \(2000\)](#) addressed the question as to whether the 13- and 17-year cicadas should be called different species. The concept of “species” varies among different systematists, but [Simon et al. \(2000\)](#) support the thesis that species should be designated as soon as they are distinguishable if it is likely that they will remain extant and isolated long enough for reproductive isolation to develop. [Marshall and Cooley \(2000\)](#) presented behavioral data and field observations to suggest that important differences do exist in male song and female response, supporting the separation of *neotredecim* and *tredecim* as species. They argue that *neotredecim* is unlikely to revert to a life cycle that is synchronized with any extant 17-year brood. Despite behavioral and genetic similarities between the 17- and 13-year broods, it is likely that reproductive isolation will be achieved between *neotredecim* and *septemdecim* due to the differences in brood timing. [Simon et al. \(2000\)](#) summarize their hypothesis as follows: “an initial allochronic event separated lineage A and B periodical cicadas and a second allochronic event placed them in secondary contact.”



**Figure 13.3** Two hypotheses for the formation of 13- and 17-year lineages of the *Magicicada* *septendecim*, *tredecim*, and *neotredecim* group. A) The evolution of *neotredecim* involved one evolutionary event (labeled 2). B) Alternatively, the evolution of *neotredecim* involved two separate events (labeled 2 and 3). The models assume that *tredecim* populations were ancestral in both cases. (Modified from Simon et al. 2000.)

Cooley et al. (2001) carry the story forward and predict that the newly described *M. neotredecim* inhabits midwestern habitat that will be unsuitable for its survival during the next “ice age.” They argue that the most southern populations of *neotredecim* have the greatest “likelihood of colonizing a refugium during the next glacial cycle … then after the next glacial retreat all undisplaced *M. neotredecim* will have gone extinct, and the pattern of character displacement linking *M. neotredecim* to an allochronic speciation event in the *M. septemdecim* lineage will have been erased.” Thus, the *Magicicada tredecim-septemdecim-neotredecim* story supports the hypothesis that instantaneous speciation can occur by shifts in reproductive timing (allochrony), although the precise mechanism by which the 4-year shift occurred remains unknown. Cooley et al. (2003) hypothesize that the 2.5% difference in mitochondrial gene sequences between *N. tredicim* and *M. neotredecim/M. septemdecim* lineages suggest that this degree of genetic distance is correlated with  $\approx$ 1 million years of separation.

Nariai et al. (2011) asked the question how life-cycle switching could occur in small numbers of these cicadas and how the population could increase. In theory, small numbers could fail to reproduce or be destroyed by predators. The Allee effect was invoked to explain the change in life cycles. The Allee effect is “a positive relationship between any component of individual fitness and either numbers or density of conspecifics” (Stephens et al. 1999). Nariai et al. (2011) evaluated the possibility that life-cycle switching between 13- and 17-year cycles can be explained by introducing a few individuals into an isolated population with the other cycle, without the need for hybridization or introgression in hybrid zones. Their model assumed that the life cycle is controlled by alleles at a Mendelian locus, with one cycle dominant over the other and neither allele has a selective advantage. The simulation models were started with different proportions of two pure populations with alleles for either the 13- or 17-year cycles. In the model, if there were no Allee effects, the mixed population exhibits the genetically dominant life cycle (either 17- or 13-years). If moderate-to-strong levels of Allee effects are present, a population of a recessive cycle “may shift entirely to the genetically dominant cycle after the introduction of a few individuals with the dominant alleles” and “all broods of the recessive cycle disappear.” Nariai et al. (2011) conclude, “Our model makes an intriguing prediction. Life cycle switching by gene introduction appears to be possible under a moderate level of Allee effects … the direction of switching depends only on genetic dominance: switching from recessive to dominant cycles.” The model relied on several assumptions: no mating preferences, neither cycle alleles have a selective disadvantage, and the net reproductive rate per generation is the same for both types.

### 13.7.4 Eradicating Medflies in California?

The Mediterranean fruit fly (Medfly), *Ceratitis capitata*, is an immensely destructive pest of agriculture. It is a native of sub-Saharan Africa, but invaded the Mediterranean basin, portions of Central and South America, Hawaii, and Australia during the past 100–200 years. *Ceratitis capitata* is able to feed on >200 host plants and can survive in a variety of climates (USDA–APHIS 2002). Females deposit up to 1000 eggs in fruits (including peaches, pears, plums, apples, apricots, avocados, citrus, cherries, figs, grapes, guavas, kumquats, loquats, nectarines) or vegetables (peppers, tomatoes), and the resulting maggots eat the fruits or vegetables, leaving them mushy and infested with mold. Medflies can develop from egg to adult in 21 days, so populations can increase exponentially in favorable climates.

*Ceratitis capitata* is a quarantined pest, meaning that when it is detected in the United States efforts are made to eradicate it (USDA–APHIS 2002). Eradication efforts are justified because this pest significantly increases crop-production costs; pesticide applications to suppress Medfly can disrupt biological control of other crop pests; and fruits grown in Medfly-infested regions cannot be exported to Medfly-free areas, thereby affecting national and international trade. Furthermore, establishment of Medfly would create serious pest problems in backyard gardens and orchards.

Eradication efforts involve surveys, regulation, and control. Surveys conducted by the USDA–APHIS and the States involve placing Medfly traps in high-risk areas, especially near international airports and seaports. If an infestation is found, additional traps are placed to determine the extent of the infestation; unfortunately, the traps are not 100% efficient and very low populations can be missed. Control methods include application of aerial and ground bait sprays, release of large numbers of sterile flies (sterile insect release method [SIRM]), and application of pesticides to the soil under infested trees to kill larvae as they enter the soil to pupate and the adults as they later leave the soil. Movement of host plants out of the infested area is prohibited and, in some cases, infested fruits may be picked and destroyed.

#### 13.7.4.1 The Controversy

Since 1975, California has grappled with a controversial problem. The periodic and repeated appearance of Medflies in traps, especially in the Los Angeles basin, has raised questions as to whether California can export agricultural products as a “Medfly-free state” and whether the eradication efforts have been successful (Carey 1991, Abate 1993, Carey 1996a,b, Myers et al. 2000). This

problem received national and international attention due to the large amounts of money involved; California exports >US\$1 billion of fruits and vegetables to other countries and ≈US\$1 billion of produce is sold domestically and in neighboring countries. Estimates of damage to California's economy due to direct damage to agriculture and to related jobs are enormous ([Abate 1993](#), [CDFA 2002](#), [USDA 2002](#)). "If California fruits were quarantined from all foreign markets because of Medfly infestation, the state would suffer a loss of 35,000 jobs and experience reductions in output of \$3.6 billion.... The worst case would be if all other states also embargo California fruits. This could result in >132,000 jobs lost, US\$13.4 billion lost in economic activity and >\$3.6 billion in lost income to California families" ([CDFA 2002](#)).

Because southern California is a gateway to Latin America and Hawaii, where Medfly is endemic, there is a constant risk of Medfly introductions. Medflies could be transported by the millions of international travelers, commercial fruit smugglers, and mailed packages. For example, more than a million passengers enter the Los Angeles airport annually from Hawaii, where Medfly is endemic; if only one visitor in a thousand illegally transported infested fruit from Hawaii, there could be 1000 opportunities annually to create an infestation in California ([CDFA 2002](#)).

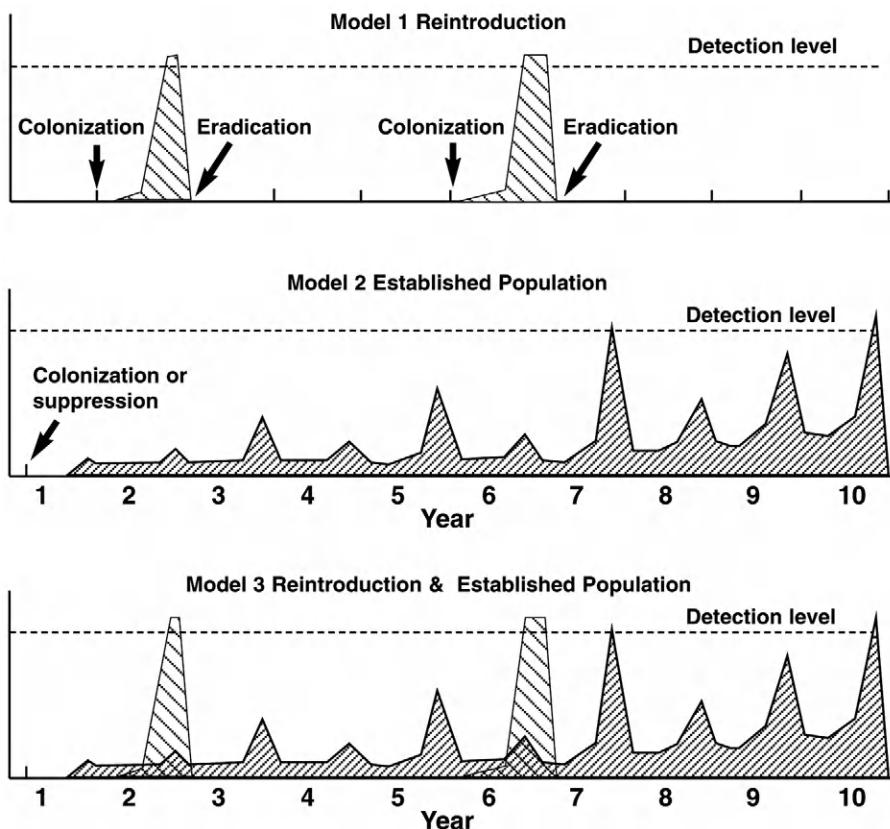
The first Medfly infestation in California was detected in Los Angeles in 1975; sterile Medfly releases and ground applications of malathion and bait were carried out and the infestation soon was declared eradicated. In 1980, a total of 180 Medflies were found and eradicated. Few Medflies were found between 1982 and 1987 in California. However, the questions really began in the late 1980s because 43 were found in Los Angeles County in 1987. Eradication efforts resulted in aerial sprays followed by the release of one million sterile Medflies per square mile per week. In 1988, 54 Medflies were found in two locations in Los Angeles County, which was treated with an aerial spray and releases of 1 million sterile Medflies per square mile per week. In 1989 and 1990, a total of 304 Medflies were found, which resulted in an eradication program that included 21 treatments in four counties covering 536 square miles. One wild Medfly was found in 1991 and trapping located an additional 24 flies; eradication efforts included trapping, ground application of malathion and bait and sterile Medfly releases over a 26-square mile area using 30 million sterile flies per week between October 1991 and August 1992. During 1992 and 1993, 202 wild Medflies were found in the Los Angeles basin and eradication efforts used trapping, ground applications of malathion and bait, and sterile releases. In 1993, 400 flies were found, leading to eradication efforts that included eight aerial applications of malathion and bait plus mass trapping. During 1994, a

"basin wide" sterile Medfly release program was initiated over a 1464-square mile area; 250,000 sterile flies per square mile per week were released over the entire area and 250,000 sterile flies were released, as well, in the areas where 73 wild flies were found in 1994; this effort was concluded in March 1996.

In 1996, a "preventative" release program was initiated over a 2155-square mile area in four counties to "prevent the development of Medfly infestations and to limit the geographic size of any that manage to start" (CDFA 2002). The releases involved at least 125,000 sterile Medflies per square mile per week and an additional 125,000 sterile flies over a high-risk area in central Los Angeles. No Medflies were found during 1996. During 1997, 24 wild Medflies were found in Los Angeles County, which led to increased efforts, including ground sprays, fruit stripping, soil drenches with pesticides under infested trees, and sterile fly releases that were increased to 500,000 per square mile per week within a 9-square mile area around each infested site.

Between 1975 and 1990, aerial spraying of malathion in baits had been used to kill the Medfly but, after considerable public concerns were raised about potential negative environmental and health effects, aerial application of this pesticide was banned in California. In the early 1990s, agricultural experts and entomologists hotly debated whether the eradication methods were effective. One "camp" believed that a combination of releasing millions of sterile Medflies, plus implementation of effective quarantines, and the use of attractants and traps to delineate the infestation zone and to monitor the effectiveness of the SIRM program was effective. Others believed that eradicating the Medfly from the Los Angeles basin was nearly impossible and that these methods were failures.

One of the significant questions in the debate is whether the Medfly was a "permanent resident" of California, especially in the Los Angeles area, or whether the ongoing outbreaks were the result of additional independent invasions ([Carey 1996a,b](#), [Headrick and Goeden 1996](#), [Myers et al. 2000](#), [Figure 13.4](#)). This issue was hotly debated and created considerable animosity. Until 1990, scientists and government officials assumed that each Medfly outbreak originated from flies that arrived from another country, hitchhiking within agricultural products imported into or smuggled into California. Many believed that the source of the invasions was from Hawaii. The other group believed, however, that the Medfly had become permanently established and was being maintained at undetectable levels, due in part to the inability of Medfly traps to detect very low-density populations.



**Figure 13.4** The top two models represent the alternative hypotheses for the presence of Medfly populations in California in the 1990s. The lowest “combination” model illustrates a situation in which both establishment and reintroductions are occurring. A more-complex situation similar to this “combination model” appears consistent with the molecular data. The top two models were redrawn from [Carey \(1991\)](#).

Molecular methods were used in attempts to answer three questions: Are species-specific diagnostics useful for ecological studies and quarantine procedures? What is the colonization history and population structure of Medflies as they moved out of Africa and into California? And, is the Medfly permanently established in the Los Angeles basin?

#### 13.7.4.2 Species-Specific Diagnostics

An important component in dealing with invasion problems is to be able to rapidly and reliably identify the introduced species. Within the Diptera, the Tephritidae contains many important agricultural pests. Of the >4000 species of tephritids found in infested fruits, 250 are considered pests ([Armstrong et al. 1997](#)). Unfortunately, it is difficult to identify immature tephritids to species, so

rearing is required to obtain accurate identifications, which can cause a significant delay in a quarantine or eradication program.

Using RAPD-PCR, [Sonvico et al. \(1996\)](#) found that it is possible to discriminate between immature stages of Medflies and *Anastrepha fraterculus*. [Haymer et al. \(1994\)](#) identified unique repetitive DNA probes that they used in slot blots or squash blots to discriminate between eggs or larvae of three tephritid species. The squash-blot procedure used a nonradioactive hybridization and detection method, making it simple and rapid to carry out and potentially allowing rapid identification of infested fruits at the earliest stage during quarantine and eradication procedures.

AFLP-PCR was used by [Kakouli-Durate et al. \(2001\)](#) to discriminate between *C. capitata* and *C. rosa*. A species-specific repetitive marker was cloned and used as a probe for genomic dot-blot hybridizations; the probes were sequenced and primers were developed. AFLP-PCR products from *C. capitata* electrophoresed on polyacrylamide gel revealed diagnostic bands after silver staining ([Kakouli-Durate et al. 2001](#)). However, the probe produced faint bands with DNA from *Bactrocera cucurbitae* and *B. oleae*, suggesting that the repetitive DNA fragment exists in low-copy number in them. The number of “outgroup” tephritids tested was limited, so the probe may only be useful to discriminate among a few species. However, the authors did investigate the method’s sensitivity and found that, because the probe sequences are found in multiple copies in the genome, sufficient DNA could be extracted from wings or legs of *C. capitata*, yet yield positive results.

Ribosomal ITS1 polymorphisms were investigated in *C. capitata* and *C. rosa* to provide species-specific probes and to investigate the differences in size among different populations of each species ([Douglas and Haymer 2001](#)). Recall that insects have multiple ribosomal genes (rDNA) and that the noncoding ITS region tends to vary sufficiently to allow discrimination of lower-level taxonomic groups. ITS1 sequences were isolated from a *C. capitata* genomic library and sequenced; sequences from *C. capitata* and from *Drosophila melanogaster* were aligned to identify conserved sequences and primers were designed to amplify across the variable ITS1 regions. No obvious size variability in the PCR products was found among *C. capitata* from Hawaii, Guatemala, Spain, Greece, Costa Rica, and Peru, and sequencing of some of these ITS1 regions revealed only single nucleotide changes, with length variation ranging from 829 to 832 bp. Thus, ITS1 sequence data would not allow a researcher to discriminate among these populations, but they do allow *C. capitata* to be identified. The same primers amplified ITS1 sequences from two populations of *C. rosa* and produced products that were 717 bp and 930 bp long, which might be used to distinguish

*C. rosa* individuals from Kenya from those from South Africa (Douglas and Haymer 2001), if additional populations from each geographic region can be tested and the utility of the length variation is confirmed.

Armstrong et al. (1997) reported efforts to identify tephritids for quarantine purposes in New Zealand using 18S and 18S + ITS regions of ribosomal DNA. These were amplified from larval DNA by the PCR and 19 species in four genera were evaluated. Restriction analysis of the 18S product provided poor resolution, even at the generic level. Digestion of the 18S + ITS PCR product generated 13 diagnostic haplotypes using four restriction endonucleases. Six of 10 *Bactrocera* species could not be diagnosed separately with this method despite analyzing the effects of 22 restriction enzymes. However, all six species are considered high risk with respect to their likely establishment in New Zealand, so a diagnosis of suspicious larvae as *Bactrocera* would result in the same response by regulatory authorities.

The above-described studies provide an overview of some of the approaches that could be taken to use molecular methods to identify *C. capitata* immature stages and to discriminate *C. capitata* from other tephritids. However, additional research is required before these approaches could be used in specific quarantine procedures and none are useful for all objectives. For example, if a DNA probe is intended to identify *C. capitata* immatures when tephritid larvae are found in infested fruit in border inspections, then the probe should be validated against all potential fruit fly species that might be introduced (recall that  $\approx 250$  species are considered pests). However, it may not be relevant to identify the larvae to the species level; rather, it may be relevant only to confirm that the larvae are from one of the pest tephritid genera, because all would be quarantined pests. In addition, the question should be answered as to how often false positives and false negatives occur under real-world conditions with the test method? In some ecological studies, it might be relevant to discriminate between immatures of two, or a few, tephritid species, which would be much easier, because the test could be validated only against these species.

#### 13.7.4.3 Geographic Origin of Medfly Populations

Efforts to determine the geographic origin of Medfly populations in new environments were difficult. Studies were conducted using RAPD-PCR, enzyme analysis, nuclear introns, mitochondrial DNA, RFLP-PCR, and microsatellites. Some projects are described briefly to illustrate that each method has strengths and weaknesses in answering a specific question.

Regulatory agencies often want to know where a pest population came from because it might allow them to prevent future invasions if they know where to

invest their inspection and detection efforts. It also is relevant if sterile Medflies are to be released in an eradication program; potentially, sterile Medflies of one “type” might not mate with an invasive wild population if there are sufficient genetic differences in behavior, or other premating isolating mechanisms. With regard to the California Medfly eradication efforts, the origin or genetic makeup of the different Medfly invasions could answer the question as to whether the expensive eradication efforts were successful.

Malacrida et al. (1996) used enzyme electrophoresis to analyze genetic similarities among 11 tephritid species. A later study focused on tracking the colonization of *C. capitata* throughout the world using samples from 17 populations (Malacrida et al. 1998). Variability at 26 polymorphic enzyme loci revealed “that the geographical dispersal of medfly from its ancestral source area (East Africa) is associated with a great reduction in variability. The pattern of decreasing variability occurs at two regional levels: in the African-Mediterranean region where the differentiation is gradual, and in the Latin American-Pacific region where some ancestral variability is still present as a consequence of recent colonization” (Malacrida et al. 1998).

The molecular data confirm that the common name “Medfly” is inappropriate; because the ancestral home of *C. capitata* is Africa, it ought to be called “Africafly.” Malacrida et al. (1998) concluded that the “population genetic changes observed in the species range are consistent with both the chronology and the historical circuitous course of the medfly colonization process.” Thus, these molecular data are congruent with what is known about its movements out of Africa.

#### 13.7.4.4 Is the Medfly Established in California?

A critical question regarding the success of eradication in California is, “Is the Medfly permanently established in California?” The question was approached by assuming that independent introductions of Medflies from different geographic sources would result in populations with unique genetic markers; if each invasive population had different markers, it would be evidence that each invasion is independent. By contrast, if the markers found in the California populations during the different “invasions” were the same, the conclusion could be that it is more likely that a single Medfly population is established in California. However, an alternative explanation for Medflies having the same markers is that multiple invasions occurred from a particular geographic source, and it would be impossible to exclude this possibility.

Obtaining definitive data to discriminate between populations depends on having markers that are diagnostic. These markers need to be validated with

large samples of flies from different geographic regions to confirm that the differences detected in the preliminary screening hold up when larger samples of, potentially more-diverse, flies are sampled. The task of identifying appropriate genetic markers has engaged a number of researchers and considerable funds.

Early attempts to discriminate between different geographic populations of the Medfly used RAPD-PCR markers (Haymer and McInnis 1994), enzyme electrophoresis (Malacrida et al. 1996), and compared RAPD-PCR and enzyme electrophoresis data (Baruffi et al. 1995). As expected, RAPD-PCR revealed larger amounts of genetic variation than enzyme electrophoresis data (Baruffi et al. 1995). The complete mitochondrial genome of the Medfly was sequenced, and different populations were found to exhibit genetic differences that are potentially useful for developing diagnostic tools (Spanos et al. 2000).

A PCR-RFLP method used by He and Haymer (1999) compared variation in intron sequences of the *glucose-6-phosphate dehydrogenase* gene among different Medfly populations. Five alleles of this locus were found in 26 populations of *C. capitata* and two restriction enzymes were used in successive digestions of the PCR products to document genotypes and allele frequencies. This approach involved amplifying intron sequences from individuals from various populations by using primers designed from cDNA sequences in GenBank. PCR products were cloned and sequenced, and allelic variants and restriction-site changes were identified. The restriction-site data allowed He and Haymer (1999) to develop a diagnostic test that did not require sequencing the PCR products. The data were analyzed using principal coordinate analysis and **Analysis of MOlecular VAriation** (AMOVA) to quantify the distribution of genetic diversity in a hierarchical manner. For some of the invasive sites, populations that "are probably acting as sources of origin" were identified (He and Haymer 1999). Five alleles tended to be associated with populations from different geographic regions: A1 was most common and found in all populations surveyed, so was not informative. A2 and A3 were widespread in samples from Greece, but only one allele tended to be prevalent in other samples (A2 was prevalent in samples from Guatemala, Peru, Florida, and southern California; A3 was prevalent in samples from Argentina). Hawaiian populations showed substantial frequencies of A4, but A4 was rare in other populations.

He and Haymer (1999) concluded "the invasive population from northern California appears similar to populations from Argentina and Costa Rica. From southern California, three of the infestations (1992–1994) are clustered with populations from Guatemala, suggesting, "that Guatemala is a possible source of

origin of these flies." The 1997 southern California infestation is "well separated from either of the two previous groupings ... these results suggest it is not appropriate to group or depict all of them as homogeneous" (He and Haymer 1999). They concluded "the extreme separation of the Hawaiian populations from these California (and Florida) infestations also suggests that Hawaii can be considered a very unlikely source." Furthermore, "Samples from the California 1997 infestation are also well separated from all other populations, suggesting that none of the worldwide populations sampled here can be considered likely sources." Thus, "the multiple infestations detected within California in recent years are not likely to represent a single, homogenous population that is similar to the "established" populations seen in Guatemala or Argentina" (He and Haymer 1999). The authors noted that additional surveys of these markers in populations from other regions of the world are desirable to improve resolution of Medfly population relationships. They concluded that analysis of these alleles in ancestral African populations, where considerably more genetic variability occurs, is desirable.

Gomulski et al. (1998) evaluated variability in the size of the first intron in the *alcohol dehydrogenase* gene to assess 16 populations from five geographical regions: Africa, the Mediterranean basin, Latin America, Hawaii, and Australia. PCR primers were developed that spanned the intron between exons 1 and 2. PCR product sizes varied from 1400 bp to 3450 bp and were grouped into four categories: short, medium, long, and very long. Most variants were found only in the African populations and only a few migrated from Africa with the colonizing populations. The results obtained were congruent with those obtained by analyzing allozyme variation and showed a gradual and large reduction in intron variability. Gomulski et al. (1998) concluded that drift, bottleneck effects, and migration were important in explaining the observed intron-size variability.

Villablanca et al (1998) analyzed multiple nuclear-gene introns in an effort to provide sufficient information to resolve the origins of the Medfly populations in California and Hawaii despite the reduction in variability expected due to founder effects and genetic bottlenecks. They chose to work with multiple introns because they concluded that mitochondrial DNA is poorly suited to studies of invasions unless the invading population is large or grows rapidly; mtDNA is subject to strong genetic drift due to its maternal and haploid mode of inheritance. They also argued that RAPD-PCR data are difficult to interpret and may lack repeatability; microsatellites are effective, but require a long period of development for each new taxon. Their review of previous work of molecular analyses of Medflies indicated, "little genetic variation has been uncovered within invading populations with both allozymes ... and mtDNA."

Villablanca et al. (1998) "found a wealth of genetic variability within invading populations." Introns evolve more quickly than the protein-coding regions of a gene and are expected to retain variation due to their diploid and biparental inheritance. The intron-sequence variation can be subjected to "phylogenetic analysis, cladistic analysis of gene flow, as well as standard population genetic and coalescence analysis of alleles" (Palumbi 1996, Roderick 1996). Medfly populations in Africa, California, Hawaii, Brazil, and Greece were evaluated using primers constructed based on Medfly sequence data from the literature; introns from four single-copy nuclear genes were amplified that have conserved positions across species. Single-copy genes were used to avoid analysis of nonspecific PCR products that could occur from multiple-gene copies or pseudogenes. The four loci were: muscle-specific actin intron 1, chorion s36 intron 1, vitellogenin 1 gamma intron 2, and Cu/Zn superoxide dismutase (SOD) intron 1. The PCR products were cloned and sequenced and were found to be specific to the targeted loci, with no evidence for pseudogenes. The sequence data were analyzed to eliminate sequences in which errors were incorporated by *Taq* polymerase, which has a misincorporation rate of approximately one per 1000 bases. Villablanca et al. (1998) eliminated these erroneous sequences by sequencing between one and three clones per individual and then identifying and removing "singletons." Singletons are variability that occurs in only one sequence of an alignment and, although "not all singletons are PCR errors, but considering them to be so results in a conservative measure of allelic diversity." The remaining sequences were analyzed phylogenetically using Templeton's network method, which allows reconstruction of phylogenies from potentially recombinant DNA fragments (Villablanca et al. 1998).

Villablanca et al. (1998) interpreted the phylogenetic analysis of four intron sequences as follows: "The phylogeny of alleles shows that there is no phylogeographic structuring at the population level. Few alleles are shared between African and invading [California, Hawaii, Brazil and Greece] populations . . . . The phylogenies of alleles, similarly, do not provide evidence that any invading population is monophyletic . . . . Although the phylogeny of alleles is not useful for phylogeographic analysis in this case, it is still essential in that it demonstrates that alleles might be shared among populations simply because all populations are ultimately derived from Africa and not because they share a common invasion history." The authors pointed out that the next step is to "sample populations more thoroughly and test for population subdivision." **Phylogeography** is the study of relationships among genotypes (phylogeny of alleles or haplotypes) from one or more populations that are examined relative to their geographical location (Roderick 1996).

Davies et al. (1999a) used the same intron loci to distinguish between "alternative hypotheses concerning the source of medfly infestations in California." In their study, intron sequences from Villablanca et al. (1998) were used, as well as newly obtained intron sequences from Medfly samples in California, Costa Rica, Guatemala, Mexico, Brazil, Peru, Greece, Hawaii, and Africa. A total of 237 sequences were obtained for the four loci in 74 individuals. The data from all Medflies in California were treated as a "single population for the purposes of statistical analysis. Under the null hypothesis that there is a resident medfly population in California, we assume that these flies, captured in the same geographic area, represent a single biological population (Carey 1991)." To assess whether the recent outbreak was due to a new invasion, the authors focused on a single fly (B-96) captured in southern California (Burbank) in 1996. AMOVA produced indices of population subdivision analogous to standard F statistics. Another program, TFPGA, was used to calculate the average theta. Yet another program, IMMANC, was used to carry out an assignment test. Davies et al. (1999a) concluded that, "Because the B-96 genotype was included in the Californian population (and not the potential source) when estimating the "resident" allele frequencies, the test is conservative with respect to the null hypothesis that B-96 is a resident—in this case of California." Davies et al. (1999a) concluded that the single B-96 Medfly studied was "less likely ( $\alpha < 0.05$ ) to be a resident of California than an immigrant from no less than four potential sources: Costa Rica, Guatemala, Mexico, or Peru." Finally, Davies et al. (1999a) concluded,

*"More work is clearly needed to explore the phylogenetic consequences of invasions and a better understanding of invasion genetic patterns will provide a deeper insight into the ecological and evolutionary processes that underlie bioinvasions. It is important to consider ... that invasions often involve a hierarchy of events, the totality of which might be termed a metainvasion. The metainvasion begins with a primary invasion, when a species first colonizes a new area from its ancestral source. Subsequently, secondary and tertiary invasions arise as the newly established populations themselves seed new areas. The genetic changes that result from these events are complex and phylogenetic analyses may be informative at some levels but not others. A primary invasion of the medfly occurred from Africa to the Mediterranean. The invasion of Latin America may be another primary invasion, direct from Africa, or a secondary invasion from the Mediterranean. Californian medfly invasions thus represent secondary or tertiary events in the global medfly metainvasion. Indeed, California may be subject to repeat invasions that could superimpose on one another."*

Although by 1998, powerful genetic tools had been brought to bear on the California Medfly colonization question and the statistical methods for

analyzing molecular data were more sophisticated, no firm conclusions could be made. [Villablanca et al. \(1998\)](#) suggested that the origin of Californian Medfly infestations might be determined through the use of microsatellites or single-strand conformation polymorphism analysis. They noted,

*"Due to its economic significance, the medfly infestation has become a model system for the study of contemporary bioinvasions and has several important lessons for other cases where limited funding is likely to restrict the amount of research effort. The genetic analysis of new bioinvasions should begin with mtDNA and allozymes; however, highly variable nuclear regions, such as introns, should also be considered. Multilocus genotyping provides a rapid method of determining the origin of invasions, whether using nonsequencing methods of screening intron variation and/or other types of markers."*

Thirty Medfly microsatellites for *C. capitata* were developed as tools for population analysis by [Bonizzoni et al. \(2000\)](#). In addition, 11 microsatellite loci were identified by RAPD-PCR and random genomic sequencing. Two additional loci were identified in GenBank, for a total of 43 microsatellites. Ten of these microsatellite sequences were used to analyze 122 Medflies from six populations (Kenya, Reunion, Madeira, South Italy, Greece, and Peru). The results obtained were "consistent with results obtained from allozyme and single-copy DNA studies with respect to the historically documented expansion of the medfly" ([Bonizzoni et al. 2000](#)). As with the allozyme data ([Malacrida et al. 1998](#)), polymorphisms decreased as flies moved from tropical Africa to the Mediterranean basin and to South America.

Microsatellite analysis was used by [Bonizzoni et al. \(2001\)](#) to resolve: was there one established population or many invasive populations in California? The 10 previously characterized microsatellite loci were used to compare 109 Medflies captured in California between 1992 and 1998 with 242 Medflies from Hawaii, Guatemala, El Salvador, Ecuador, Brazil, Argentina, and Peru, using between six and 30 flies per sample site. Their data analysis used a method that accounts for heterogeneity in the size of samples to estimate allelic richness. The frequency of each allele per locus, the observed heterozygosity and deviations from Hardy-Weinberg expectations were computed using several methods ([Bonizzoni et al. 2001](#)). Genetic divergence between individuals, and within and between populations was estimated in terms of shared bands between individuals. Relationships between populations were given in dendograms obtained from the dissimilarity index and Nei's unbiased genetic distance ( $D_A$ ). Trees were constructed using the neighbor-joining method of [Felsenstein \(1993\)](#), and bootstrap values for the tree were obtained using the "gene frequency" option within the program SEQBOOT. The Kenyan sample was used as the out-group because it is the

most differentiated in the Medfly range. An estimation of the probability that the California Medflies are immigrants from South America, Central America, or Hawaii was determined using the IMMANC test ([Bonizzoni et al. 2001](#)).

[Bonizzoni et al. \(2001\)](#) had three main conclusions: "(i) among the Latin American and Pacific samples, the Guatemalan flies are most closely related genetically to the California flies, according to the majority of the tests applied; (ii) the Californian infestations are structured with the San Diego infestation being the most differentiated; and (iii) the fact that flies captured between 1992 and 1997 in the Los Angeles basin appear to be genetically related, supports the hypothesis that an endemic population has been formed in this area." Furthermore, "Hawaiian flies show relatively low similarities with Californian flies," which indicates Hawaii is not the source of the Californian Medflies tested, as found by [He and Haymer \(1999\)](#).

[Bonizzoni et al. \(2001\)](#) concluded that the situation could be more complex than expected: "Within at least some of the Los Angeles basin samples, there is considerable evidence for genetic homogeneity. Based on this, the possibility of an endemic population in California cannot be excluded. It is entirely possible that independent infestations of this pest from the same geographical region, overlaid on an existing endemic population, have acted together to create this unique situation."

#### *13.7.4.5 The End?*

A California Department of Food and Agriculture press release in June 2001 summarized eradication costs since 1975 in California: "more than \$256 million in state and federal funds have been spent eradicating infestations of the pest, primarily in Southern California and the Bay Area." The report concluded that the 5-year test program to stop new infestations by releasing sterile Medflies in the Los Angeles basin was "enormously successful" because "Between 1987 and 1994, an average of 7.5 Medfly infestations were discovered each year in California. Since the preventative Release Program began in 1996, there has been just one infestation for the entire five-year period."

Over the life of the 5-year "preventative program" that began in 1996, nearly 75 billion sterile Medflies were released from airplanes over >2000 square miles of the Los Angeles basin. This preventative program cost California ≈US\$7.4 million annually and the U.S. government an additional US\$7.4 million (CDFA Legislative Report, March 2000). The report highlighted the fact that, "Since 1994, California's Medfly strategy in Southern California has shifted from a reactive approach, focused on detecting and eradicating early infestations, to a proactive approach that emphasizes preventative measures."

In 2010, the Medfly controversy erupted again ([Chen 2010](#)). [Carey \(2010\)](#) reasserted that the Medfly was established in California and stated, "To the extent possible, scientific questions and policy questions should not be conflated. Whereas the primary responsibility of scientists is to pursue scientific truths, the principal responsibility of policy makers is to formulate policy on the basis of these truths." [Carey \(2010\)](#) reiterated that repeated invasions were unlikely the source of Medflies in California. [Liebhold et al. \(2010\)](#) responded to Carey by reviewing an earlier paper in which they evaluated 190,101 interceptions of alien insects at U.S. airports between 1984 and 2000, with 43% of the interceptions consisting of fruit infested with insects (especially mangoes). Tephritidae consisted of 33% of the interceptions. However, "While there is little doubt that Medflies have consistently arrived in California at relatively high rates in fruit associated with baggage, the question of whether the species has persisted for many years or whether it has been repeatedly introduced and then eradicated remains, in our opinion, unresolved" ([Liebhold et al. 2010](#)).

[Diamantidis et al. \(2008\)](#) compared biological traits of wild Medflies obtained from Kenya, Greece, Portugal, Brazil, Guatemala, and Hawaii under identical laboratory conditions. The F<sub>1</sub> generation of each population was studied and showed differences in life span, age-specific oviposition, and differences in sexual behavior, although Medflies from different areas were able to interbreed. However, the divergences in life history and behavioral traits suggest considerable evolutionary changes have occurred as this species has colonized new geographic areas.

Using a demographic system model, [Gutierrez and Ponti \(2011\)](#) evaluated the effects of temperature on Medfly's potential distribution in Arizona, California, and Italy. Their model predicts that the area of potential favorability is the south coastal region of California, especially the urban areas of the Los Angeles basin. Other regions of California might be suitable for some years, but not continuously over several years and their analysis "suggests that temperatures in much of CA are simply outside of medfly's thermal envelope. ... and "if the fly is established in CA, its permanence is tenuous. In contrast, Medfly is established in Italy, and our model predicts the wide geographic favorability observed there" ([Gutierrez and Ponti 2011](#)).

Classical biological control has been suggested as a possible management tool for the Medfly in California ([Headrick and Goeden 1996](#)). However, a biological-control program typically would not be conducted unless there is a good probability that the Medfly is established (which would create economic and political problems in California).

#### 13.7.4.6 Some Lessons Learned

The Medfly case study illustrates several points: molecular tools vary in their sensitivity, ease of use, cost, and time to develop. Despite differences in the methods used, the various researchers agreed on some key findings: Medfly populations in the Los Angeles area were unlikely to have come from Hawaii; the population present in the Los Angeles area may be due either to the presence of an established population or to multiple infestations from the same source, or both.

Molecular markers of six Medflies from the San Diego population were clearly different from the markers found in Medflies collected in the Los Angeles area. The San Diego Medflies sampled most closely resembled Medflies from Hawaii. Analysis of the infestation area by entomologists correlated the San Diego infestation with a family that had just returned from a trip to Hawaii. Thus, combining molecular and ecological data may provide more information than relying on a single approach (A. Malacrida, personal communication).

It should be clear that insect population-genetic structure can be quite complex; simply assuming that molecular markers will allow an unequivocal conclusion as to the population's geographic origin may be unrealistic, especially if primary, secondary, and tertiary invasions have occurred ([Figure 13.4](#)). Genetic variability and structuring is dependent upon events in the population's history, including bottlenecks, drift, selection and hybridization. The Medfly case study shows that increasingly refined molecular methods are available, and the statistical and other analyses used to reach conclusions are becoming more sophisticated.

#### 13.7.5 Plant Defenses to Insect Herbivory

Plants face a variety of biotic (bacteria, fungi, insects, and other herbivores) and abiotic (drought, heat, salinity, and UV damage) stresses in their environment ([Strauss and Agrawal 1999](#)). In response, plants evolved both constitutive and inducible defenses that have a genetic basis ([Zheng and Dicke 2008](#)). Direct defenses include synthesis of secondary plant compounds that affect attraction and deterrence of insects, or inhibit insect growth and development. Induced defenses include proteinase inhibitors and polyphenol oxidases that inhibit the digestive enzymes or reduce the nutritive content of the plant. Indirect defenses also include releases of volatiles that signal the location of pests on infested plants to their natural enemies. There are many genes involved in plant-defense mechanisms, making it difficult to analyze their role using traditional genetic methods.

DNA microarrays can be used to obtain significant advances in our understanding of plant defenses against insect herbivores ([Schenk et al. 2000](#), Baldwin

et al. 2001, Reymond 2001, Kempema et al. 2007). Analysis of  $\approx$ 7000 *Arabidopsis* genes (which is 25–30% of this plant's genome) suggested that  $\approx$ 300 of the 7000 genes (4.3%) evaluated are involved in defense (Maleck et al. 2000). Comparison of gene transcript profiles after plants were exposed to different stresses revealed that some genes have overlapping roles in defense; thus, microarray analyses can elucidate how plants respond to multiple stressors.

Microarrays of *Arabidopsis* expressed sequence tags (ESTs) have become available from several core laboratory sources since this species' genome was sequenced (Reymond 2001). The microarrays allow simultaneous hybridization of probes to an array of immobilized DNA fragments that correspond to a specific gene. After scanning the microarray with a laser scanner, the signal for each fragment reflects the abundance of the corresponding messenger RNA in the sample (Maleck et al. 2000). For example, Kempema et al. (2007) found that feeding by nymphs of type B of the silverleaf whitefly, *Bemisia tabaci*, up-regulated 700 transcripts and down-regulated 556. The responses were qualitatively and quantitatively different from those induced by chewing insects or by aphids. Phloem-feeding insects such as *Bemisia* feed for long periods but cause minimal tissue damage.

Many plants produce volatile organic compounds (VOCs) in response to feeding damage caused by herbivores (Dicke 1999). These VOCs may influence neighboring plants to respond rapidly to wounding, or to insect or plant pathogen attack. Some VOCs attract predators and parasitoids to the insect herbivore and DNA microarrays will allow researchers to analyze which genes are involved in this type of plant defense. DNA microarrays allow researchers "to determine the extent to which VOCs can elicit defense-related transcripts in neighboring plants" (Arimura et al. 2000a,b).

It appears that a complex network of interdependent signaling pathways convey molecular messages in *Arabidopsis* that identify the type of pest, which allows the plant to mount an appropriate response (Reymond 2001). Integrating the information on plant responses, obtaining an understanding of the communications that take place between the different defense response pathways, and obtaining a complete catalog of response genes should be achievable. Thus, the global analysis of plant gene expression on microarrays and the complete sequencing of the *Arabidopsis* and rice genomes could revolutionize the analysis of insect–plant interactions. Microarray research will bring together ecologists, molecular biologists, and plant scientists (Maleck et al. 2000, Baldwin et al. 2001). Some caution is needed, however, because conducting and analyzing microarray experiments requires careful consideration of experimental design and statistical analysis (Kerr and Churchill 2001a,b, Quackenbush 2001).

### 13.7.6 Origins of Insect Populations

It is often the case that pest managers and regulatory agencies want to know the origin of an invasive pest. Knowing the origin may allow the regulatory authorities to develop improved inspection methods for cargo and improved quarantine procedures. Scientists may want to predict the distribution of the pest in the new environment based on the origin of the pest population. If classical biological control is to be conducted, it is desirable to know the origin of the pest so that the most effective natural enemies can be identified for evaluation, importation, release, and establishment. The following examples illustrate the approaches taken to study the origins of several insect populations.

#### 13.7.6.1 Emerald Ash Borer

The emerald ash borer (EAB), *Agrilus planipennis* (Buprestidae), is a wood-boring pest of ash that was discovered to have invaded North America in 2002. It spread from its original infestations in Michigan and Ontario, Canada rapidly, so that as of 2011 it can be found in 15 states in the United States and in two provinces in Canada. Larvae of the EAB have killed tens of millions of ash trees in North America by feeding under the outer bark on phloem and outer sapwood. Adults exit the trees within 1 or 2 years to attack new trees.

The EAB is known to occur in Asia, including China, Japan, Korea, Mongolia, Taiwan, and the Russian Far East. Beetles from 17 areas in Asia and from 7 localities in North America were analyzed using mitochondrial DNA sequences, AFLP fingerprints, and two microsatellite loci ([Bray et al. 2011](#)). The COI mitochondrial sequences indicated that all North American beetles could have come from China or South Korea. The 108 AFLP fingerprints indicated that the North American populations most resembled beetles in Hebei and Tianjin City and were most likely the result of a single introduction from China. The genetic diversity of the North American beetles was reduced compared to the beetles studied from Asia. [Bray et al. \(2011\)](#) concluded that the EAB probably invaded North America only once, even though the first sites were geographically distant (or that the invasions were from the same source because it was not possible to distinguish the Ontario population from the other North American populations tested). However, the authors suggest additional analysis of beetles from other locations in Asia is desirable. The results obtained have allowed searches for natural enemies of the EAB in Asia, and releases of three parasitoid species have been made.

#### 13.7.6.2 Domesticated Populations of *Aedes aegypti*

The mosquito *Aedes aegypti* is a vector of dengue and yellow fever viruses and is thought to have originated in sub-Saharan Africa, breeding in tree holes in

the forest and feeding on wild animals (Brown et al. 2011). Populations of this mosquito, especially those that have invaded areas outside Africa, have developed the ability to breed in human habitats and feed on humans. Brown et al. (2011) hypothesized that all “domesticated” populations are genetically related and represent a single domestication event. The alternative hypothesis was that association with humans developed multiple times independently within *Ae. aegypti* populations. Brown et al. (2011) sampled 24 populations by using 12 polymorphic microsatellite loci and found two distinct genetic clusters, one of which included all populations outside of Africa that were “domesticated,” and the other that included both domestic and wild populations in Africa. The authors concluded that domestication occurred in Africa independently from domestication elsewhere. The estimates of genetic diversity supported the hypothesis that *Ae. aegypti* originated in Africa.

#### 13.7.6.3 Bumblebees from Greenhouses Invade Wild Conspecific Populations

Bumblebees (*Bombus* species) are used to pollinate greenhouse crops, especially tomatoes and peppers, because they are efficient and easy to handle. The colonies are mass reared and sold by commercial companies and are transported around the world. Concerns have been raised about the effects of such large-scale releases on endemic populations of the same species. Other concerns include fears that the commercial bees may displace native species, or that they may spread diseases into wild populations. The effect of releasing bumblebees into greenhouses in a geographic area where conspecific endemic bees are present was analyzed by comparing microsatellite DNA from greenhouse and wild bumblebee populations (*Bombus terrestris*) in Poland (Kraus et al. 2010). The results indicated that there is “strong genetic introgression from the sampled greenhouse populations into the adjacent populations.” However, wild populations more distant from the greenhouses had less genetic admixture than those populations close to the greenhouses. The impact of such introgression on the endemic bees is unknown.

#### 13.7.6.4 The Tobacco Aphid in the New World

The tobacco aphid, *Myzus persicae nicotianae*, is an invasive pest in Chile (Zepeda-Paulo et al. 2010). Seven microsatellite loci were sampled from American and European countries to identify the potential source population(s). The data suggest that the tobacco aphid came from Europe and/or Asia to North America and then to South America. The authors present evidence that there were multiple introductions into North America and that a strong bottleneck occurred when the aphids were introduced into South America.

### 13.7.6.5 Populations of *Aedes taeniorhynchus* in the Galapagos Islands: Rapid Evolution of an Invasive Species

*Aedes taeniorhynchus* is thought to have colonized the Galapagos Islands ≈200,000 years ago. The islands in the Galapagos typically have coastal habitats that are rocky and contain mangroves but have a very dry inland zone. Highlands on the larger islands, such as Santa Cruz, obtain more rain and have a lush vegetation (Bataille et al. 2010). Bataille et al. (2010) used microsatellite data to compare coastal and highland mosquito populations and to determine whether gene flow is occurring between the two habitats. They discovered that the coastal and highland populations were highly differentiated from each other all year round, with some gene flow detected only during periods of increased precipitation. The results support the hypothesis that selection, arising from ecological differences between the two habitats, is driving adaptation and divergence in *A. taeniorhynchus*. Indeed, the authors speculated that the two populations on Santa Cruz might represent an example of incipient speciation.

### 13.7.6.6 Industrial Melanism in Peppered Moths

Industrial melanism in the moth *Biston betularia* in nineteenth century Great Britain is famous for illustrating a rapid evolutionary response to an altered environment. The ancestral form of the peppered moth (*typica*) is white with dark speckles but, after the Industrial Revolution, a darker form (*carbonaria*) became more common due to natural selection in sooty environments. The *carbonaria* form displaced the light-colored moths in the polluted woodlands of Europe, reportedly due to selective predation by birds on the lighter moths when located on sooty trees (Cook et al. 2012, Luiggi 2012). The black form of the moth was known to be due to a single-locus dominant allele, but the biochemical basis of this phenotype remained unknown until van't Hof et al. (2011) mapped the melanism to a 200-kb region of a chromosome and further narrowed it to a single nucleotide polymorphism (SNP) marker. Analyses of moths from the field indicated that the same SNP marker is the basis of all *carbonaria* moths in the United Kingdom, and all are derived from a single ancestral haplotype that “coincides with major wing-patterning loci in other lepidopteran systems, suggesting the existence of basal color-patterning regulators in this region.”

## 13.8 Applied Pest Management

Intraspecific variation and genetic changes in both pest and beneficial arthropods influence pest management strategies and tactics in different ways. The fact that pest and natural enemy populations change genetically has been

discussed often. Molecular-genetic techniques allow for rapid assessment of intraspecific genetic variation, changes in host preferences, insecticide-resistance levels, and pest and natural enemy biotypes significant in biological control of pest arthropods and weeds. The following examples illustrate some of the diversity of methods and goals used when molecular tools are applied for applied pest-management programs. However, molecular methods are not a panacea and [Curtis \(2002\)](#) recently raised a relevant question (see Section 13.8.5).

### 13.8.1 Monitoring Biotypes, Species, and Cryptic Species

Real-time PCR assays based on TaqMan technology and RFLP analyses of the COI mitochondrial gene were used to distinguish the B and Q biotypes (cryptic species) of *Bemisia tabaci* ([Papayiannis et al. 2009](#)). The RFLP assay and real-time PCR assay were equivalent except that the real-time PCR was more sensitive and more rapid.

Distinguishing early immature stages of the screwworms *Cochliomyia hominivorax* and *C. macellaria* (Calliphoridae) is difficult, but important, in the eradication and exclusion program to prevent reintroduction of screwworms into the United States, Mexico, and Central America ([Alamalakala et al. 2009](#)). AFLP-PCR was evaluated for strain and species identifications and all ten primers tested resulted in AFLP banding patterns. The 10-primer combinations generated 10–35 bands per individual and resulted in 52 useful bands, with seven bands found in both species. 22 bands were specific for *C. macellaria* populations, 10 were specific for *C. hominivorax*, and 13 diagnostic for different populations of *C. hominivorax*. The authors concluded, "AFLP-PCR is a good tool for differentiating species and has tremendous potential for studies of intraspecific genetic variation." Diagnostic bands could be isolated from the gels and sequenced to develop specific primers for future screens.

Although barcoding of the COI gene is often useful in identifying species, it does not always work. [Whitworth et al. \(2007\)](#) found that it could not identify species of the blowfly genus *Protocalliphora*. Assignment of unknown individuals to species was impossible for 60% of the species.

Barcode was useful in identifying bees (Apoidea) in Nova Scotia, Canada ([Sheffield et al. 2009](#)). A survey of the fauna of this region not only identified the known bee species but also revealed two undescribed species, one each in the genera *Ceratina* and *Andrena*. It allowed identification of the two sexes in dimorphic species so that synonymies could be resolved.

[Lefort et al. \(2012\)](#) wanted to identify scarab larvae without killing them, so that they could be studied in the laboratory. They discovered it was possible to

use frass (excreta) or larval exuviae as sources of DNA for multiplex PCR for identifying the larvae of the two species. There was sufficient DNA present in the frass as long as 7 days after it was excreted to produce a PCR product.

### 13.8.2 Monitoring Vectors of Disease

Monitoring mosquitoes that are vectors of disease for the presence of the pathogen is an important component in managing diseases affecting human and animal populations. Various approaches have been tested, including clinical diagnoses of symptoms or detection of the pathogen in the vectors. Sentinel animals are used to determine whether a disease outbreak is impending. Virus surveillance of the mosquitoes is based on isolating the virus from mosquitoes collected in traps. Hall-Mendelin et al. (2010) developed a novel method to detect viruses in mosquitoes that involves attracting the mosquitoes to a carbon dioxide-baited trap where they can feed on honey-soaked nucleic-acid preservation cards in the trap. The mosquitos expectorate the virus into the honey while feeding, and the cards can then be analyzed by reverse-transcriptase PCR. Hall-Mendelin et al. (2010) detected Ross River and Barmah Forest viruses in these cards taken from traps in Australia and found the viral RNA was preserved for at least 7 days. The use of the cards reduced handling of mosquitoes or processing them to extract viral RNA.

A heteroduplex PCR reaction was developed by Lee et al. (2002) to identify what bird species *Culex tarsalis* mosquitoes had fed on among 16 possible hosts. The assay used primers amplifying a fragment of the cytochrome B gene from the 16 bird species and this gene could be detected 7 days after feeding by the mosquitoes. Information on host range is important in understanding the relationship between mosquitoes and their normal hosts and determining how pathogens get transmitted to humans or other animals. Epidemics in humans or other animals often can only occur when “bridge species” of mosquitoes transmit the virus (or other pathogen) to them because many mosquitoes are restricted in their host range and rarely bite humans. Thus, knowing the natural source of the pathogen can be important in monitoring for and predicting potential epidemics.

### 13.8.3 Pesticide Resistances and Pest Management

Pest managers would like to identify pesticide resistances in pests and improve their ability to design novel pesticides (Perry et al. 2011). Perry et al. (2011) argue that resistance studies in *Drosophila melanogaster* could elucidate biochemical and genetic mechanisms underlying resistance in many insects because genes and metabolic activities in all insects are conserved and the advances in

genomics, metabolomics, and structural biology could result in improved pesticides. Insecticides could be designed against target sites in pests, but allow the survival of biological control agents. However, “The *D. melanogaster* system does not offer a panacea. Research into this model and the major insect pests needs to proceed in parallel, but it is certain that research in *D. melanogaster* will accelerate progress in improving the way in which insecticides are used to control insect pests” (Perry et al. 2011).

Genome analyses of pest species will also aid in understanding resistance to pesticides. Strode et al. (2008) evaluated the detoxification genes in the mosquito *Aedes aegypti* and discovered 235 members of the cytochrome P450, glutathione transferase, and carboxy/cholinesterase families, which is 58% and 36% more genes compared with *D. melanogaster* and the mosquito *Anopheles gambiae*, respectively. The authors used genomic data to construct a microarray containing unique oligos for these putative resistance genes and compared their expression level in resistant and susceptible strains. Candidate genes were identified in two families, the CYP9 P450s and the Epsilon GSTs. The “Detox Chip” could allow pest managers to monitor for resistance to pesticides in dengue and yellow fever control programs.

#### **13.8.4 Monitoring Pest-Population Biology**

Diapause is an important component of the biology of pest species and understanding when and where it occurs, and when diapause development is completed is important to pest managers. Diapause in 870 individuals of the Colorado potato beetle, *Leptinotarsa decemlineata*, was evaluated under field conditions for three years using multiplex PCR to monitor the expression of five genes (Yocum et al. 2010). The authors discovered that the experimental results on diapause attributes obtained in the laboratory were not directly applicable to some of the beetles in the field and concluded that these differences could be important in managing the pest. Differences in the environmental cues obtained by the beetles in the laboratory and field indicated, “laboratory investigations are necessary but are not sufficient to reflect the complexity of field diapause regulation.”

#### **13.8.5 The “So What?” Test**

Curtis (2002) reviewed three areas in medical entomology where molecular methods had been applied. These included identifying complexes of sibling malaria-vector species, evaluating insecticide resistance in vectors of malaria, and developing nonsusceptibility to pathogens in mosquitoes using transgenic methods. Curtis (2002) questioned the number of instances in which “molecular

taxonomic markers for these characteristics add to what can be measured simply and directly?" and noted "There are very few instances where control programmes are guided by data on sibling species compositions of vector populations."

Likewise, [Curtis \(2002\)](#) noted that "Much emphasis is placed on studies of biochemical or molecular resistance mechanisms, but it was painstaking and direct field testing, rather than these molecular or biochemical studies, that revealed "... resistance to pyrethroid insecticides in malaria vectors in South Africa. [Curtis \(2002\)](#) concluded, "that molecular methods in medical entomology should not be an end in themselves. They should be adopted only after careful investigation shows that they can pass the "so what?" test, that is, **could they solve real problems that are not already solvable by simpler and cheaper methods.**"

There are genuine benefits to molecular methods, although they are not panaceas. Because the field is yet young, it is difficult to know *a priori* whether a particular molecular tool will aid in solving a specific problem. However, it is appropriate to ask whether there are more-rapid and less-expensive methods that can resolve the question.

## Relevant Journals

*Biochemical Systematics and Ecology*, Elsevier Press

*Conservation Genetics*, Springer Publishers

*Molecular Ecology*, Blackwell Publ.

*Molecular Ecology Notes*, now *Molecular Ecology Resources*, Blackwell Publ.

## References Cited

- Abate, T., (1993). Is the invading medfly eligible for California residency. *Bioscience* **43**: 3–6.
- Agusti, N., De Vicente, M.C., and Gabarra, R., (1999). Development of sequence characterized region (SCAR) markers of *Helicoverpa armigera*: a new polymerase chain reaction-based technique for predator gut analysis. *Mol. Ecol.* **8**: 1474–1476.
- Alamalakala, L., Skoda, S.R., and Foster, J.E., (2009). Amplified fragment length polymorphism used for inter- and intraspecific differentiation of screwworms (Diptera: Calliphoridae). *Bull. Entomol. Res.* **99**: 139–149.
- Aquadro, C.F., Noon, W.A., and Begun, D.J., (1992). RFLP analysis using heterologous probes. In: Hoelzel, A.R., Ed., *Molecular Genetic Analysis of Populations. A Practical Approach*, IRL Press, Oxford, UK, pp. 115–158.
- Archak, S., Meduri, E., Kumar, P.S., and Nagaraju, J., (2007). InSatDb: a microsatellite database of fully sequenced insect genomes. *Nucleic Acids Res.* **35**: D36–D39.
- Arias, R.S., Blanco, C.A., Portilla, M., Snodgrass, G.L., and Scheffler, B.E., (2011). First microsatellites from *Spodoptera frugiperda* (Lepidoptera: Noctuidae) and their potential use for population genetics. *Ann. Entomol. Soc. Am.* **104**: 576–587.
- Arimura, G., Tashiro, K., Kuhara, S., Nishioka, T., Ozawa, R., and Takabayashi, J., (2000a). Gene responses in bean leaves induced by herbivory and by herbivore-induced volatiles. *Biochem. Biophys. Res. Commun.* **277**: 305–310.

- Arimura, G., Ozawa, R., Shimoda, T., Nishioka, T., Boland, W., and Takabyashi, J., (2000b). Herbivory-induced volatiles elicit defence genes in lima bean leaves. *Nature* **406**: 512–515.
- Armstrong, K.F., Cameron, C.M., and Frampton, E.R., (1997). Fruit fly (Diptera: Tephritidae) species identification: a rapid molecular diagnostic technique for quarantine application. *Bull. Entomol. Res.* **87**: 111–118.
- Arnheim, N., White, T., and Rainey, W.E., (1990). Application of PCR: organismal and population biology. *Bioscience* **40**: 174–182.
- Ashley, M.V., and Dow, B.D., (1994). The use of microsatellite analysis in population biology: background, methods and potential applications. In: Schierwaer, B., Stret, B., Wagner, G.P., and DeSalle, R., Eds., *Molecular Ecology and Evolution Approaches and Applications*, Birkhauser Verlag, Basel, Switzerland, pp. 187–201.
- Atkinson, L., and Adams, E.S., (1997). Double-strand conformation polymorphism (DSCP) analysis of the mitochondrial control region generates highly variable markers for population studies in a social insect. *Insect Mol. Biol.* **6**: 369–376.
- Attimonelli, M., Altamura, N., Benne, R., Brennicke, A., Cooper, J.M., D'Ellia, D., and de Montalvo, A., et al. (2000). MitBASE: a comprehensive and integrated mitochondrial DNA database. The present status. *Nucleic Acids Res.* **28**: 148–152.
- Baldwin, I.T., Halitschke, R., Kessler, A., and Schittko, U., (2001). Merging molecular and ecological approaches in plant-insect interactions. *Curr. Opin. Plant Biol.* **4**: 351–358.
- Balloux, F., and Lugon-Moulin, N., (2002). The estimation of population differentiation with microsatellite markers. *Mol. Ecol.* **11**: 155–165.
- Baruffi, Damiani, L.G., Guglielmino, C.R., Bandi, C., Malacrida, A.R., and Gasperi, G., (1995). Polymorphism within and between populations of *Ceratitis capitata*: comparison between RAPD and multilocus enzyme electrophoresis data. *Heredity* **74**: 425–537.
- Bataille, A., Cunningham, A.A., Cruz, M., Cedeno, V., and Goodman, S.J., (2010). Seasonal effects and fine-scale population dynamics of *Aedes taeniorhynchus*, a major disease vector in the Galapagos Islands. *Mol. Ecol.* **19**: 4491–4504.
- Behura, S.K., (2006). Molecular marker systems in insects: current trends and future avenues. *Mol. Ecol.* **15**: 3087–3113.
- Berticat, C., Dubois, M.P., Marquine, M., Chevillon, C., and Raymond, M., (2000). A molecular test to identify resistance alleles at the amplified esterase locus in the mosquito *Culex pipiens*. *Pest Management Sci.* **56**: 727–731.
- Black, W.C., and Vontas, J.G., (2007). Affordable assays for genotyping single nucleotide polymorphisms in insects. *Insect Mol. Biol.* **16**: 377–387.
- Black, W.C., DuTeau, N.M., Puterka, G.J., Nechols, J.R., and Pettorini, J.M., (1992). Use of the random amplified polymorphic DNA polymerase chain reaction (RAPD-PCR) to detect DNA polymorphisms in aphids (Homoptera: Aphididae). *Bull. Entomol. Res.* **82**: 151–159.
- Black, W.C., Baer, C.F., Antolin, M.F., and DuTeau, N.M., (2001). Population genomics: genome-wide sampling of insect populations. *Annu. Rev. Entomol.* **46**: 441–469.
- Blacket, M.J., Semeraro, L., and Malipatil, M.B., (2012). Barcoding Queensland fruit flies (*Bactrocera tryoni*): impediments and improvements. *Mol. Ecol. Res.* **12**: 428–436.
- Bohonak, A.J., (1999). Dispersal, gene flow, and population structure. *Quart. Rev. Biol.* **74**: 21–45.
- Bonizzoni, M., Malacrida, A.R., Guglielmino, C.R., Gomulski, L.M., Gasperi, G., and Zheng, L., (2000). Microsatellite polymorphism in the Mediterranean fruit fly. *Ceratitis capitata*. *Insect Mol. Biol.* **9**: 251–261.
- Bonizzoni, M., Zheng, L., Guglielmino, C.R., Haymer, D.S., Gasperi, G., Gomulski, L.M., and Malacrida, A.R., (2001). Microsatellite analysis of medfly bioinfestations in California. *Mol. Ecol.* **10**: 2524–2525.
- Bossart, J.L., and Pashley-Powell, D., (1998). Genetic estimates of population structure and gene flow: limitations, lessons and new directions. *Trends Ecol. Evol.* **13**: 202–206.

- Bray, A.M., Bauer, L.S., Poland, T.M., Haack, R.Z., Cognato, A.I., and Smith, J.J., (2011). Genetic analysis of emerald ash borer (*Agrilus planipennis* Fairmaire) populations in Asia and North America. *Biol. Invasions* **13**: 2869–2887.
- Brookfield, J.F.Y., (1992). DNA fingerprinting in clonal organisms. *Mol. Ecol.* **1**: 21–26.
- Brown, J.E., McBride, C.S., Johnson, P., Ritchie, S., Paupy, C., and Bossin, H., et al. (2011). Worldwide patterns of genetic differentiation imply multiple 'domestications' of *Aedes aegypti*, a major vector of human diseases. *Proc. R. Soc. Lond. B* **278**: 2446–2454.
- Bruford, M.W., Hanotte, O., Brookfield, J.F.Y., and Burke, T., (1992). Single-locus and multilocus DNA fingerprinting. In: Hoelzel, A.R., Ed., *Molecular Genetic Analysis of Populations. A Practical Approach*, IRL Press, Oxford, UK, pp. 225–269.
- Brun-Barale, A., Bouvier, J.C., Paruon, D., Berge, J.B., and Benoit, S., (2005). Involvement of a sodium channel mutation in pyrethroid resistance in *Cydia pomonella* L, and development of a diagnostic test. *Pest Manag. Sci.* **61**: 549–554.
- Buckley, T.R., Simon, C., Flook, P.K., and Misof, B., (2000). Secondary structure and conserved motifs of the frequently sequenced domains IV and V of the insect mitochondrial large subunit rRNA gene. *Insect Mol. Biol.* **9**: 565–580.
- Burke, T., (1989). DNA fingerprinting and other methods for the study of mating success. *Trends Ecol. Evol.* **4**: 139–144.
- California Dept. of Food and Agriculture, (2000). Food and Agric. Report to the Legislature. *Mediterranean Fruit Fly Preventative Release Program*: 18.
- California Department of Food and Agriculture. (2002). The Mediterranean Fruit Fly Fact Sheet. <[www.cdfa.ca.gov/plant/factsheets/MedFlyFactSheet.pdf](http://www.cdfa.ca.gov/plant/factsheets/MedFlyFactSheet.pdf)>
- Carey, J.R., (1991). Establishment of the Mediterranean fruit fly in California. *Science* **253**: 1369–1373.
- Carey, J.R., (1996a). The incipient Mediterranean fruit fly population in California Implications for invasion biology? *Ecology* **77**: 1690–1697.
- Carey, J.R., (1996b). The future of the Mediterranean fruit fly *Ceratitis capitata* invasion of California: a predictive framework. *Biol. Conserv.* **78**: 35–50.
- Carey, J.R., (2010). The Mediterranean fruit fly (*Ceratitis capitata*). *Am. Entomol. Fall*: 158–163.
- Casquet, J., Thebaud, C., and Gillespie, R.G., (2012). Chelex without boiling, a rapid and easy technique to obtain stable amplifiable DNA from small amounts of ethanol-stored spiders. *Mol. Ecol. Res.* **12**: 136–141.
- Cavalli-Sforza, L.L., (1998). The DNA revolution in population genetics. *Trends Genet.* **14**: 60–65.
- Chandra, S.B.C., Vik, J.L., and Kapatral, V., (2006). Comparative insect mitochondrial genomes: differences despite conserved genome synteny. *Afr. J. Biotechnol.* **5**: 1308–1318.
- Chen, I., (2010). From Medfly to moth: Raising a buzz of dissent. *Science* **327**: 134–136.
- Clark, P.L., Molina-Ochoa, J., Martinelli, S., Skoda, S.R., Isenhour, D.J., Lee, D.J., Krumm, J.T., and Foster, J.E., (2007). Population variation of the fall armyworm, *Spodoptera frugiperda*, in the Western Hemisphere. *J. Insect Sci.* **7**: 5.
- Coates, B.S., Bayles, D.O., Wanner, K.W., Robertson, H.M., Hellmich, R.L., and Sappington, T.W., (2011). The application and performance of single nucleotide polymorphism markers for population genetic analyses of Lepidoptera. *Front. Genet.* **2** (38): 1–10.
- Cook, L.M., Grant, B.S., Saccheri, I.J., and Mallet, J., (2012). Selective bird predation on the peppered moth: the last experiment of Michael Majerus. *Biol. Lett.* **8** (4): 609–612.
- Cooley, J.R., Simon, C., Marshall, D.C., Slon, K., and Ehrhardt, C., (2001). Allochronic speciation, secondary contact, and reproductive character displacement in periodical cicadas (Hemiptera: *Magicicada* spp.): genetic, morphological and behavioural evidence. *Mol. Ecol.* **10**: 661–671.
- Cooley, J.R., Simon, C., and Marshall, D.C., (2003). Temporal separation and speciation in periodical cicadas. *Bioscience* **53**: 151–157.
- Cornuet, J.M., Piry, S., Luikart, G., Estoup, A., and Solignac, M., (1999). New methods employing multilocus genotypes to select or exclude populations as origins of individuals. *Genetics* **153**: 1989–2000.

- Crozier, R.H., (1977). Evolutionary genetics of the Hymenoptera. *Annu. Rev. Entomol.* **22**: 263–288.
- Cullings, K.W., and Vogler, D.R., (1998). A 5.8S nuclear ribosomal RNA gene sequence database: applications to ecology and evolution. *Mol. Ecol.* **7**: 919–923.
- Curtis, C., (2002). Molecular medical entomology and the ‘so what?’ test. *Trends Ecol. Evol.* **17**: 102.
- Davies, N., Villalba, F.X., and Roderick, G.K., (1999a). Bioinvasions of the Medfly *Ceratitis capitata*: source estimation using DNA sequences at multiple intron loci. *Genetics* **153**: 351–360.
- Davies, N., Villalba, F.X., and Roderick, G.K., (1999b). Determining the source of individuals: multilocus genotyping in nonequilibrium population genetics. *Trends Ecol. Evol.* **14**: 17–21.
- Diamantidis, A.D., Carey, J.R., and Papadopoulos, N.T., (2008). Life-history evolution of an invasive tephritid. *J. Appl. Entomol.* **132**: 695–705.
- Dick, M., D.M., Bridge, W.C., Wheeler and R., DeSalle. (1993). Collection and storage of invertebrate samples. pp. 51–65. In: *Methods in Enzymology*. Vol. **224**, Academic Press.
- Dicke, M., (1999). Evolution of induced indirect defence of plants. In: Tollrian, R., and Harvell, C.D., Eds., *The Ecology and Evolution of Inducible Defences*, Princeton University Press, Princeton, NJ, pp. 62–88.
- Doolittle, R.F., Ed., (1990). *Molecular Evolution: Computer Analysis of Protein and Nucleic Acid Sequences*, Vol. **183**, *Methods in Enzymology*, Academic Press, NY.
- Douglas, L.J., and Haymer, D.S., (2001). Ribosomal ITS1 polymorphisms in *Ceratitis capitata* and *Ceratitis rosa* (Diptera: Tephritidae). *Ann. Entomol. Soc. Am.* **94**: 726–731.
- Dowling, T.E., Moritz, C., and Palmer, J.D., (1990). Nucleic acids II: restriction site analysis. In: Hillis, D.M., Moritz, C., Eds., *Molecular Systematics*, Sinauer, Sunderland, MA, pp. 250–317.
- Edwards, O.W., and Hoy, M.A., (1993). Polymorphism in two parasitoids detected using random amplified polymorphic DNA (RAPD) PCR. *Biol. Control* **3**: 243–257.
- Edwards, O.W., and Hoy, M.A., (1995). Random amplified polymorphic DNA markers to monitor laboratory-selected, pesticide-resistant *Trioxys pallidus* (Hymenoptera: Aphidiidae) after release into three California walnut orchards. *Environ. Entomol.* **24**: 487–496.
- Ellegren, H., (2004). Microsatellites: simple sequences with complex evolution. *Nat. Rev. Genet.* **5**: 435–445.
- Erlich, H.A., Ed., (1989). *PCR Technology, Principles and Applications of DNA Amplification*, Stockton Press, New York.
- Estoup, A.M., Solignac, M., Harry, and Cornuet, J.M., (1993). Characterization of (GT)<sub>n</sub> and (CT)<sub>n</sub> microsatellites in two insect species: *Apis mellifera* and *Bombus terrestris*. *Nucleic Acids Res.* **21**: 1427–1431.
- Estoup, A., and Angers, B., (1998). Microsatellites and minisatellites for molecular ecology: theoretical and empirical considerations. In: Carvalho, G.R., Ed., *Advances in Molecular Ecology*, IOS Press, Amsterdam, The Netherlands, pp. 55–86.
- Feder, J.L., Chilcote, C.A., and Bush, G.L., (1988). Genetic differentiation between sympatric host races of the apple maggot fly *Rhagoletis pomonella*. *Nature* **336**: 61–64.
- Felsenstein, J., (1993). *PhyliP (Phylogeny Inference Package)*. v. 3.Sc. Distributed by the author. Department of Genetics, University of Washington, Seattle, WA.
- Ferraris, J.D., and Palumbi, S.R., (1996). *Molecular Zoology. Advances, Strategies, and Protocols*. Wiley-Liss, New York.
- ffrench-Constant, R.H., Anthony, N., Aronstein, K., Rocheleau, T., and Stilwell, G., (2000). Cyclodiene insecticide resistance: from molecular to population genetics. *Annu. Rev. Entomol.* **48**: 449–466.
- FLYBase Consortium, (2002). The FLYBase database of the *Drosophila* genome projects and community literature. *Nucleic Acids Res.* **30**: 106–108.
- Foltan, P., Sheppard, S., Konvicka, M., and Symondson, W.O.C., (2005). The significance of facultative scavenging in generalist predator nutrition: detecting decayed prey in the guts of predators using PCR. *Mol. Ecol.* **14**: 4147–4158.
- Freeland, J.R., Kirk, H., and Petersen, S.D., (2011). *Molecular Ecology*, 2nd Ed., Wiley-Blackwell, Oxford, UK.

- Gibson, G., (2002). Microarrays in ecology and evolution: a preview. *Mol. Ecol.* **11**: 17–24.
- Gibson, G., and Muse, S.V., (2002). *A Primer of Genome Science*. Sinauer, Sunderland, MA.
- Gilbert, D.G., (2002). euGenes: a eukaryote genome information system. *Nucleic Acids Res.* **30**: 145–148.
- Gilad, Y., Pritchard, J.K., and Thornton, K., (2009). Characterizing natural variation using next-generation sequencing technologies. *Trends Genet.* **25**: 471–863.
- Gomulski, L.M., Bourtzis, K., Brogna, S., Morandi, P.A., Bonvicini, C., and Sebastiani, F., et al. (1998). Intron size polymorphism of the Adh(1) gene parallels the worldwide colonization history of the Mediterranean fruit fly. *Ceratitis capitata*. *Mol. Ecol.* **7**: 1729–1741.
- Goodman, S.J., (1997). R<sub>ST</sub> Calc: a collection of computer programs for calculating estimates of genetic differentiation from microsatellite data and determining their significance. *Mol. Ecol.* **6**: 881–885.
- Goodnight, K.F., and Queller, D.C., (1999). Computer software for performing likelihood tests of pedigree relationship using genetic markers. *Mol. Ecol.* **8**: 1231–1234.
- GrandeSbarau, J., de Azeredo-Espin, A.M.L., and Lessinger, A.C., (2005). Conservation and versatility of a new set of primers for long-PCR amplification of complete insect mitochondrial genomes based on *Haematobia irritans* mtDNA sequences. *Mol. Ecol. Notes* **5**: 885–887.
- Greenstone, M.H., Weber, D.C., Coudron, T.A., Payto, M. E., and Hu, J.S., (2012). Removing DNA contamination from arthropod predators destined for molecular gut-content analysis. *Mol. Ecol. Res.* **12**: 404–469.
- Gribskov, M., and Devereux, J., Eds., (1991). *Sequence Analysis Primer*, Stockton Press, NY.
- Gutierrez, A.P., and Ponti, L., (2011). Assessing the invasive potential of the Mediterranean fruit fly in California and Italy. *Biol. Invasions* **13** (12): 2661–2676.
- Hadrys, H., Balick, M., and Schierwater, B., (1992). Applications of random amplified polymorphic DNA (RAPD) in molecular ecology. *Mol. Ecol.* **1**: 55–63.
- Hadrys, H., Schierwater, B., Dellaporta, S.L., DeSalle, R., and Buss, L.W., (1993). Determination of paternity in dragonflies by random amplified polymorphic DNA fingerprinting. *Mol. Ecol.* **2**: 79–87.
- Hagerman, P.J., (1990). Sequence-directed curvature of DNA. *Annu. Rev. Biochem.* **59**: 755–781.
- Hall-Mendelin, S., Ritchie, S.A., Johansen, C.A., Zborowski, P., Cortis, G., and Dandridge, S., et al. (2010). Exploiting mosquito sugar feeding to detect mosquito-borne pathogens. *Proc. Natl. Acad. Sci. USA* **107**: 11255–11259.
- Hamm, C.A., Aggarwal, D., and Landis, D.A., (2010). Evaluating the impact of non-lethal DNA sampling on two butterflies, *Vanessa cardui* and *Satyrodes eurydice*. *J. Insect Conserv.* **14**: 11–18.
- Hartl, D.L., (1981). *A Primer of Population Genetics*. Sinauer, Sunderland, MA.
- Hartl, D.L., and Clark, A.G., (1997). *Principles of Population Genetics*, Third Ed., Sinauer, Sunderland, MA.
- Haymer, D.S., (1994). Arbitrary (RAPD) primer sequences used in insect studies. *Insect Mol. Biol.* **3**: 191–194.
- Haymer, D.S., and McInnis, D.O., (1994). Resolution of populations of the Mediterranean fruit fly at the DNA level using random primers for the polymerase chain reaction. *Genome* **37**: 244–248.
- Haymer, D.S., Tanaka, T., and Teramae, C., (1994). DNA probes can be used to discriminate between tephritid species at all stages of the life cycle (Diptera: Tephritidae). *J. Econ. Entomol.* **87**: 741–746.
- He, M., and Haymer, D.S., (1999). Genetic relationships of populations and the origins of new infestations of the Mediterranean fruit fly. *Mol. Ecol.* **8**: 1247–1257.
- Headrick, D.H., and Goeden, R.D., (1996). Issues concerning the eradication or establishment and biological control of the Mediterranean fruit fly, *Ceratitis capitata* (Wiedemann) (Diptera: Tephritidae), in California. *Biol. Control* **6**: 412–421.

- Helyar, S.J., Hemmer-Hansen, J., Bekkevold, D., Taylor, M.I., Ogden, R., and Limborg, M.T., et al. (2011). Application of SNPs for population genetics of nonmodel organisms: new opportunities and challenges. *Mol. Ecol. Res.* **11** (Suppl. 1): 123–136.
- Hewitt, G.M., (2001). Speciation, hybrid zones and phylogeography—or seeing genes in space and time. *Mol. Ecol.* **10**: 537–549.
- Hickson, R.E., Simon, C., Cooper, A., Spicer, G.S., Sullivan, J., and Penny, D., (1996). Conserved sequence motifs, alignment, and secondary structure for the third domain of animal 12S rRNA. *Mol. Biol. Evol.* **13**: 150–160.
- Hoelzel, A.R. and G.A. Dover (1991). Statistical interpretation of variation and genetic distance. Pp. 31–46 In: *Molecular Genetic Ecology*. IRL Press, Oxford, UK.
- Hoelzel, A.R., and Bancroft, D.R., (1992). Statistical analysis of variation. In: Hoelzel, A.R., Ed., *Molecular Genetic Analysis of Populations. A Practical Approach*, IRL Press, Oxford, UK, pp. 297–305.
- Hoelzel, A.R., and Green, A., (1992). Analysis of population-level variation by sequencing PCR-amplified DNA. In: Hoelzel, A.R., Ed., *Molecular Genetic Analysis of Populations. A Practical Approach*, IRL Press, Oxford, UK, pp. 159–187.
- Hoffmann, A.A., Sgro, C.M., and Lawler, S.H., (1995). Ecological population genetics: the interface between genes and the environment. *Annu. Rev. Genet.* **29**: 349–370.
- Hoogendorn, M., and Heimpel, G.E., (2001). PCR-based gut content analysis of insect predators: using ribosomal ITS-1 fragments from prey to estimate predation frequency. *Mol. Ecol.* **10**: 2059–2067.
- Howard, D.J., and Berlocher, S.H., Eds., (1998). *Endless Forms: Species and Speciation*, Oxford University Press, New York.
- Hoy, M.A., Jeyaprakash, A., and Nguyen, R., (2001). Long PCR is a sensitive method for detecting *Liberobacter asiaticum* in parasitoids undergoing risk assessment in quarantine. *Biol. Control* **22**: 278–287.
- Hoy, M.A., Jeyaprakash, A., Clarke-Harris, D., and Rhodes, L., (2007). Molecular and field analysis of the fortuitous establishment of *Lipolexis oregmae* (Hymenoptera: Aphidiidae) in Jamaica as a natural enemy of the brown citrus aphid. *Biocontrol Sci. Technol.* **17**: 473–482.
- Huffaker, C.B., and Gutierrez, A.P., Eds., (1999). *Ecological Entomology*, Second edition, J. Wiley & Sons, New York.
- Innis, M.A., Gelfand, D.H., Ninsky, J.J., and White, T.J., Eds., (1990). *PCR Protocols, A Guide to Methods and Applications*, Academic, San Diego, CA.
- Jeffreys, A.J., (1987). Highly variable minisatellites and DNA fingerprints. *Biochem. Soc. Trans.* **15**: 309–317.
- Jeyaprakash, A., and Hoy, M.A., (2000). Long PCR improves *Wolbachia* DNA amplification: wsp sequences found in 76% of 63 arthropod species. *Insect Mol. Biol.* **9**: 393–405.
- Jinbo, U., Kato, T., and Ito, M., (2011). Current progress in DNA barcoding and future implications for entomology. *Entomol. Sci.* **14**: 107–124.
- Juen, A., and Traugott, M., (2005). Detecting predation and scavenging by DNA gut-content analysis: a case study using a soil insect predator-prey system. *Oecologia* **142**: 344–352.
- Kakouli-Durate, T., Casey, D.G., and Burnell, A.M., (2001). Development of a diagnostic DNA probe for the fruit flies *Ceratitis capitata* and *Ceratitis rosa* (Diptera: Tephritidae) using amplified fragment-length polymorphism. *J. Econ. Entomol.* **94**: 989–997.
- Kambhampati, S., Black, W.C., and Rai, K.S., (1991). Geographic origin of the US and Brazilian *Aedes albopictus* inferred from allozyme analysis. *Heredity* **67**: 85–94.
- Kambhampati, S., Black, W.C., and Rai, K.S., (1992). A RAPD-PCR based method for the identification and differentiation of mosquito populations: techniques and statistical analyses. *J. Med. Entomol.* **29**: 939–945.
- Karl, S.A., and Avise, J.C., (1993). PCR-based assays of Mendelian polymorphisms from anonymous single-copy nuclear DNA: techniques and applications for population genetics. *Mol. Biol. Evol.* **10**: 342–361.

- Kashi, Y., King, D., and Soller, M., (1997). Simple sequence repeats as a source of quantitative genetic variation. *Trends Genet.* **13**: 74–78.
- Kellenberger, E., (1994). Genetic ecology: a new interdisciplinary science, fundamental for evolution, biodiversity and biosafety evaluations. *Experientia* **50**: 429–437.
- Kempema, L.A., Cui, X., Holzer, F.M., and Walling, L.L., (2007). *Arabidopsis* transcriptome changes in response to phloem-feeding silverleaf whitefly nymphs. Similarities and distinctions in responses to aphids. *Plant Physiol.* **143**: 849–865.
- Kerr, M.K., and Churchill, G.A., (2001a). Bootstrapping cluster analysis: assessing the reliability of conclusions from microarray experiments. *Proc. Natl. Acad. Sci. USA* **98**: 8961–8965.
- Kerr, M.K., and Churchill, G.A., (2001b). Experimental design for gene expression microarrays. *Biostatistics* **2**: 183–201.
- King, R.A., Read, D.S., Traugott, M., and Symondson, W.O.C., (2008). Molecular analysis of predation: a review of best practice for DNA-based approaches. *Mol. Ecol.* **17**: 947–963.
- Kirby, L.T., (1990). *DNA Fingerprinting. An Introduction*. Stockton Press, New York.
- Kraus, F.B., Szentgyorgyi, H., Rozej, E., Rhode, M., Moron, D., Woyciechowski, M., and Moritz, R.F.A., (2010). Greenhouse bumblebees (*Bombus terrestris*) spread their genes into the wild. *Conserv. Genet.* **12**: 187–192.
- Lambert, D.M., (1995). The new science of molecular ecology. *N. Z. J. Ecol.* **9**: 93–96.
- Lanave, C., Licciulli, F., De Robertis, M., Marolla, A., and Attimonelli, M., (2002). Update of AMmtDB: a database of multi-aligned Metazoa mitochondrial DNA sequences. *Nucleic Acids Res.* **30**: 174–175.
- Landry, B.S., Dextraze, L., and Boivin, G., (1993). Random amplified polymorphic DNA markers for DNA fingerprinting and genetic variability assessment of minute parasitic wasp species (Hymenoptera: Mymaridae and Trichogrammatidae) used in biological control programs of phytophagous insects. *Genome* **36**: 580–587.
- Lee, J.H., Hassan, H., Hill, G., Cupp, E.W., Higazi, T.B., and Mitchell, C.J., et al. (2002). Identification of mosquito avian-derived blood meals by polymerase chain reaction-heteroduplex analysis. *Am. J. Trop. Med. Hyg.* **66**: 599–604.
- Lee, W., Park, J., Choi, J., Jung, K., Park, B., and Kim, D., et al. (2009). IMGD: an integrated platform supporting comparative genomics and phylogenetics of insect mitochondrial genomes. *BMC Genomics* **10**: 148.
- Lefort, M.C., Boyer, S., Worner, S.P., and Armstrong, K., (2012). Noninvasive molecular methods to identify live scarab larvae: an example of sympatric pest and nonpest species in New Zealand. *Mol. Ecol. Res.* **12**: 389–395.
- Lester, L.J., and Selander, R.K., (1979). Population genetics of haplodiploid insects. *Genetics* **92**: 1329–1345.
- Lewter, J.A., and Szalanski, A.L., (2007). Molecular identification of the fall armyworm, *Spodoptera frugiperda* (J.E. Smith) (Lepidoptera: Noctuidae) using PCR-RFLP. *J. Agric. Urban Entomol.* **24**: 51–57.
- Liebhold, A.M., Work, T.T., McCullough, D.G., and Cavey, J.F., (2010). Response to Carey (2010). *Am. Entomol. Fall*: 184–185.
- Loxdale, H.D., and Luschai, G., (1998). Molecular markers in entomology. *Bull. Entomol. Res.* **88**: 577–600.
- Lu, Y.J., Adang, M.J., Isenhour, D.J., and Kochert, G.D., (1992). RFLP analysis of genetic variation in North American populations of the fall armyworm moth *Spodoptera frugiperda* (Lepidoptera: Noctuidae). *Mol. Ecol.* **1**: 199–208.
- Lu, Y., Kochert, G.D., Eisenhour, D.J., and Adang, M.J., (1994). Molecular characterization of a strain-specific repeated DNA sequence in the fall armyworm, *Spodoptera frugiperda* (Lepidoptera: Noctuidae). *Insect Mol. Biol.* **3**: 123–130.
- Luiggi, C., (2012). Moth man. *Scientist* **26** (5): 21–22.

- Luikart, G., and England, P.R., (1999). Statistical analysis of microsatellite DNA data. *Trends Ecol. Evol.* **14**: 253–256.
- Lunt, D.H., Whipple, L.E., and Hyman, B.C., (1998). Mitochondrial DNA variable number tandem repeats (VNTRs): utility and problems in molecular ecology. *Mol. Ecol.* **7**: 1441–1455.
- Lynch, M., and Crease, T.J., (1990). The analysis of population survey data on DNA sequence variation. *Mol. Biol. Evol.* **7**: 377–394.
- MacPherson, J.M., Eckstein, P.E., Scoles, G.J., and Gajadhar, A.A., (1993). Variability of the random amplified polymorphic DNA assay among thermal cyclers, and effects of primer and DNA concentration. *Mol. Cell. Probes* **7**: 293–299.
- Malacrida, A.R., Guglielmino, C.R., Adamo, P.D., Torti, C., Marinoni, F., and Gasperi, G., (1996). Allozyme divergence and phylogenetic relationships among species of tephritid flies. *Heredity* **76**: 592–602.
- Malacrida, A.R., Marinoni, F., Torti, C., Gomulski, L.M., Sebastiani, F., Vonvuccini, C., Gasperi, G., and Guglielmino, C.R., (1998). Genetic aspects of the worldwide colonization process of *Ceratitis capitata*. *Genetics* **89**: 501–507.
- Maleck, K., Levine, A., Eulgem, T., Morgan, A., Schmid, J., and Lawton, K.A., et al. (2000). The transcriptome of *Arabidopsis* during systemic acquired resistance. *Nat. Genet.* **26**: 403–410.
- Marshall, D.C., (2001). Periodical cicada (Homoptera: Cicadidae) life-cycle variations, the historical emergence record, and the geographic stability of brood distributions. *Ann. Entomol. Soc. Am.* **94**: 386–399.
- Marshall, D.C., and Cooley, J.R., (2000). Reproductive character displacement and speciation in periodical cicadas, with description of a new species, 13-year *Magicicada neotredecim*. *Evolution* **54**: 1313–1325.
- Martin, A.P., and Simon, C., (1988). Anomalous distribution of nuclear and mitochondrial DNA markers in periodical cicadas. *Nature* **336**: 6196.
- May, B., (1992). Starch gel electrophoresis of allozymes. In: Hoelzel, A.R., Ed., *Molecular Genetic Analysis of Populations. A Practical Approach*, IRL Press, Oxford, UK, pp. 1–27.
- McMichael, M., and Prowell, D.P., (1999). Differences in amplified fragment-length polymorphisms in fall armyworm (Lepidoptera: Noctuidae) host strains. *Ann. Entomol. Soc. Am.* **92**: 175–181.
- Menken, S.B.J., (1991). Does haplodiploidy explain reduced levels of genetic variability in Hymenoptera. *Proc. Exp. Appl. Entomol. N.E.V. Amsterdam* **2**: 172–178.
- Mita, K.M., Kasahara, S., Sasaki, Y., Nagayasu, T., and Yamada, 2004. The genome sequence of the silkworm, *Bombyx mori*. *DNA Res.* **11**: 27–35.
- Mitton, J.B., (1994). Molecular approaches to population biology. *Annu. Rev. Ecol. Syst.* **25**: 45–69.
- Moritz, C., Dowling, T.E., and Brown, W.M., (1987). Evolution of animal mitochondrial DNA: relevance for population biology and systematics. *Annu. Rev. Ecol. Syst.* **18**: 192–269.
- Moulton, M.J., Song, H., and Whiting, M.F., (2010). Assessing the effects of primer specificity on eliminating numt coamplification in DNA barcoding: a case study from Orthoptera (Arthropoda: Insecta). *Mol. Ecol. Res.* **10**: 615–627.
- Mount, D.W., (2004). *Bioinformatics. Sequence and Genome Analysis*, Second Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor Press, NY.
- Mueller, U.G., and Wolfenbarger, L.L., (1999). AFLP genotyping and fingerprinting. *Trends Ecol. Evol.* **14**: 389–393.
- Munoz-Torres, M., Reese, J.T., Childers, C.P., Bennett, A.K., Sundaram, J.P., and Childs, K.L., et al. (2011). Hymenoptera genome database: integrated community resources for insect species of the order Hymenoptera. *Nucleic Acids Res.* **39 (Suppl. 1)**: D658–D662.
- Murphy, R.W., Sites, J.W., Buth, D.G., and Hafler, C.H., (1990). Proteins I: Isozyme electrophoresis. In: Hillis, D.M., and Moritz, C., Eds., *Molecular Systematics*, Sinauer, Sunderland, MA, pp. 45–126.
- Myers, J.H., Simberloff, D., Kuris, A.M., and Carey, J.R., (2000). Eradication revisited: dealing with exotic species. *Trends Ecol. Evol.* **15**: 316–320.

- Nadeau, N.J., and Jiggins, C.D., (2010). A golden age for evolutionary genetics? Genomic studies of adaptation in natural populations. *Trends Genet.* **26**: 484–492.
- Nagoshi, R.N., and Meagher, R.L., (2008). Review of fall armyworm (Lepidoptera: Noctuidae) genetic complexity and migration. *Fla. Entomol.* **91**: 546–554.
- Nagoshi, R.N., Silvie, P., and Meagher, R.L., (2007). Comparison of haplotype frequencies differentiate fall armyworm (Lepidoptera: Noctuidae) corn-strain populations from Florida and Brazil. *J. Econ. Entomol.* **100**: 954–961.
- Nagoshi, R.N., Fleischer, S., and Meagher, R.L., (2009). Texas is the overwintering source of fall armyworm in central Pennsylvania: Implications for migration into the northeastern United States. *Environ. Entomol.* **38**: 1546–1554.
- Nariai, Y., Hayashi, S., Morita, S., Umemura, Y., Tainaka, K., Sota, T., Cooley, J.R., and Yoshimura, J., (2011). Life cycle replacement by gene introduction under an Allee effect in periodical cicadas. *PLoS ONE* **6** (4): e18347.
- Nei, M., (1972). Genetic distance between populations. *Am. Nat.* **106**: 283–292.
- Neve, G., and Meglecz, E., (2000). Microsatellite frequencies in different taxa. *Trends Ecol. Evol.* **15**: 376–377.
- Odum, E.P., (1971). *Fundamentals of Ecology*, 3rd Ed., W. B. Saunders, Philadelphia, PA.
- O'Rorke, R., Lavery, S., and Jeffs, A., (2012). PCR enrichment techniques to identify the diet of predators. *Mol. Ecol. Res.* **12**: 5–17.
- Packer, L., and Owen, R.E., (1992). Variable enzyme systems in the Hymenoptera. *Biochem. Syst. Ecol.* **20**: 1–7.
- Palumbi, S.R., (1996). Macrospatial genetic structure and speciation in taxa with high dispersal abilities. In: Ferraris, J., and Palumbi, S.R., Eds., *Molecular Zoology: Advances, Strategies, and Protocols*, Wiley, New York, pp. 101–108.
- Pamilo, P., Viljakainen, L., and Vihavainen, A., (2007). Exceptionally high density of NUMTs in the honeybee genome. *Mol. Biol. Evol.* **24**: 1340–1346.
- Papaiyannis, L.C., Brown, J.K., Seraphides, N.A., Hadjistylli, M., Ioannou, N., and Katis, N.I., (2009). A real-time PCR assay to differentiate the B and Q biotypes of the *Bemisia tabaci* complex in Cyprus. *Bull. Entomol. Res.* **99**: 573–582.
- Pashley, D.P., (1986). Host-associated genetic differentiation in fall armyworm (Lepidoptera: Noctuidae): a sibling species complex? *Ann. Entomol. Soc. Am.* **79**: 898–904.
- Pashley, D.P., (1988). Quantitative genetics, development, and physiological adaptation in host strains of fall armyworm. *Evolution* **42**: 93–102.
- Pashley, D.P., (1989). Host-associated differentiation in armyworms (Lepidoptera: Noctuidae): an allozymic and mitochondrial DNA perspective. In: Loscale, H.D., den Hollander, J., Eds., *Electrophoretic Studies on Agricultural Pests*, Clarendon Press, Oxford, UK, pp. 103–114.
- Pashley, D.P., McMichael, M., and Silvain, J.F., (2004). Multilocus genetic analysis of host use, introgression, and speciation in host strains of the fall armyworm (Lepidoptera: Noctuidae). *Ann. Entomol. Soc. Am.* **97**: 1034–1044.
- Pasteur, N., Pasteur, G., Bonhomme, F., Catalan, J., and Britton-Davidian, J., (1988). *Practical Isozyme Genetics*. Ellis Horwood, Chichester, UK.
- Penner, G.A., Bush, A., Wise, R., Kim, W., Domier, L., Kasha, K., Laroche, A., Scoles, G., Molar, S.J., and Fedak, G., (1993). Reproducibility of random amplified polymorphic DNA (RAPD) analysis among laboratories. *PCR Methods Appl.* **2**: 341–345.
- Perry, T., Batterham, P., and Daborn, P.J., (2011). The biology of insecticidal activity and resistance. *Insect Biochem. Mol. Biol.* **31**: 411–422.
- Persad, A.B., Jeyaprakash, A., and Hoy, M.A., (2004). High-fidelity PCR assay discriminates between immature *Lipolexis oregmae* and *Lysiphlebus testaceipes* (Hymenoptera: Aphidiidae) within their aphid hosts. *Fla. Entomol.* **87**: 18–24.
- Price, P.W., (1997). *Insect Ecology*, Third edition, John Wiley, New York.

- Price, P.W., Denno, R.F., Eubanks, M.D., Finke, D.L., and Kaplan, I., (2011). *Insect Ecology. Behavior, Populations and Communities*. Cambridge University Press, Cambridge, UK.
- Quackenbush, J., (2001). Computational analysis of microarray data. *Nat. Rev. Genetics* **2**: 418–427.
- Real, L.A., Ed., (1994). *Ecological Genetics*, Princeton University Press, Princeton, NJ.
- Reymond, P., (2001). DNA microarrays and plant defense. *Plant. Physiol. Biochem.* **39**: 313–321.
- Rieseberg, L.H., (1998). Molecular ecology of hybridization. In: Carvalho, G.R., Ed., *Advances in Molecular Ecology*, IOS Press, Amsterdam, The Netherlands, pp. 243–265.
- Roderick, G.K., (1996). Geographic structure of insect populations: gene flow, phylogeography and their uses. *Annu. Rev. Entomol.* **41**: 263–290.
- Roehrdanz, R.L., and DeGrugillier, M.E., (1998). Long sections of mitochondrial DNA amplified from fourteen orders of insects using conserved polymerase chain reaction primers. *Ann. Entomol. Soc. Am.* **91**: 771–778.
- Rougerie, R., Smith, M.A., Fernandez-Triana, J., Lopez-Vaamonde, C., Ratnasingham, S., and Hebert, P.D.N., (2011). Molecular analysis of parasitoid linkages (MAPL): gut contents of adult parasitoid wasps reveal larval hosts. *Mol. Ecol.* **20**: 179–186.
- Rowntree, J.K., Shuker, D.M., and Preziosi, R.F., (2011). Forward from the crossroads of ecology and evolution. *Phil. Trans. R. Soc. London. B* **366**: 1322–1328.
- Rozas, J., and Rozas, R., (1997). DnaSP version 2.0: a novel software package for extensive molecular population genetics analysis. *CABIOS* **13**: 307–311.
- Saad, F.A., Halliger, B., Muller, C.R., Roberts, R.G., and Danieli, G.A., (1994). Single base substitutions are detected by double strand conformation analysis. *Nucleic Acids Res.* **22**: 4352–4353.
- Schenk, P.M., Kazan, K., Wilson, I., Anderson, J.P., Richmond, T., and Somerville, S.C., et al. (2000). Coordinated plant defense responses in *Arabidopsis* revealed by microarray analysis. *Proc. Natl. Acad. Sci. USA* **97**: 11655–11660.
- Schnabel, A., Beerli, P., Estoup, A., and Hillis, D., (1998). A guide to software packages for data analysis in molecular ecology. In: Carvalho, G.R., Ed., *Advances in Molecular Ecology*, IOS Press, Amsterdam, The Netherlands, pp. 291–303.
- Schowalter, T.D., (2011). *Insect Ecology. An Ecosystem Approach*, 3rd Ed., Elsevier, Amsterdam, The Netherlands.
- Schwartz, M.K., Tallmon, D.A., and Luikart, G., (1998). Review of DNA-based census and effective population size estimators. *Anim. Conserv.* **1**: 293–299.
- Sheffield, C.S., Hebert, P.D.N., Kevan, P.G., and Packer, L., (2009). DNA barcoding a regional bee (Hymenoptera: Apoidea) fauna and its potential for ecological studies. *Mol. Ecol. Res.* **9** (Suppl. 1): 196–207.
- Sheppard, S.K., Bell, J., Sunderland, K.D., Fenlon, J., Skervin, D., and Symondson, W.O.C., (2005). Detection of secondary predation by PCR analyses of the gut contents of invertebrate generalist predators. *Mol. Ecol.* **14**: 4461–4468.
- Simon, C., Franke, A., and Martin, A., (1991). The polymerase chain reaction: DNA extraction and amplification. In: Hewitt, G. M., Johnston, A.W.B., Young, J.P.W., Eds., *Molecular Techniques in Taxonomy*, Springer-Verlag, Berlin, Germany, pp. 329–355.
- Simon, C., McIntosh, C., and Deniega, J., (1993). Standard restriction fragment length analysis of the mitochondrial genome is not sensitive enough for phylogenetic analysis or identification of 17-year periodical cicada broods (Hemiptera: Cicadidae): the potential for a new technique. *Ann. Entomol. Soc. Am.* **86**: 228–238.
- Simon, C., Frati, F., Beckenbach, A., Crespi, B., Liu, H., and Flook, P., (1994). Evolution, weighting and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Ann. Entomol. Soc. Am.* **87**: 651–701.
- Simon, C.J., Tang, S., Dalwadi, G., Staley, J., Deniega, and Unnasch, T.R., (2000). Genetic evidence for assortative mating between 13-year cicadas and sympatric “17-year cicadas with 13-year life cycles” provides support for allochronic speciation. *Evolution* **54**: 1326–1336.

- Slatkin, M., (1987). Gene flow and the geographic structure of a natural population. *Science* **236**: 787–792.
- Slatkin, M., and Barton, N.H., (1989). A comparison of three indirect methods for estimating average levels of gene flow. *Evolution* **43**: 1349–1368.
- Smith, M.A., Fernandez-Triana, J., Roughley, R., and Hevert, P.D.N., (2009). DNA barcode accumulation curves for understudied taxa and areas. *Mol. Ecol. Res.* **9** (Suppl. 1): 208–216.
- Softlinks (2007). Available at [www.stanford.edu/group/rosenberglab/adze.html](http://www.stanford.edu/group/rosenberglab/adze.html)
- Song, H., Buhay, J.E., Whiting, M.F., and Crandall, K.A., (2008). Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified. *Proc. Natl. Acad. Sci. USA* **105**: 13486–13491.
- Sonvico, A., Manso, F., and Quesada Allue, L.A., (1996). Discrimination between the immature stages of *Ceratitis capitata* and *Anastrepha fraterculus* (Diptera: Tephritidae) populations by random amplified polymorphic DNA polymerase chain reaction. *J. Econ. Entomol.* **89**: 1208–1212.
- Southwood, T.R.E., (1978). *Ecological Methods*, Second Edition, Halsted, New York.
- Southwood, T.R.E., and Henderson, P.A., (2000). *Ecological Methods*, Third Edition, Blackwell Sciences, Oxford, UK.
- Spanos, L., Koutroumbas, G., Kotsyfakis, M., and Louis, C., (2000). The mitochondrial genome of the Mediterranean fruit fly. *Insect Mol. Biol.* **9**: 139–144.
- Speight, M.R., Hunter, M.D., and Watt, A.D., (2008). *Ecology of Insects: Concepts and Applications*. Wiley-Blackwell, Oxford, UK.
- Spotter, A., Gupta, P., Nurnberg, G., Reinsch, N., and Bienefeld, K., (2012). Development of a 44 K SNP assay focusing on the analysis of a Varroa-specific defence behaviour in honey bees (*Apis mellifera carnica*). *Mol. Ecol. Res.* **12**: 323–332.
- Stephens, P.A., Sutherland, W.J., and Freckleton, R.P., (1999). What is the Allee effect? *Oikos* **87**: 184–190.
- Stinchcombe, J.R., and Hoekstra, H.E., (2008). Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. *Heredity* **100**: 158–170.
- Strauss, S.Y., and Agrawal, A.A., (1999). The ecology and evolution of plant tolerance to herbivory. *Trends Ecol. Evol.* **14**: 179–185.
- Strode, C., Wondji, C.S., David, J.P., Hawkes, N.J., Lumjuan, N., and Nelson, D.R., et al. (2008). Genomic analysis of detoxification genes in the mosquito *Aedes aegypti*. *Insect Biochem. Mol. Biol.* **38**: 113–123.
- Sunnucks, P., (2000). Efficient genetic markers for population biology. *Trends Ecol. Evol.* **15**: 199–203.
- Symondson, W.O.C., and Liddell, J.E., Eds., (1996). *The Ecology of Agricultural Pests. Biochemical Approaches*. Chapman & Hall, London, UK.
- Szymanski, M., Barciszewska, M.Z., Erdmann, V.A., and Barciszewski, J., (2002). 5S ribosomal RNA database. *Nucleic Acids Res.* **30**: 176–178.
- Tabachnick, W.J., and Black, W.C., (1995). Making a case for molecular population genetic studies of arthropod vectors. *Parasitol. Today* **11**: 27–29.
- Tang, J., and Unnasch, T.R., (1997). Heteroduplex analysis in medical entomology: a rapid and sensitive sequence-based tool for population and phylogenetic studies. *Parasitol. Today* **13**: 271–274.
- Taylor, M.F.J., McKechnie, S.W., Pierce, N., and Kreitman, M., (1993). The lepidopteran mitochondrial control region: structure and evolution. *Mol. Biol. Evol.* **10**: 1259–1272.
- Techen, N., Arias, R.S., Glynn, N.C., Pan, Z., Khan, I., and Scheffler, B.E., (2010). Optimized construction of microsatellite-enriched libraries. *Mol. Ecol. Res.* **10**: 5078–5515.
- Tegelstrom, H., (1992). Detection of mitochondrial DNA fragments. In: Hoelzel, A.R., Ed., *Molecular Genetic Analysis of Populations. A Practical Approach*, IRL Press, Oxford, UK, pp. 89–113.

- Templeton, A.R., (1998). Nested clade analyses of phylogeographic data: testing hypotheses about gene flow and population history. *Mol. Ecol.* **7**: 381–397.
- Thomas, M.A., and Klapoetke, R., (2004). Genomics for the ecological toolbox. *Trends Ecol. Evol.* **19**: 439–445.
- Tittiger, C., (2004). Functional genomics and insect chemical ecology. *J. Chem. Ecol.* **30**: 2335–2357.
- Toth, G., Gaspari, Z., and Jurka, J., (2000). Microsatellites in different eukaryotic genomes: survey and analysis. *Genome Res.* **10**: 967–981.
- Unigene Microsatellite Database ([veenuash.info/](http://veenuash.info/))
- U.S. Department of Agriculture, Animal and Plant Health Inspection Service (2002). The Mediterranean fruit fly. <<http://www.aphis.usda.gov/oa/pubs/fsmedfly.html>>
- van't Hof, A.E., Edmonds, N., Dalikova, M., Marec, F., and Saccheri, I.J., (2011). Industrial melanism in British peppered moths has a singular and recent mutational origin. *Science* **332**: 958–960.
- Villalba, F.X., Roderick, G.K., and Palumbi, S.R., (1998). Invasion genetics of the Mediterranean fruit fly: variation in multiple nuclear introns. *Mol. Ecol.* **7**: 547–560.
- Walker, A.M., and Hoy, M.A., (2003). Responses of *Lipolexis oregmae* (Hymenoptera: Aphidiidae) to different instars of *Toxoptera citricida* (Homoptera: Aphididae). *J. Econ. Entomol.* **96**: 1685–1692.
- Wang, J., and Caballero, A., (1999). Developments in predicting the effective size of subdivided populations. *Heredity* **82**: 212–226.
- Wang, R., Kafatos, F.C., and Zheng, L., (1999). Microsatellite markers and genotyping procedures for *Anopheles gambiae*. *Parasitol. Today* **15**: 33–37.
- Weber, D.C., and Lundgren, J.G., (2009). Detection of predation using qPCR: effect of prey quantity, elapsed time, chaser diet, and sample preservation on detectable quantity of prey DNA. *J. Insect Sci.* **9**: 41.
- Weir, B.S., (1990). *Genetic Data Analysis. Methods for Discrete Population Genetic Data*. Sinauer, Sunderland, MA.
- Weir, B.S., and Cockerham, C.C., (1984). Estimating F-statistics for the analysis of population structure. *Evolution* **38** (1): 358–1370.
- Weiss, P., (1950). Perspectives in the field of morphogenesis. *Quart. Rev. Biol.* **25**: 177–198.
- White, P.S., and Densmore, L.D., (1992). Mitochondrial DNA isolation. In: Hoelzel, A.R., Ed., *Molecular Genetic Analysis of Populations: A Practical Approach*, IRL Press, Oxford, UK, pp. 29–58.
- Whitfield, C.W., Band, M., Bonaldo, M.F., Kumar, C.G., and Liu, L., et al. (2002). Annotated expressed sequence tags and cDNA microarrays for studies of brain and behavior in the honey bee. *Genome Res.* **12**: 555–566.
- Whitworth, T.L., Dawson, R.D., Magalon, H., and Baudry, E., (2007). DNA barcoding cannot reliably identify species of the blowfly genus *Protocalliphora* (Diptera: Calliphoridae). *Proc. R. Soc. London B* **274**: 1731–1739.
- Wilson, R.J., Goodman, J.L., Strelets, V.B., and The FlyBase Consortium, 2008. FlyBase: integration and improvements to query tools. *Nucleic Acids. Res.* **36**: D588–D593.
- Wuyts, J., Van de Peer, Y., Windelsmans, T., and De Wachter, R., (2002). The European database on small subunit ribosomal RNA. *Nucleic Acids Res.* **30**: 183–185.
- Yocom, G.D., Rinehart, J.P., and Larson, M.L., (2010). Monitoring diapause development in the Colorado potato beetle, *Lepinotarsa decemlineata*, under field conditions using molecular biomarkers. *J. Insect Physiol.* **57**: 645–652.
- Zaidi, R.H., Jaal, Z., Hawkes, N.J., Hemingway, J., and Symondson, W.O.C., (1999). Can multiple-copy sequences of prey DNA be detected amongst the gut contents of invertebrate predators. *Mol. Ecol.* **8**: 2081–2087.

- Zane, L., Bargelloni, L., and Patarnell, T., (2002). Strategies for microsatellite isolation: a review. *Mol. Ecol.* **11**: 1–16.
- Zepeda-Paulo, F.A., Simon, J.C., Ramirez, C.C., Fuentes-Contreras, E., Margaritopoulos, J.T., and Wilson, A.C.C., et al. (2010). The invasion route for an insect pest species: the tobacco aphid in the New World. *Mol. Ecol.* **19**: 4738–4752.
- Zheng, S.J., and Dicke, M., (2008). Ecological genomics of plant-insect interactions: from gene to community. *Plant Physiol.* **246**: 823–827.

# Genetic Modification of Pest and Beneficial Insects for Pest-Management Programs

## Chapter Outline

- 14.1 Overview 662
- 14.2 Introduction 664
- 14.3 Why Genetically Modify Insects? 664
  - 14.3.1 Beneficial Insects 664
  - 14.3.2 Pest Insects 667
- 14.4 Why Use Molecular-Genetic Methods? 668
- 14.5 What Genetic Modification Methods are Available? 672
  - 14.5.1 Transposable-Element (TE) Vectors and Transgenesis 672
  - 14.5.2 Paratransgenesis (Genetic Modification of Symbionts) 675
  - 14.5.3 Viral Vectors 677
  - 14.5.4 Transfer of *Wolbachia* from Another Arthropod 679
  - 14.5.5 Site-Specific Modifications 679
  - 14.5.6 No Vectors 680
  - 14.5.7 RNAi to Control Pests 680
    - 14.5.7.1 RNAi in Crop Plants 684
- 14.6 Methods to Deliver Exogenous Nucleic Acids into Arthropod Tissues 685
- 14.7 What Genes are Available? 686
- 14.8 Why are Regulatory Signals Important? 686
- 14.9 How are Modified Arthropods Identified? 687
- 14.10 How to Deploy Genetically Modified Pest and Beneficial Arthropods 689
- 14.11 Potential Risks Associated with Releases of Genetically Modified Arthropods 692
  - 14.11.1 Could Gene Silencing Reduce Program Effectiveness? 692
  - 14.11.2 Relative Risks 693
  - 14.11.3 General Risk Issues 693
  - 14.11.4 Horizontal Transfer (HT) 693
- 14.12 Permanent Releases of Genetically Modified Arthropods into the Environment 697
  - 14.12.1 Models to Predict? 697
- 14.13 Regulatory Issues: Releases of Genetically Modified Arthropods 700
- 14.14 Conclusions 700
- References Cited 715

## 14.1 Overview

Genetic modification of pest and beneficial arthropods can be achieved by multiple methods. Initially, the only methods available were artificial selection or hybridization. After it became possible to develop genetically modified *Drosophila* using transposable-element (TE) vectors, scientists focused on transgenesis (insertion of exogenous genes into the nuclear genome) using TE vectors. That approach has limitations (possible lack of stability and randomness of the insertion of the transgene), as well as a potential for horizontal transfer (HT). Once RNA interference (RNAi) was discovered, its possible use to control pests began to be explored. Paratransgenesis (the insertion of genetically modified symbionts into an insect) has been studied as a method for control of vector-transmitted diseases. Another genetic modification involves the introduction of novel endosymbionts into species previously lacking that symbiont, with the goal of interrupting the transmission of pathogens such as dengue or malaria by mosquitoes. Homing endonucleases, TAL effector nucleases (TALENs), and zinc-finger nucleases offer the promise of specific genome modifications, although this research has not yet produced modified arthropods for deployment in pest-management programs. Molecular-genetic methods could improve genetic-control programs, especially the sterile insect technique, in which males of pest species are mass reared, sterilized by irradiation, and released to mate with wild females, resulting in suppression or eradication of the population over time. Producing sterile males or producing only females using molecular tools could improve the efficiency and cost-effectiveness of such programs. Other goals include producing honey bees (*Apis mellifera*) that are disease resistant, and silk moths (*Bombyx mori*) that produce improved silk or novel drugs. Natural enemies used in biological control programs could be modified to enhance their effectiveness by altering their sex ratio, host range, temperature and relative humidity tolerances, or diapause attributes.

Genetic manipulation of the nuclear genome with recombinant DNA methods requires methods for efficient and stable insertion of exogenous genes, as well as the availability of useful genes, appropriate promoters, and other regulatory elements to obtain effective expression of the inserted gene in both space and time. The resultant transgenic insects should be contained in the laboratory with effective procedures until permits have been obtained from appropriate regulatory authorities that would allow their release into the environment for either short-term evaluation or permanent release. Potential risk issues to be resolved before the release of a transgenic arthropod include whether: the inserted gene(s) (or trait) is stable and the likelihood the modified genes can be transferred horizontally to other populations or species (and the consequences if that did occur). Another concern is whether the released insects will perform as expected with

regard to their geographic distribution, host or prey specificity, or other biological attributes. Will the released insects have unintended environmental effects? In the case of short-term releases, can the released arthropods be recovered from the release site? Assessments of fitness and host specificity are relatively easy to assess in the laboratory, but evaluations of potential risks such as horizontal transfer (HT) or unintended effects on ecosystem function are more difficult.

Risk issues associated with paratransgenesis appear similar to those of transgenic arthropods, except that the potential risk of HT could be greater due to the relatively high rate of HT in the insect gut of transgenes into environmental microorganisms taken in during feeding. Risk issues associated with the insertion of a novel symbiont (such as *Wolbachia*) into a new host are perhaps lower, although it is unknown how stable such infections will be in stressful natural environments or how rapidly resistance to the effects of the symbiont could develop in the host, which could result in the loss of efficacy.

Several steps are involved in a program designed to control pest arthropods through genetic modifications, whether they be transgenesis, paratransgenesis, RNAi, or insertion of a novel symbiont. First, the target species must be identified as a significant pest for which conventional control tactics are ineffective. Genetic manipulation is more expensive and time consuming than other pest-management tactics. The increased costs are due, in part, because genetic manipulations generate concerns about risk. The genetic engineer next will want to ask, How best can our knowledge about the pest's physiology, ecology, or behavior be used against it? How will the genetically modified strain be deployed in a pest-management program?

If the ultimate deployment method requires mass rearing of very large numbers of high-quality insects, mass-rearing methods and release models must be developed. The manipulated strain may be released into greenhouses or field cages for evaluation of stability, fitness, and efficacy as a first step. Permission to release a genetically modified arthropod will have to be obtained from (likely several) regulatory agencies. Short-term releases may be made into small plots or field cages. Currently, releases of genetically modified arthropods into the environment in the United States must be short-term experiments, and the researcher is expected to retrieve all released arthropods from the environment at the end of the experiment (or the released arthropods should be sterile). At present (2012), there are no guidelines in the United States regarding permanent releases of genetically modified arthropods (GMAs). International efforts at regulation of GMAs also remain limited; some nations have biosafety laws, but many do not. Many pest-management programs, especially those involving *replacement* of pest

populations by a GMA, will require permanent establishment of the GMA in the environment and may require the use of "drive" mechanisms. Several drive mechanisms, such as release of active TEs, *Medea*, or *Wolbachia* strains that cause cytoplasmic incompatibility, have been proposed as drive elements.

## 14.2 Introduction

This chapter provides an overview of the progress in developing genetically modified arthropods (GMAs) for use in pest-management programs or for producing products for use by humans (silk, drugs, honey, pollination services). Enormous efforts have been made in the past few years to develop and test GMAs, especially genetically modified mosquitoes. The use of RNAi is being evaluated as a method of controlling pest arthropods, and mosquitoes containing *Wolbachia* isolated from another insect species have been tested in the field. Because the area of research is diverse and is expanding rapidly, it is difficult to include all relevant citations. However, key references provide the reader an entry to the literature and [Table 14.1](#) lists some of the diverse research projects being conducted involving GMAs.

## 14.3 Why Genetically Modify Insects?

### 14.3.1 Beneficial Insects

Domesticated and semidomesticated insects have been modified by traditional breeding methods for hundreds of years. Artificial selection has improved disease resistance and silk production in *Bombyx mori* ([Yokoyama 1979](#), [Gopinathan 1992](#)) and disease resistance and pollination attributes in *Apis mellifera* ([Rothenbuhler 1979](#)).

The genetic modification of silk moths to produce improved silk or to produce proteins used in drugs is an active area of research ([Goldsmith et al. 2005](#), [Royer et al. 2005](#), [Tatematsu et al. 2010](#), [Zhao et al. 2010](#)). [Teule et al. \(2012\)](#) noted that silk fibers are used as sutures, and could be used for wound dressings, artificial ligaments, tendons, and other applications in human medicine. Spiders produce silk of superior mechanical properties compared to the silk produced by *B. mori*, but spiders cannot be mass reared as efficiently. Silk-protein genes were cloned from a spider and a chimeric *B. mori*-spider silk gene was developed and inserted into silk moths with a *piggyBac* vector ([Teule et al. 2012](#)). The composite silk fibers produced by the transgenic silk moth had improved mechanical properties and the silk was produced in the same place and at the same time as native silk proteins due to the use of an appropriate promoter. [Grenier et al. \(2004\)](#) developed new *B. mori* strains to reduce risk and improve the stability of transgenic lines by producing lines that reproduced by parthenogenesis (to reduce loss or

**Table 14.1: Selected Examples of Genetic Modifications of Arthropods Using Multiple Methods.**

Project goals	Selected reference(s)
<b>Gal4-US targeted gene expression</b>	
Gene expression in <i>Aedes aegypti</i> evaluated at a specific insertion site	Kokoza and Raikhel 2011
<b>Gene knockout using Transcription Activator-Like Effector Nucleases (TALENs)</b>	
Modification of <i>Gryllus bimaculatus</i> genome	Watanabe et al. 2012
<b>Gut microbe affects malaria transmission</b>	
Naturally occurring gut microbe reduces infection of <i>Anopheles gambiae</i> by <i>Plasmodium</i>	Cirimotich et al. 2011
<b>Homing endonucleases for precise modification of the genome</b>	
Transformation of <i>Ae. aegypti</i> Transformation of <i>An. gambiae</i>	Traver et al. 2009 Windbichler et al. 2011
<b>Medea as a driver for population replacement</b>	
Medea tested for population replacement in <i>Drosophila</i> Medea is a maternal-effect dominant embryonic arrest construct that kills all progeny not inheriting the Medea- from the maternal or paternal genome. Such elements will spread if released into a population above a threshold frequency.	Chen et al. 2007
<b>Paratransgenesis</b>	
Transgenic symbiont from <i>Rhodnius</i> , the vector of Chagas disease Viral paratransgenesis in <i>An. gambiae</i> Paratransgenesis of sandfly by gut symbiont ( <i>Bacillus subtilis</i> containing a green fluorescent protein gene)	Durvasula et al. 2008 Ren et al. 2008 Hurwitz et al. 2011b
<b>Produce novel silk or proteins</b>	
Chimeric spider-silk moth silk produced in <i>Bombyx mori</i> Production of therapeutic proteins in <i>B. mori</i>	Teule et al. 2012 Royer et al. 2005, Tatematsu et al. 2010, Zhao et al. 2010
<b>RIDL for control of pest insects</b>	
Lethality driven by the dominant tetracycline-controlled trans-activator that results in lethality of progeny of transgenic <i>D. melanogaster</i> Late-acting dominant lethal system for mosquito control <i>Ae. aegypti</i> modified by RIDL also contain genes that make females flightless Conditional embryonic lethality in Medfly, <i>Ceratitis capitata</i> Conditional embryonic lethality in Caribbean fruit fly, <i>Anastrepha suspensa</i>	Horn and Wimmer 2003 Phuc et al. 2007 Fu et al. 2010 Schetelig et al. 2009a Schetelig and Handler 2012a
<b>RNAi for gene silencing or pest control</b>	
Control of western corn rootworm in corn Control of aphids	Baum et al. 2007 Jaubert-Posamai et al. 2007, Borgio 2010, Pitino et al. 2011

(Continued)

**Table 14.1: (Continued)**

<b>Project goals</b>	<b>Selected reference(s)</b>
Silencing a bollworm P450 gene impairs gossypol tolerance All-male progeny produced by knockdown of <i>tra</i> <sup>+</sup> in <i>Tribolium</i> Chitin synthase genes silenced by RNAi introduced by nanoparticle-mediated feeding of <i>An. gambiae</i> RNAi in <i>Apis mellifera</i>	Mao et al. 2007 Shukla and Palli 2012 Zhang et al. 2010  Hunter et al. 2010, Liu et al. 2010, Jarosch and Moritz 2011
<b>Silk moth transformation</b>	
Artificial parthenogenesis and voltinism control to improve safety of transgenic strains	Grenier et al. 2004
Protein production by transgenic silk moths	Royer et al. 2005, Zhao et al. 2010
Protein production in middle silk gland	Tatematsu et al. 2010
Production of chimeric spider-silkworm silk	Teule et al. 2012
<b>Sterile insect projects</b>	
Transgenic sexing strain for vector control	Catteruccia et al. 2005
Female-specific lethality	Fu et al. 2007
Genetic sexing strains of Lepidoptera	Marec et al. 2005
Fluorescent red marker gene for transgenic Caribbean fruit flies	Nirmala et al. 2011
Production of spermless <i>An. gambiae</i>	Thailavil et al. 2011
<b>Transgenic flying vaccinator</b>	
<i>An. stephensi</i> expresses a <i>Leishmania</i> vaccine candidate protein in salivary glands; mice bitten by the mosquito raised antibodies	Yamamoto et al. 2010
<b>Transgenic insects</b>	
Green fluorescent protein as a marker in <i>Ae. aegypti</i>	Pinkerton et al. 2000
Block <i>Plasmodium</i> transmission in transgenic <i>Ae. aegypti</i>	Kokoza et al. 2010
Relative fitness of <i>An. stephensi</i> lines and site-specific integration	Amenya et al. 2010
Two novel midgut-specific promoters for <i>An. stephensi</i>	Nolan et al. 2011
Transformation of <i>Lucilia cuprina</i> and <i>L. sericata</i> germ line	Concha et al. 2011
Transformation of <i>Anastrepha ludens</i>	Meza et al. 2011
Transformation of <i>Bactrocera tryoni</i>	Raphael et al. 2011
<b><i>Wolbachia</i> injected into mosquitoes</b>	
<i>Wolbachia</i> induces resistance to dengue virus in <i>Ae. aegypti</i>	Evans et al. 2009, Moreira et al. 2009a, Bian et al. 2010, Frentiu et al. 2010, Hoffmann et al. 2011, Walker et al. 2011, Yeap et al. 2011, Lu et al. 2012, Pan et al. 2012
<i>Wolbachia</i> wMel blocks dengue in <i>Ae. albopictus</i>	Blagrove et al. 2012
<i>Wolbachia</i> wMelPop inserted into <i>An. gambiae</i>	Jin et al. 2009
<i>Wolbachia</i> reduces filarial competence in mosquitoes	Kambris et al. 2009

(Continued)

**Table 14.1: (Continued)**

Project goals	Selected reference(s)
<i>Wolbachia</i> shortens life and affects feeding of <i>Ae. aegypti</i>	McMenniman et al. 2009, Moreira et al. 2009b
<b>Zinc-finger nucleases for genome modification</b>	
Silk moth modification	Takasu et al. 2010
Zinc-finger nucleases modify <i>G. bimaculatus</i> genes	Watanabe et al. 2012

movement of the transgenes by the *piggyBac* vector), lines that had the ability to develop without diapause (to screen the putative transformant lines), or the ability to diapause (to be able to store eggs over time for future use). Because silk moths cannot survive in the wild on their own, transgenic silk moths that escape from mass-rearing facilities are unlikely to establish in the environment.

Natural enemies of pest insects and mites have been modified by traditional breeding methods and by hybridization of different strains to achieve hybrid vigor (Hoy 1976, 1990a,b, 2003, Stouthamer et al. 1992, Heilmann et al. 1994, Whitten and Hoy 1999). A pesticide-resistant predatory mite, developed by traditional breeding methods, was deployed in an integrated mite management program in almonds in California and provided proof-of-principle that genetically modified natural enemies could be incorporated into practical pest-management programs (Hoy 1985). The predators provided effective control of spider mites, reduced the need for costly pesticides, reduced crop-production costs, saving almond growers ≈US\$22 million per year due to fewer applications of pesticides to control the spider mites (Headley and Hoy 1987).

Genetic improvement of natural enemies for biological control of pest insects and mites by traditional genetic methods has involved selecting for resistance to pesticides, lack of diapause, and increased tolerance to temperature extremes, although modification of other traits, such as dispersal rate and sex ratio, theoretically could result in improved biological control. Future projects could result in genetically modified natural enemies with improved traits, especially now that the genomes of several natural enemy species have been (or are being) sequenced and genes and their regulatory elements are being identified.

### 14.3.2 Pest Insects

During the past 40 years, irradiation or chemicals have been used to sterilize a number of pest insects for use in genetic-control programs (Wright and Pal 1967, Pal and Whitten 1974, Curtis 1979, LaChance 1979, Whitten 1979, Tan 2000).

This approach to genetic control has been called the **Sterile Insect Release Method (SIRM)** or the **Sterile Insect Technique (SIT)**. Male insects are mass reared and sterilized, usually by irradiation, and released. By releasing approximately 100 sterile males for each wild male, wild females should mate most often with sterile males (assuming equal fitness), resulting in reduced numbers of progeny. [Why does the program manager need to have a good estimate of the population size of the wild species in order to release the appropriate number of sterile males?] The SIRM program is most effective when females mate only once, but can be successful with species that mate more than once if sufficient numbers are released over a sufficiently long time. Serious pests, including the Mediterranean and Caribbean fruit flies, mosquitoes, and the New World screwworm, *Cochliomyia hominivorax*, have been controlled or eradicated in these genetic-control programs ([Tan 2000](#)).

A significant example of the SIRM is the eradication of the New World screwworm from the United States (see Chapter 10, Box 10.1). Later the program was expanded to eliminate *C. hominivorax* from Central America to provide a buffer zone to preclude its reintroduction into the United States. Benefits of the SIRM program in 1996 to U.S., Mexican, and Central American cattle producers were estimated to be US\$796 million, US\$292 million, and US\$77.9 million, respectively ([Wyss 2000](#)). The benefit-to-cost ratios for the eradication programs ranged from an average of 12.2 to 1 for Central America to 18 to 1 for the U.S. and Mexican programs ([Wyss 2000](#)). In addition, screwworm eradication has a significant human and wildlife health component not included in these calculations.

#### 14.4 Why Use Molecular-Genetic Methods?

Traditional genetic methods have limitations and molecular-genetic methods offer new opportunities for improving pest-management programs. Such potential benefits were perceived by many shortly after the seminal publication on *P*-element transformation of *Drosophila* by Rubin and Spradling in 1982. For example, significant benefits could accrue if molecular-genetic methods allowed sterile insects to be produced without incurring the negative effects of irradiation. During the sterilization process, the insect's whole body is irradiated, which produces damage in all tissues. As a result, the SIRM requires that large numbers of insects be reared for release. Often, pest populations first are reduced by pesticide applications or through natural seasonal (winter) mortality so that the number of insects that have to be reared can be reduced. Rearing huge numbers of sterile insects requires effective mass-rearing methods and costly facilities. Another issue is whether females should be reared and released; eliminating females using genetic methods is another desirable goal.

Molecular methods could allow unique markers, such as red or green fluorescent protein, to be inserted into sterile insects, which would allow SIRM program managers to more easily discriminate between released sterile males and wild fertile males caught in the traps used to monitor the progress of the SIRM program (Handler and McCombs 2000, Higgs and Lewis 2000, Handler and Harrell 2001). Marking methods using fluorescent dusts or dyes are not satisfactory because they can reduce fitness of the insects and the dusts/dyes do not always persist, which could lead program managers to conclude that more wild insects are present in the field than is true.

Other significant benefits could be obtained if molecular-genetic methods make it possible to control the sex of insects being reared in SIRM programs, to introduce lethal genes or genetic loads into pest populations, or to produce vectors of human and animal diseases that are unable to transmit diseases such as malaria, dengue, yellow fever, and sleeping sickness.

Molecular-genetic techniques could make genetic improvement of beneficial insects, such as silk moths, honey bees, or natural enemies more efficient and less expensive (Beckendorf and Hoy 1985, Walker 1989, Hoy 1990a,b, Heilmann et al. 1994, Beckage 1998). Once a useful gene has been cloned, it could be inserted into many beneficial species in a relatively short time. Furthermore, recombinant DNA methods broaden the number and type of genes potentially available for use; no longer is a project dependent upon the intrinsic genetic variability of the species under study.

After the development of *P*-element-mediated transformation of *D. melanogaster* in 1982, many speculated about the role that molecular genetics could play in the genetic control of vectors of human and animal diseases or pests of agricultural crops (Crampton et al. 1990, Eggleston 1991, Richards 1993, Curtis 1994, Gwadz 1994, Beaty 2000, Beard et al. 1992, 1993, 2000, Blair et al. 2000, Collins 1994, Durvasula et al. 1997, Collins et al. 2000, Aksoy et al. 2001, James 2000, 2001, Robinson and Franz 2000). Some considered transgenic technology to be a new and vitally important pest-management tool for the control of serious pests that cannot be controlled by any other means. Others expressed reservations about the goals and the methods to be used (Spielman 1994, Curtis and Townson 1998, Curtis 2000, Knols et al. 2006) (discussed further in the sections on Risk Analysis [Section 14.11] and Regulatory Issues [Section 14.13]).

There are limitations to molecular-genetic methods at present that require additional research. For example, traits primarily determined by single major genes are most appropriate for manipulating insects by recombinant DNA techniques at present. Methods for manipulating and stabilizing traits that are

determined by complex genetic mechanisms are not yet feasible with insects, although such methods could be developed using procedures developed by plant molecular geneticists as models.

Genetic manipulation of arthropods involves several steps and requires substantial investments of time and resources to answer the questions in Table 14.2. A successful project outcome requires that we have a thorough knowledge of the biology, ecology, and behavior of the target species. Identifying one or more specific traits that, if altered, potentially would achieve the goals of the project is a critically important planning step. If the nuclear genome of the target species is to be modified, suitable genes must be identified and cloned and appropriate regulatory sequences must be identified so that the inserted gene will be expressed at appropriate levels in the correct tissues and at a relevant time.

**Table 14.2: Questions to Answer when Developing a Genetic Manipulation Project if it is to be Deployed Successfully.**

PHASE I. Defining the problem and planning the project
<ul style="list-style-type: none"> <li>• What genetic trait(s) limit effectiveness of beneficial species or might reduce damage caused by the pest?             <ul style="list-style-type: none"> <li>– Do we know enough about the biology, behavior, genetics, and ecology of the target species to answer this question?</li> <li>– Is the potential trait determined by single or multiple genes?</li> </ul> </li> <li>• Can alternative control tactics be made to work more effectively and inexpensively than genetic manipulation projects, and are they more environmentally friendly?             <ul style="list-style-type: none"> <li>– The costs of genetic-manipulation projects are high and the time to develop a functional program can be quite long.</li> <li>– Transgenic or paratransgenic technology may not be appropriate if traditional genetic or other control methods can be used because issues surrounding risk assessment of releasing transgenic arthropods into the environment for permanent establishment have not been resolved.</li> </ul> </li> <li>• How will the genetically manipulated strain be deployed?             <ul style="list-style-type: none"> <li>– Will releases be inoculative and some type of selection or drive system used to replace the wild strain?</li> <li>– Will the desired genes be introgressed (introduced) into the wild population? What selection mechanism will be used?</li> <li>– Will augmentative releases of very large numbers be required?</li> <li>– Will multiple releases be required over many years?</li> </ul> </li> <li>• What risk issues should be considered in planning?             <ul style="list-style-type: none"> <li>– If pesticide resistance genes are used as a selectable marker or useful trait for beneficial species, is there a possibility of the resistance gene moving to a pest?</li> <li>– What is known about the potential for horizontal transfer?</li> <li>– If TE or viral vectors are used in the transformation process, what risks might they pose if the transgenic strain is released into the environment?</li> <li>– What health or other hazards might be imposed on human subjects if the genetically modified strain were released?</li> </ul> </li> <li>• What advice do the relevant regulatory authorities give regarding your plans to develop a genetically modified strain?             <ul style="list-style-type: none"> <li>– Which agencies are relevant to consult for your project?</li> </ul> </li> </ul>

(Continued)

**Table 14.2: (Continued)**

<b>PHASE II. Developing the genetically manipulated strain and evaluating it in the laboratory</b>
<ul style="list-style-type: none"> <li>• Where will you get your gene(s)? <ul style="list-style-type: none"> <li>– Should the transgene(s) sequence be modified to optimize expression in the target species if it is from a species with a different codon bias?</li> </ul> </li> <li>• Is it important to obtain a high level of expression in particular tissues or life stages? <ul style="list-style-type: none"> <li>– Where can you get the appropriate regulatory sequences?</li> </ul> </li> <li>• How can you maintain or restore genetic variability in your genetically modified arthropod? <ul style="list-style-type: none"> <li>– Because paratransgenic or transgenic methods typically involve substantial inbreeding to obtain pure lines, can you outcross the manipulated strain with a field population to improve its adaptation to the field or otherwise increase genetic variability?</li> </ul> </li> <li>• What methods can you use to evaluate “fitness” in artificial laboratory conditions that will best predict effectiveness in the field? <ul style="list-style-type: none"> <li>– Have life-table analyses and laboratory studies of the stability of the trait under no selection been correlated with efficacy in the field?</li> <li>– Is it possible to carry out competitive population cage studies?</li> </ul> </li> <li>• Do you have adequate containment methods to prevent premature release of the genetically modified strain into the environment? <ul style="list-style-type: none"> <li>– Have these containment methods been reviewed by appropriate regulatory authorities?</li> </ul> </li> <li>• Do you have adequate rearing methods developed for carrying out field tests? <ul style="list-style-type: none"> <li>– Are artificial diets available to reduce rearing costs?</li> <li>– Are quality control methods available to maintain quality during mass rearing?</li> </ul> </li> <li>• What release rate will be required to obtain the goals you have set? <ul style="list-style-type: none"> <li>– Do you have an estimate of the absolute population density of the target species in your field test?</li> <li>– What release model are you applying: inundative, inoculative, introgression, complete population replacement?</li> </ul> </li> <li>• Have you tested for mating biases, partial reproductive incompatibilities or other population genetic problems if interbreeding will be required?</li> <li>• Have you obtained approval from the appropriate regulatory authorities to release the genetically modified strain into the greenhouse or small plot? <ul style="list-style-type: none"> <li>– Can you contain it in the release site?</li> <li>– Can you retrieve it from the release site at the end of the experiment?</li> <li>– Can you mitigate if unexpected problems arise?</li> </ul> </li> <li>• How will you measure effectiveness of the modified strain in the field trials?</li> </ul>
<b>PHASE III. Field evaluation and eventual deployment in practical pest management project</b>
<ul style="list-style-type: none"> <li>• If the small-scale field trial results obtained in phase II were promising, questions remain to be asked before the deployment of the manipulated strain. <ul style="list-style-type: none"> <li>– Are mass-rearing methods adequate?</li> <li>– Is the quality-control program in place?</li> <li>– Is the release model feasible?</li> <li>– Were there unexpected reproductive incompatibilities between the released and wild populations?</li> </ul> </li> <li>• If permanent releases are planned, have the risk issues been evaluated? Are there possible methods of mitigation if something goes wrong?</li> <li>• How will the program be evaluated for effectiveness?</li> <li>• Will the program be implemented by the public or private sector?</li> <li>• What will the program cost and what are the benefits?</li> <li>• What inputs will be required to maintain the effectiveness of the program over time?</li> </ul>

Modified from Hoy (2000a).

Stable transformation of the nuclear genome involves incorporating the genetic information into the germ line (ovaries and testes) so that the new genetic information is transmitted to succeeding generations (=transgenesis). Several approaches have been successful in inserting exogenous nucleic acids into the genomes of insects (Table 14.3). Typically, multiple transgenic lines are developed and evaluated to determine which are most fit and stable in the laboratory; such evaluations should be conducted in a manner to preclude accidental release into the environment by using appropriate containment conditions and procedures (Hoy et al. 1997, Young et al. 2000, National Academy of Sciences 2004). If laboratory tests indicate that the transgenic strain is fit and the trait is stably and appropriately expressed, the transgenic strain(s) may be evaluated in greenhouses or small field plots to confirm their efficacy and fitness under more-natural conditions (Table 14.2).

The goals of genetic manipulation of the nuclear genomes of pest and beneficial insects are different, although most of the steps are similar (Table 14.2). The goal is to *reduce* or eliminate the pest population, or its impact; by contrast, the goal is to *enhance* or increase the population of beneficial arthropods, or their impact, in the field.

## 14.5 What Genetic Modification Methods are Available?

### 14.5.1 Transposable-Element (TE) Vectors and Transgenesis

Inserting nucleic acids into insects can be accomplished using several different techniques (Table 14.3). If the inserted DNA is incorporated into the chromosomes in the cells that give rise to the ovaries and testes, the new genetic material could be transmitted faithfully and indefinitely to successive generations (stable germ-line transformation). However, not all gene insertions remain stable due to gene silencing, DNA-repair mechanisms, or instability due to the movement of the transposable-element (TE) vector.

Initial research on stable transformation methods were accomplished with *Drosophila melanogaster* when it was discovered that the *P* element could be genetically manipulated to serve as a vector to carry foreign genes into the chromosomes of germ-line cells (Rubin and Spradling 1982, Spradling and Rubin 1982, see Chapter 9). The genes carried by the *P*-element vector became stably integrated into the chromosomes of *D. melanogaster* and were expressed. This served as the model for several years but, ultimately, the *P* element failed to transform insects other than *Drosophila* species. As a result other TEs, such as *Hermes*, *hobo*, *mariner*, *Minos* and *piggyBac*, were isolated from insects and genetically modified for use as vectors (Atkinson et al. 2001).

**Table 14.3: Methods to Deliver Nucleic Acids into Arthropods or their Cells.**

Method of delivering DNA	Example(s) (selected references)
<b>Artificial chromosomes</b>	
Insert genes into artificial chromosome, insert chromosome into genome	Not yet achieved with insects, but feasible with yeast, human (Kazuki et al. 2011), and mouse chromosomes (Peterson et al. 1997)
<b>Biostatic methods</b>	
Gene gun	<i>D. melanogaster</i> (Baldarelli and Lengel 1990) <i>Anopheles gambiae</i> eggs (Mialhe and Miller 1994) <i>Bombyx mori</i> salivary glands (Horard et al. 1994)
<b>Electroporation</b>	
Electric current punches holes in membranes, letting DNA in	<i>B. mori</i> eggs (Shamila and Mathavan 1998); <i>D. melanogaster</i> , transient expression (Kamdar et al. 1992); <i>Helicoverpa zea</i> , <i>Musca domestica</i> (Leopold et al. 1996)
Insert ds RNA into eggs and nymphs	RNAi of <i>Ixodes scapularis</i> (Karim et al. 2010)
<b>Feeding of ds RNA for RNAi</b>	
Feeding ds RNA in plants or diets Feeding <i>Escherichia. coli</i> genetically modified to produce ds RNA	Borgio 2010, Huvenne and Smagghe 2010 Li et al. 2011
<b>Microinject eggs after dechorionation</b>	
Method originating with <i>D. melanogaster</i> (Santamaría 1986), modified for each egg type	<i>B. mori</i> (Nikolaev et al. 1993, Nagaraju et al. 1996) <i>M. domestica</i> (Yoshiyama et al. 2000) <i>Pectinophora gossypiella</i> (Peloquin et al. 1997)
<b>Microinject abdomens of females</b>	
Maternal microinjection of plasmids containing transgenes	<i>Metaseiulus occidentalis</i> (Presnail and Hoy 1992, 1994) <i>Cardiochiles diaphaniae</i> (Presnail and Hoy 1996)
<b>Microinject testes</b>	
<b>Nanotechnology methods</b>	Lundin et al. 2009
<b>Nuclear transplantation</b>	
<i>D. melanogaster</i> (Zalokar 1981) Chimeric larvae of honey bee produced (Omholt et al. 1995)	
<b>Soaking insects or cells</b>	
Transfer of ds RNA for RNAi	Whyard et al. 2009, Huvenne and Smagghe 2010, Pitino et al. 2011
<b>Sperm-mediated transformation</b>	
In vitro association of DNA with sperm (Atkinson et al. 1991)	
<b>Insert DNA by artificial insemination</b>	
<i>Apis mellifera</i> (Robinson et al. 2000)	

*hobo*, isolated from *D. melanogaster*, can function as a vector in several drosophilids, as well as in several tephritids (Handler and Gomez 1996). The stability and effectiveness of the transformed insect strains developed for pest-management programs depends on the stability of the inserted gene within the

genome. As a result, it is desirable to examine the genome of the population undergoing transformation to be sure that endogenous TE elements related to the vector are lacking before conducting transformation experiments with a TE vector. For example, tephritid flies transformed with *hobo* vectors were unstable and excision was stimulated by heat shocks that presumably elicited the production of an endogenous *hobo*-like transposase (Atkinson et al. 1993). The excision rate was 8- to 10-fold higher than that seen for hosts lacking endogenous transposase. It appears that it may be insufficient to examine the target insect for endogenous versions of *hobo*, however. Laboratory assays indicated that *hobo* transposase functioned with both *hobo* and *Hermes* substrates (Sundararajan et al. 1999). The TE family that contains *hobo* includes elements from plants, fungi, fish, insects, and humans. Thus, it appears to move horizontally relatively easily and *hobo* appears to have invaded *D. melanogaster* populations sometime after 1960 (Bonnivard et al. 2000). The broad host range makes *hobo* desirable as a vector, but this is considered a negative attribute from the point of view of risk assessment.

The *mariner* element initially was isolated from *Drosophila mauritiana*, but is extremely widespread among arthropods (Robertson 1993, 1995, Robertson and Lampe 1995) and is present in nematodes (Grenier et al. 1999, Leroy et al. 2000), flatworms (Garcia-Fernandez et al. 1995), and hydras (Robertson 1997). It is found in mammals, including humans (Auge-Gouillou et al. 1995, Oosumi et al. 1995, Robertson and Martos 1997). A *mariner* vector was used to transform the chicken (Sherman et al. 1998), the zebrafish (Fadool et al. 1998), and the protozoan *Leishmania major*, which indicates that *mariner* has a general ability to “parasitize the eukaryotic genome” (Gueiros-Filho and Beverley 1997). The broad host range of *mariner* raises the question as to whether there is a risk that active *mariner* elements (or other TEs) purposefully released in insect-control programs as “drivers” could move horizontally, even to humans? At least two different subfamilies of *mariner* were found in the human genome, which suggests our genome was invaded more than once. In fact, Plasterk et al. (1999) engineered a *mariner* element to make it more active in human cells. It (*Sleeping Beauty*) has 25-fold higher levels of activity in human cells than the standard *mariner*. However, the presence of ancient and degenerated *mariner* elements in the human genome indicates that humans developed resistance to *mariner*, suggesting the risk could be low. Rates of transformation of arthropods with *mariner* have been low (Coates et al. 1995, Lampe et al. 2000, Wang et al. 2000), perhaps because many insect species have, over evolutionary time, developed the ability to suppress its damaging effects. Genetic modification of the *mariner* element could improve it as an arthropod vector (Lampe et al. 1999).

*Minos* was discovered in *Drosophila hydei* and can produce stable germ-line transformation in insects (Loukeris et al. 1995, Catteruccia et al. 2000), and can transform human cell lines (Klinakis et al. 2000). To reduce potential risk with genetically modified insects for use in pest-management programs, it may be necessary to eliminate TE vector sequences after transforming an insect, even if the element has been “disabled.” See the discussion below on “conversion,” which indicates that, under some circumstances, even disabled TE vectors can become active.

The *piggyBac* element was isolated from a nucleopolyhedrosis virus infecting cell cultures of the moth *Trichoplusia ni* (Fraser 2000), and it has a broad host range. Today, it is the most-used vector for genetic modification of the nuclear genome of insects.

#### 14.5.2 Paratransgenesis (Genetic Modification of Symbionts)

The genetic modification of symbionts of insects is called **paratransgenesis** (Miller et al. 1987, Aksoy et al. 2008, Coutinho-Abreu et al. 2010). Coutinho-Abreu et al. (2010) evaluated the status and challenges of paratransgenesis, with the goals of reducing insect vectorial capacity or to eliminate disease-causing agents. Beard et al. (1992, 1993, 2000) demonstrated that genetic engineering of insect gut symbionts is feasible by transforming a bacterial symbiont of *Rhodnius prolixus*, the Chagas-disease vector. The extracellular symbiont lives in the gut lumen and is transmitted from adult to progeny by contamination of eggshells or of food with infected feces. The modified symbionts had ampicillin and thiostrepton resistance genes, as well as genes coding for cecropin A and related pore-forming molecules (Richards 1993, Beard et al. 2000). The antibiotic resistance genes provided a selective advantage to the transgenic symbionts so they could survive antibiotics in the blood meal. The cecropin A and related molecules make holes in cell membranes, perhaps leading to lysis of the Chagas disease-causing agent.

Field trials were conducted in Guatemala to determine whether the engineered symbiont of *Rhodnius* could be transmitted to *Rhodnius* by a simulated triatomine-fecal material called CRUZIGARD (Hurwitz 2011a,b). Hurwitz et al. (2011b) acknowledged, “deployment of genetically altered lines of bacteria to target field populations of triatomine bugs may have profound environmental consequences.” To assess the possible risk of horizontal gene transfer, a model was developed that attempted to simulate the environmental conditions of such a project (Matthews et al. 2011). Durvasula et al. (2008) also transformed a bacterial symbiont of *Triatoma infestans*, another vector of the Chagas disease-causing agent in South America, with a single-chain antibody.

Symbionts of tsetse flies (*Glossina* species), which are vectors of both animal and human African sleeping sickness, have been transformed (Richards 1993, Cheng and Aksoy 1999, Aksoy et al. 2001, 2008, Hurwitz et al. 2011a). Proposals have been made to release tsetse flies carrying transgenic symbionts so the released flies could replace or outcompete native populations, but fail to transmit the disease. Because the host range of these bacteria is narrow, horizontal transfer (HT) of the transformed bacteria is less likely. One of the symbionts, *Sodalis*, can be cultured and genetically modified (Aksoy et al. 2008). *Sodalis* is a secondary symbiont that is found in the hemolymph, midgut, and milk gland and is transmitted vertically through the milk glands. Because *Sodalis* is in proximity to the trypanosome parasite and is resistant to proteins produced by the trypanosome, paratransgenic *Sodalis* can be introduced by injection and potentially reduce transmission of the disease-causing agent. The endemic *Sodalis* could be eliminated with specific antibiotics without eliminating the essential *Wigglesworthia* symbiont. Deployment of the paratransgenic tsetse flies could be achieved using a drive system to replace wild flies with the flies unable to vector the trypanosome. An alternative is to use the SIT in order to eliminate wild-type tsetse and follow it with releases of paratransgenic flies that cannot transmit the pathogen.

Extracellular bacteria isolated from the gut of the walnut husk fly, *Rhagoletis completa*, was transformed with enhanced green fluorescent protein (GFP) and zeomycin resistance genes (Peloquin et al. 2000). This modified bacterium in theory could deliver proteins into the gut that could enhance the nutrition of the flies, improving their vigor and competitiveness for potential SIT programs.

Ren et al. (2008) discovered, cloned, and characterized a densovirus able to infect the mosquito *Anopheles gambiae*, and introduced the GFP gene into the virus. The virions were highly infectious and disseminated to and expressed GFP in the midgut, fat body, and ovaries. Ren et al. (2008) concluded that this virus could be used as part of a paratransgenic malaria-control strategy if anti-*Plasmodium* peptides or insect-specific toxins could be introduced into the virus. Cirimotich et al. (2011) found a naturally occurring *Enterobacter* in wild *An. gambiae* populations in Zambia. The bacterium “renders the mosquito resistant to infection with the human malaria parasite *Plasmodium falciparum* by interfering with parasite development before invasion of the midgut epithelium.” The question is: Can this bacterium be transmitted to wild *An. gambiae* populations elsewhere to prevent malaria transmission?

Sand flies (Psychodidae) transmit leishmaniasis (also known as kala azar), causing nearly 100,000 deaths per year in the Indian state of Bihar (Hurwitz et al. 2011a).

DDT resistance in the sand fly *Phlebotomus argentipes* makes it difficult to manage the vector. *Bacillus subtilis*, a commensal bacterium found in the gut of the sand fly, was transformed using GFP and added to the food of the larvae. The bacteria were transstadially transmitted and adults emerging from the treated food carried large numbers of modified bacteria in their gut lumens. This indicates that paratransgenic manipulation of the sand fly is possible once genes are discovered that can kill or inactivate the leishmaniasis-causing pathogen.

Rio et al. (2004) outlined the requirements for use of paratransgenesis for control of human or animal disease (Table 14.4). As you can see, paratransgenesis projects require extensive efforts in both development and risk assessment.

#### 14.5.3 Viral Vectors

Several types of viruses have been modified to serve as vectors in insects or insect cells (Burns 2000, Carlson et al. 2000, Olson 2000, Terzian et al. 2000, Webb 2000). In some cases, these viruses are intended to yield stable transformation, but others are expected to result in a short-term transformation of the infected tissues.

Nuclear polyhedrosis viruses (NPVs), also known as baculoviruses, have double-stranded, circular DNA genomes contained within a rod-shaped protein coat.

**Table 14.4: Requirements for Effective Control of Human or Animal Disease by Paratransgenesis in Arthropods.**

- Naturally occurring symbionts must be isolated and cultured.
- Transmission mode and population dynamics of the symbiont throughout the host life cycle should be known.
- A transformation system that produces stable phenotypes is needed.
- Anti-pathogenic genes (transgenes) such as antimicrobial peptides or transmission-blocking monoclonal antibodies should be identified, cloned, and evaluated.
- Transgenes are expressed by the symbiont without fitness costs.
- It is possible to insert the transgenic symbiont into hosts (lacking normal symbionts).
- The transgene products are produced in the tissues or organs that can affect the pathogen.
- The potential emergence of resistance by the pathogen to the transgene product should be evaluated.
- The repopulated insects (with the transgenic symbiont) should be as fit as the wild-type insect hosts.
- An effective gene-driven mechanism(s) should be available to replace field populations with the paratransgenic insects.
- An assessment of the potential spread of the transgenic symbionts and barriers to dispersal to nontarget species should be conducted.
- An in-depth analysis of the potential for harmful effects of symbionts or of transgene products to humans and animals should be conducted.
- Any implementation project should be evaluated for environmental impacts.

Baculoviruses infect many pest insects and several have been used as biological pesticides, including *Autographa californica* NPV and *Lymantria dispar* NPV (Shuler et al. 1994). These, and the *Bombyx mori* NPV, were exploited as vectors to carry exogenous DNA into insect cells (Miller 1988, Iatrou and Meidinger 1990, Yu et al. 1992). Baculovirus vectors can produce a high level of commercially important proteins in insect cell cultures (Frommer and Ninnemann 1995, Jones and Morikawa 1996). The host range of baculovirus vectors has been found to include human liver cells, suggesting that baculoviruses could be used for gene therapy in humans (Hofmann et al. 1995). Recombinant baculoviruses can integrate into Chinese hamster ovary chromosomes in cell cultures, suggesting they could be used as gene vectors for transforming mammals in a stable manner (Merrihew et al. 2001).

Densoviruses (Parvoviridae) are linear single-stranded DNA molecules that apparently are restricted to arthropods and can be used to deliver genes into mosquitoes for laboratory studies of gene expression. Densovirus also might be used for biological control programs directed against mosquitoes (Beaty and Carlson 1997, Carlson et al. 2000).

Retroviral vectors were developed by genetically modifying the Moloney murine leukemia virus so that it contains the G-envelope protein from vesicular stomatitis virus (Burns 2000). These retroviral vectors have a very wide host range (are pan-tropic), but are considered to be stable once inserted into the host genome because they lack the genetic information needed to propagate. These vectors have been evaluated for human gene therapy and can be used to transform fish, cows, clams, and amoebae, as well as lepidopteran and dipteran cells (Burns 2000). The viruses have been used to study promoter function and regulation in insect cells (Matsubara et al. 1996, Jordan et al. 1998, Burns 2000). Silk moth embryos infected with pseudotyped retroviral particles carrying the GFP construct yielded larvae that contained these viral vector sequences, indicating the virus integrated into the genome (Komoto et al. 2000).

The *gypsy* element in *D. melanogaster* is an infectious retrovirus; it was the first retrovirus to be identified in invertebrates (Kim et al. 1994, Bucheton 1995). *gypsy* normally is repressed (prevented from moving) by a gene in *Drosophila* called *flamenco*. Apparently, *gypsy* elements invaded *D. melanogaster* a long time ago and *D. melanogaster* survived the invasion because variants of the *flamenco* gene were able to suppress the activity of the invading *gypsy* (Pelisson et al. 1997). This is a fine example of the ability of insects to evolve resistance to invading nucleic acids (including TEs) that cause genetic damage.

Alphaviruses (Togaviridae) have a single-stranded RNA genome and have been genetically engineered as expression vectors (Beaty and Carlson 1997). These

viruses can be grown in mammalian cells and the viruses produced can infect either mosquitoes or mosquito cell cultures. Infection is sustained and Sindbis infection was used to express an antisense form of a dengue protein in *Ae. aegypti* adults, making the mosquitoes unable to transmit the viral disease (Olson et al. 1996, Olson 2000). The Sindbis virus can be fed to mosquitoes, allowing expression of transgenes in the midgut (Olson et al. 2000). No arthropods that have been genetically modified using these viruses have been released into the environment.

#### 14.5.4 Transfer of *Wolbachia* from Another Arthropod

Transfer of *Wolbachia* into a novel arthropod host can result in unstable or stable infections (Werren and Bartos 2001). For example, Kawai et al. (2009) introduced *Wolbachia* into planthoppers by injecting cultured *Wolbachia* into nymphs. One line of *Laodelphax striatellus* lost its infection within several generations. In the second line, the infection was present for >7 years, but the frequency of infection declined from ≈80% to <10% after 40–60 generations. Heath et al. (1999) found that *Wolbachia* can be transmitted to a parasitic wasp (*Leptopilina boulardi*) from its host, *Drosophila simulans*, although the infection was lost after three generations.

Some strains of *Wolbachia* reduce the rate of transmission of diseases by mosquitoes, either by reducing lifespan or by inhibiting the replication of mosquito-borne pathogens such as filarial worms, dengue, or chikungunya viruses in *Aedes aegypti* and of *Plasmodium* parasites in *Ae. aegypti* and *An. gambiae* (Min and Benzer 1997, Cook et al. 2008, Kambris et al. 2009, McMeniman et al. 2009, Moreira et al. 2009a,b, Bian et al. 2010, Hussain et al. 2011). One of the most-advanced efforts to use *Wolbachia* to control disease transmission by the mosquito *Aedes aegypti* has been studied by Scott O'Neill's group in Australia, as described in Box 14.3 in Section 14.13.

Hancock et al. (2011) discuss methods for introducing *Wolbachia* into wild mosquito populations for control of mosquito-borne diseases. They modeled different introduction scenarios involving the releases of infected mosquitoes, included seasonal fluctuations in the size of populations, and concluded that releases of mostly males allow the infection to spread after the introduction of a few females. This strategy was expected to cause substantial reductions in transmission without increasing risk of disease.

#### 14.5.5 Site-Specific Modifications

The ability to introduce cloned genes into the germ line at a predictable chromosomal site is especially desirable, because it should reduce the likelihood of position

effects on gene expression. Genes introduced by TE and viral vectors insert more or less randomly into the chromosomes, making it difficult to predict how well the transgene will be expressed. The use of zinc-finger nucleases, homing endonucleases, and TAL effector nucleases (TALENs) to genetically modify arthropod genomes in a specific manner was described in Chapter 9. To date, these methods are being used as tools to evaluate gene function (Wright et al. 2006, Miller et al. 2007, Beumer et al. 2008, Traver et al. 2009, Takasu et al. 2010, Urnov et al. 2010, Carroll 2011, Windbichler et al. 2011, Watanabe et al. 2012), but could be used to develop genetically modified arthropods for pest-management programs. At present, however, the many issues regarding their stability and their potential risks and relevant regulations have not been resolved. However, the potential risks of HT by these methods could be greatly reduced compared with the use of TE vectors.

Another method for accomplishing precise insertion is based on a system found in the yeast *Saccharomyces cerevisiae*. A gene for yeast recombinase, FLP recombinase, and two inverted recombination target sites (FRTs) that are specifically recognized by the FLP recombinase have been cloned. The FLP-FRT system has been modified to insert DNA into a specific site in a *Drosophila* chromosome (Konsolaki et al. 1992, Simpson 1993, Golic et al. 1997). If the FRT sites are inserted into other insects, the system could reduce concerns about unstable transformation that are elicited by TE vectors. Because a stable FRT site must be present in the genome, a number of different lines carrying FRT sites in different chromosomal locations will have to be evaluated to determine which site permits better expression of the introduced genes. The FRT system is introduced into *D. melanogaster* using TE vectors, so vector sequences may have to be removed to preclude subsequent movement.

#### **14.5.6 No Vectors**

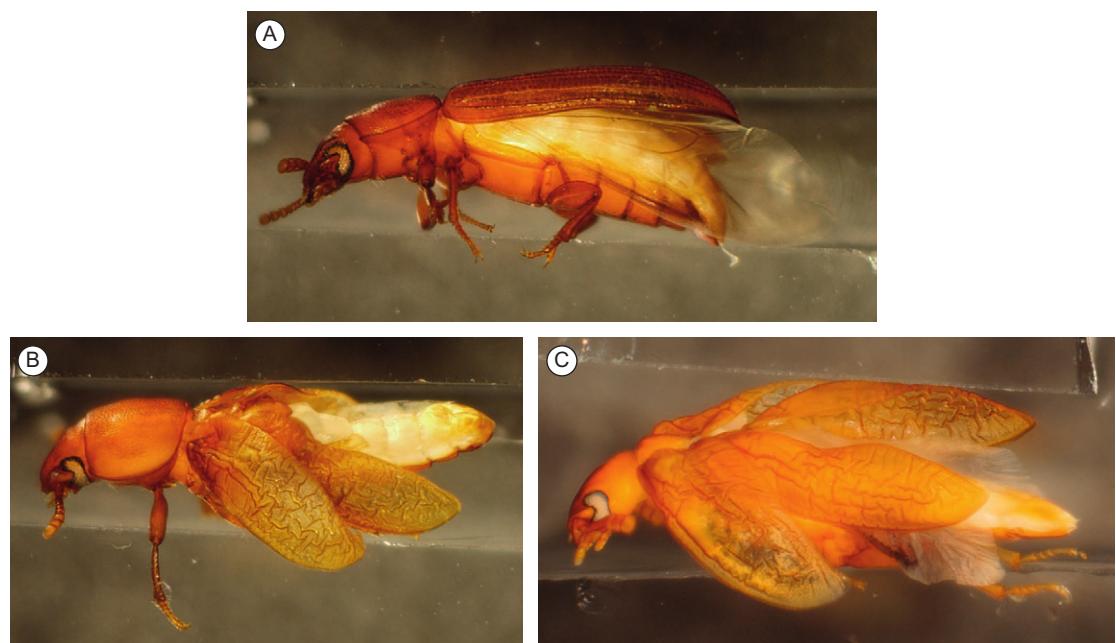
A few experiments have delivered linear or circular plasmid DNA into the genomes of arthropods without using a specific vector (Presnail and Hoy 1992, 1996). This approach has the advantage of eliminating potential risks of introducing TE sequences into the insect genome, which could result in increased stability of the inserted genes in the genome. It assumes that the inserted gene is no more likely than any other gene to be moved by endogenous TEs or viruses. How integration occurs is unknown, but could involve nonhomologous recombination or DNA-repair mechanisms.

#### **14.5.7 RNAi to Control Pests**

RNAi was described in Chapter 9 as a method for analyzing gene function in *D. melanogaster*. However, it has potential for use in managing pest arthropods, as well (De la Fuente et al. 2007, Whyard et al. 2009, Belles 2010, Huvenne

and Smagghe 2010, Perrimon et al. 2010, Yang et al. 2011, Yu et al. 2012). RNAi methods for pest control could involve developing transgenic insects that carry small hairpin RNAs (shRNAs), small hairpin microRNAs (shmiRNAs), or long ds RNAs because they allow for controlled or continuous expression of small transcripts in the cell that contain both the sense and antisense strand complementary to the targeted mRNA (Perrimon et al. 2010). These will need to be maintained as stable extrachromosomal copies or stably integrated in the genome as transgenes. shmiRNAs are considered more effective in knocking down target mRNAs than shRNAs.

Systemic RNAi occurs in some insects, but not others (Price and Gatehouse 2008, Tomoyasu et al. 2008). Whyard et al. (2009) showed that ingested ds RNAs can act as species-specific insecticides when fed to *D. melanogaster*, *Tribolium castaneum*, *Acyrtosiphon pisum*, and *Manduca sexta* when each insect was fed species-specific ds RNAs targeting vATPase genes. The ds RNAs had to be encapsulated into liposomes to ensure uptake in the *D. melanogaster* diet, but the other species tested did not need this treatment.



**Figure 14.1** A) Normal *Tribolium castaneum* adult. B) RNAi of the Hox gene *ultrabithorax* in larvae leads to adults with two pairs of elytra. C.) RNAi of *Scr* (*Sex combs reduced*) results in beetles with two pairs of elytra and membranous hind wings. (Photos courtesy of Y. Tomoyasu.)

**Table 14.5: Selected Examples of RNAi Experiments with Arthropods other than *Drosophila*.**

Order and species	Traits modified/Method introduced	Reference(s)
<b>Acari</b>		
<i>Tetranychus urticae</i>	<i>Distalless</i> gene/Injected	Khila and Grbic 2007
<i>Varroa destructor</i>	<i>Glutathione S-transferase</i> /Soak	Campbell et al. 2010
<i>Haemaphysalis longicornis</i>	Engorgement, mortality, egg hatch affected/Injected	Aung et al. 2011
<i>Ixodes scapularis</i>	Multiple genes/Electroporated	Karim et al. 2010
<b>Blattaria</b>		
<i>Blattella germanica</i>	<i>Acetylcholinesterase</i> activity and sensitivity to insecticides/Injected	Revuelta et al. 2009
	Function of RXR in molting confirmed/Injected	Martin et al. 2006
	Ecdysone receptor function/Injected	Cruz et al. 2006
<b>Coleoptera</b>		
<i>Diabrotica virgifera virgifera</i>	Multiple traits/Fed	Baum et al. 2007
<i>Tribolium castaneum</i>	Development/Injected	Bucher et al. 2002, Tomoyasu and Denell 2004, Tomoyasu et al. 2008
<b>Diptera</b>		
<i>Aedes aegypti</i>	Vacuolar ATPase/Fed	Coy et al. 2012
<i>Bactrocera dorsalis</i>	Four genes/ <i>E. coli</i> expressing ds RNA fed to flies	Li et al. 2011
<i>Culex pipiens</i>	<i>heatshock 90</i> /Rehydration of mosquito larvae	Lopez-Martinez et al. 2012
<i>Glossina morsitans</i>	Immunity genes/Fed and Injected	Walshe et al. 2009
<i>Haematobia irritans</i>	Multiple genes/Injected	Torres et al. 2011
<b>Hemiptera</b>		
<i>Acyrthosiphon pisum</i>	Two marker genes/Injected	Jaubert et al. 2007
<i>Bemisia tabaci</i>	Toxicity of 5 genes/Fed	Upadhyay et al. 2011
<i>Myzus persicae</i>	Two genes/Fed	Pitino et al. 2011
<i>Oncopeltus fasciatus</i>	Morphology affected by <i>nubbin</i> / Parental RNAi	Hrycaj et al. 2008
<i>Rhodnius prolixus</i>	<i>Nitrophorin 2</i> in salivary gland	Araujo et al. 2006
<b>Hymenoptera</b>		
<i>Apis mellifera</i>	<i>Vitellogenin</i> gene function/Injected <i>Relish</i> regulates antimicrobial peptide genes/Injected	Amdam et al. 2003 Schluns and Crozier 2007
<i>Nasonia vitripennis</i>	Israeli acute paralysis virus silenced/Fed Parental RNAi to confirm gene function Larval RNAi produced <i>cinnabar</i> eye color	Maori et al. 2009 Lynch and Desplan 2006 Werren et al. 2009

(Continued)

**Table 14.5: (Continued)**

Order and species	Traits modified/Method introduced	Reference(s)
<b>Isoptera</b>		
<i>Reticulitermes</i>	Caste determination by <i>hexamerin</i> /Injected	Zhou et al. 2006
<i>R. flavipes</i>	Toxicity induced by 2 genes/Feeding	Zhou et al. 2008
<b>Lepidoptera</b>		
<i>Epiphyas postvittana</i>	Larval gut carboxylesterase gene, Pheromone-binding protein, Injected	Turner et al. 2006
<i>Helicoverpa armigera</i>	Cytochrome P450 gene/Fed	Mao et al. 2007
<i>Plutella xylostella</i>	Pheromone production/Injected	Lee et al. 2011
<b>Orthoptera</b>		
<i>Gryllus bimaculatus</i>	Analysis of segmentation/Injected <i>caudal+</i> needed for gnathal and thoracic patterning/Injected <i>Sulfakinin</i> , a satiety effector/Injected Analysis of eye development/Injected	Miyawaki et al. 2004 Shinmyo et al. 2005 Meyering-Vos and Muller 2007 Takagi et al. 2012

RNAi has not been successful in all species tested, and Garbutt et al. (2012) suggested that an analysis of the persistence of ds RNA in the hemolymph of insects could be a predictor of success with RNAi. They showed that ds RNA was rapidly degraded in *Manduca sexta*, a species that is variably sensitive to RNAi, but persisted much longer in *Blattella germanica* hemolymph, which is highly sensitive to RNAi, based on a quantitative reverse-transcription PCR-based assay. This suggests that rapid degradation of ds RNA by extracellular nucleases may be one cause of insensitivity to RNAi.

Table 14.5 lists some of the RNAi experiments that have been conducted with arthropods other than *Drosophila*. RNAi has been successful in diverse species, ranging from spider mites (*Tetranychus urticae*) and ticks to aphids, honey bees, termites, cockroaches, *Tribolium*, and Lepidoptera. Delivery of the ds RNA has been achieved by injection (Cruz et al. 2006, Martin et al. 2006, Aung et al. 2011), feeding ds RNA (Maori et al. 2009, Walshe et al. 2009, Whyard et al. 2009, He et al. 2011), feeding *Escherichia coli* expressing ds RNA (Li et al. 2011), soaking (Lopez-Martinez et al. 2012), electroporation (Karim et al. 2010), or by transmission from parent to progeny (Bucher et al. 2006, Khila and Grbic 2007, Li et al. 2011). Figure 14.1 shows *Tribolium castaneum* adults that have been genetically modified using RNAi. This species has a systemic response and RNAi is serving as a method to understand gene function and development in this beetle.

Improvement in delivery of ds RNAs is highly desirable (Yu et al. 2012) and experiments with placing the ds RNA into liposomes or other protectants may be needed to enhance stability of the ds RNA. In some cases, RNAi is systemic (transferred into cells throughout the body), whereas in other cases it is not. Sometimes the knockdown of gene expression is brief and sometimes it is lengthy. At present it is difficult to predict the outcome of RNAi experiments (Terenius et al. 2011). Additional research is needed to select appropriate target genes and targeted regions within the gene. The optimal length of the ds RNA is another issue that needs to be resolved to enhance efficacy of the RNAi response (Saleh et al. 2006, Whyard et al. 2009). Table 14.6 lists some of the relevant concerns regarding making RNAi a useful pest-management tool in the future.

#### 14.5.7.1 RNAi in Crop Plants

RNAi is both systemic and heritable in plants. The small interfering RNAs (siRNAs) can move between cells through channels in cell walls and thus can be distributed throughout the plant (Gordon and Waterhouse 2007, Perrimon et al. 2010). Although much RNAi work in plants involves improving the production or nutritional value of crops, some efforts are being made to make crop plants resistant to insects (Baum et al. 2007, Mao et al. 2007, Price and Gatehouse 2008, Huvenne and Smagghe 2010, Zha et al. 2011).

These efforts are just beginning and require additional research before this technology can become a commercial success. Issues that need to be resolved before transgenic crops can be used effectively in agricultural pest management include resolving the concentration of ds RNA needed to induce optimal silencing (note that 100% suppression of gene expression is unlikely). The sequence used will determine whether off-target effects occur in the plant, or in other insects. The ideal situation would have the target pest, and only the target pest, affected at an appropriate level. It also will be critical to resolve the effective length of the ds RNA; this determines how well the ds RNA is taken up and how efficiently the silencing is obtained. ds RNA that is too short or too long can cause nontarget effects or result in poor uptake, respectively. The persistence of the silencing effect also is important; transient expression may result in reduced control. Another issue is which life stage of the pest to target; younger stages often exhibit greater silencing effects. To date, RNAi directed against Lepidoptera appears very difficult to achieve, which is unfortunate because this order contains serious pest species (although lepidopterans may be well controlled by the insertion of *Bacillus thuringiensis* [Bt] toxin genes into crop plants). An important target would be insects that feed on phloem such as aphids, leafhoppers, psyllids, and whiteflies; Bt toxin genes cannot control these pests. At present, it is not known whether ds RNAs can be transmitted in phloem. Finally, many RNAi experiments result in a knockdown of gene function in the pest, but not a complete silencing of the gene. Thus, the effect

**Table 14.6: Successes and Concerns with RNAi in Insects.**

Successes
<p>RNAi allows scientists to evaluate the function of genes in many insects. Species that have had their genomes sequenced readily allow effective ds RNAs to be designed for many different genes.</p> <p>RNAi can be used in cell cultures or in living organisms. Systemic RNAi in living organisms is most relevant for determining the function of genes involved in development, reproduction, behavior, immunology, and other aspects of insect biology. Systemic RNAi involves the spread of RNAi effects from cell to cell and tissue to tissue in the living insect.</p> <p>By 2010, RNAi had been used in species of arthropods in the orders Orthoptera, Blattaria, Isoptera, Hemiptera, Coleoptera, Neuroptera, Hymenoptera, Lepidoptera, and Diptera, as well as in some mites and ticks (Acari), indicating RNAi can provide functional genomics data for many arthropods.</p>
Concerns
<p>RNAi may result in off-target effects, leading to knockdown of genes with structural similarity to the delivered ds RNA.</p> <p>RNAi may inhibit gene expression, but <i>rarely</i> eliminates it. It is not a knockout, but a <i>knockdown</i> procedure that can be transient. If the effects are too transient, a phenotype may not be detected.</p> <p>Arthropods vary in their ability to develop a systemic response to RNAi, probably for several reasons including: efficient degradation of alien ds RNA by some species; deficient amplification and spread of the ds RNA signal; the tissue is not permeable to the ds RNAs; the gene counteracts RNAi by increasing transcription rates; the target mRNA is too transient or is protected from RNases. It appears that “less-derived species” are more likely to express systemic responses.</p> <p>The mechanisms underlying amplification and spreading of the RNAi signal are not fully understood in arthropods.</p> <p>Methods for delivering ds RNA to arthropods vary, including injection, feeding, electroporation, or soaking. The amount of ds RNA required to elicit a response also varies.</p> <p>For RNAi to function in insect control in crop plants, the insect must take up the ds RNA of a target gene through feeding; the ds RNA should pass through the gut lumen into the gut cells; and, if the target gene is expressed outside the gut, the silencing signal will have to spread (systemic RNAi). The nucleotide sequence of the ds RNA should not function in other insects (nontarget effects). If the silencing effect is transient, effective control may not be achieved. The optimal life stage of the insect to be targeted should be determined.</p> <p>If crops are produced that are resistant to major pest insects, resistance to RNAi could develop. Questions include: should RNAi function continuously and in all plant parts at high concentrations? Should RNAi genes be “stacked” to reduce the likelihood of resistance in the pests?</p> <p>Other issues relate to the risk assessment of RNAi plants; are there risks unique to RNAi mechanisms?</p>

Derived from Belles (2010), Huvenne and Smagghe (2010), and Center for Environmental Risk Assessment (2011).

may be transient, which could induce resistance in the target pest. One solution to delay resistance in the pest would involve stacking target genes (pyramiding).

## 14.6 Methods to Deliver Exogenous Nucleic Acids into Arthropod Tissues

A variety of methods have been evaluated for delivering genes and vectors into arthropods to achieve transformation (Table 14.3). Current methods include

microinjection of TE vectors and other vectors into dechorionated insect eggs or microinjection of plasmids directly into the testes of males or the abdomen of female mites or insects (maternal microinjection). Less frequently, DNA has been delivered by soaking eggs in DNA, using sperm to carry foreign DNA into eggs of the honey bee, by using microprojectiles (gene gun technology) to insert DNA into insect eggs, electroporation of DNA into insect eggs, and transplanting nuclei and cells. Future methods might involve inserting artificial chromosomes into the insect genome, especially if multiple genes are to be inserted.

### 14.7 What Genes are Available?

Cloned DNA can be isolated from the same or other species. It is technically feasible to insert genes from microorganisms into arthropods and have the DNA transcribed and translated, although coding sequences isolated from microorganisms must be attached to promoters (controlling elements) and other regulatory sequences derived from a eukaryotic organism so that the gene can be expressed in insects. The regulatory sequences determine when a gene will be transcribed, at what level, in what tissues, and how long the messenger RNA can be used for translation. Considerable research has been conducted to identify regulatory sequences that regulate genes in specific tissues.

It also may be possible to isolate a gene from the species being manipulated, alter it, and reinsert it into the germ line, although this approach has not yet been attempted in insects other than *Drosophila*. This approach has received increased interest in plant breeding using recombinant DNA methods because it decreases concerns about risks. The complete sequence of genomes from multiple arthropod species (see Chapter 7) increases our ability to identify interesting genes in other arthropods using sequence similarity of conserved regions.

Many relevant traits probably are determined by multiple genes, which make them difficult to manipulate at present. Inserting traits determined by multiple genes into an insect by recombinant DNA methods has not been achieved, but there are several methods by which several foreign genes can be introduced into a transgenic plant that might be adapted for use in insects (Halpin et al. 2001, Hunt and Maiti 2001).

### 14.8 Why are Regulatory Signals Important?

Genes consist of coding segments that determine the amino-acid sequences in the enzyme or structural proteins produced. However, whether a protein-coding region is transcribed and translated in a specific tissue is determined by regulatory sequences in the DNA, including promoters, enhancers and multiple types

of regulatory RNAs. Some of these regulatory elements are in proximity to the coding region, while others may be located far away. The stability of messenger RNA is influenced by signals within the mRNA that can influence the amount of protein produced in the cytoplasm. It is crucial to obtain expression of the inserted gene at appropriate times and levels, and in the targeted tissues.

A factor that may be important in maintaining the inserted DNA in the transgenic arthropod over time is the presence of "origins of replication" that regulate DNA replication of the chromosomes. If a transgene is inserted into a region of the chromosome far from a site where an origin of replication occurs naturally, the gene could be lost over time because it is not replicated. Alternatively, the transgene may be lost due to DNA-repair mechanisms, especially when the arthropod is placed into a stressful environment.

A protein-coding sequence from a prokaryote such as *E. coli* can be used to form a DNA construct that will function in an arthropod if regulatory sequences from an arthropod are used (regulatory sequences from prokaryotes do not function in arthropods). Because regulatory sequences may vary from species to species, the source of regulatory sequences chosen may be as important, or even more important, as the source of the protein-coding sequences. Furthermore, some regulatory sequences allow genes to be expressed only in particular tissues or in response to particular stimuli (such as heat shock), whereas other genes are expressed in most tissues most of the time. If it is important that the inserted gene function in a tissue- or stimulus-specific manner, it is essential to identify tissue- or stimulus-specific promoters.

Initially, the number of suitable regulatory sequences available for genetic manipulation of arthropods was limited, but that is changing with the sequencing of diverse arthropod genomes. Project goals will dictate what type of regulatory sequences might be most useful. In some cases, a low constitutive production of transgenic proteins will be useful, while in other cases high levels of protein production will be required after induction by a specific cue. Researchers may have to evaluate the tradeoffs between high protein production and the subsequent effect these have on fitness of the transgenic arthropod strain based on the specific goals of each program. The presence of introns sometimes may be necessary for high levels of transgene expression in transgenic insects ([Duncker et al. 1997](#)).

## 14.9 How are Modified Arthropods Identified?

After inserting the desired genes and regulatory elements into the nuclear genome of an arthropod, the next issue is how to detect whether the transgene

has in fact been incorporated into the germ line. Because transformation methods are relatively inefficient, a screening method is needed to rapidly and consistently identify transformed individuals. This process is relatively simple in *Drosophila*, where a wealth of genetic information and visible markers, such as eye color genes, can be used to identify transgenic individuals. Many pest and beneficial arthropods lack such extensive genetic information or markers. It is possible to use specific primers and the PCR to confirm that the DNA sequences are present, but this only indicates the DNA is present—not that it has been incorporated into the genome. Southern-blot analyses can confirm the integration and number of integrations. Ideally, the insertion site of the gene can be determined by inverse PCR and sequencing to obtain flanking sequences, as well.

Identifying transformed individuals also could be achieved using a pesticide-resistance gene as a selectable marker. However, the release of pesticide-resistant pest arthropods into the environment would create concerns about risk. Researchers working with the Medfly resolved to forego using resistance genes as selectable markers for this reason (Ashburner et al. 1998). Concerns about releasing pesticide-resistant natural enemies might be lower, except that there is a possibility of HT of pesticide resistance genes from beneficial to pest species. No one knows at present how to quantify this potential risk.

Another option is to use antibiotic resistance genes as selectable markers to identify transgenic arthropods. However, HT of antibiotic resistance genes into microbes in the environment could result in increased levels of antibiotic resistance in pathogenic microbes. Antibiotic resistance genes are no longer considered safe for release into the environment in transgenic crops and methods have been developed to remove them (Yoder and Goldsbrough 1994, Witte 1998, Ebinuma et al. 2001, Matthews et al. 2001). It is probably desirable to eliminate unneeded marker genes from arthropods before their permanent release into the environment.

Another potential marker is based on the  $\beta$ -galactosidase gene (the *lacZ* construct) isolated from *E. coli*, which can be detected by an assay that produces a blue color in the transformed insects and mites. This construct has been present in a number of organisms released into the environment and risks associated with the release of this construct are considered low (Hoy 2000a,b). Eye color (Zwiebel et al. 1995, White et al. 1996, Cornel et al. 1997, Ke et al. 1997, Berghammer et al. 1999, Sarkar and Collins 2000) and red or green fluorescent protein genes (Higgs and Lewis 2000, Horn et al. 2000, Pinkerton et al. 2000) also are considered safe selectable markers if transgenic arthropods are to be released into the environment. Unfortunately, transgenic arthropods with

mutant eye color genes may exhibit abnormal behavior, which could reduce their effectiveness in the field. The effects of GFP on vision could be important if the GFP is expressed in the eyes (Horn et al. 2000). Normal behavior often is crucial to the function of released genetically modified arthropods.

#### 14.10 How to Deploy Genetically Modified Pest and Beneficial Arthropods

A critically important step is consideration of *how* to use a genetically modified arthropod strain in pest-management programs (Table 14.2, Phase III). Ideally, the questions outlined in Table 14.7 should be considered when *initiating* the project, because genetic manipulation projects of beneficial or pest arthropods are neither rapid, inexpensive, nor simple. Understanding the biology and ecology of the target species is essential (Scott et al. 2008).

The efficacy of a “drive” mechanism (such as *Wolbachia* or active TEs) combined with a “driven gene” to control a pest population has not yet been demonstrated in any practical pest-management program. Although some small-scale experimental releases have occurred (with nontransgenic insects), this type of pest population manipulation raises several questions regarding risk and effectiveness. Clearly, the driver and the gene to be driven should be strongly linked if the combined system is to spread through a population (Kidwell and Ribeiro 1992, Brookfield and Badge 1997, Curtis 2000, Braig and Yan 2002, James 2004, 2005, Sinkins and Gould 2006, Chen et al. 2007, Huang et al. 2007, Benedict et al. 2008).

Genetic control (SIRM or SIT) programs usually require repeated releases of large numbers of sterile pests. The insects produced must be free of diseases, vigorous, competitive, and free of genetic deterioration caused by inbreeding or inadvertent selection for laboratory adaptations (Bush 1979, Benedict et al. 2009). Large-scale rearing of insects is difficult and expensive. If the goal is to release pests, such as mosquitoes, that are able to reproduce and may bite humans and domestic animals, it also will be important to have the consent and cooperation of the human inhabitants of the release area (Aultman et al. 2000, Marshall et al. 2010). Past genetic-control programs involving releases of mosquito vectors of disease have elicited concern by inhabitants of the affected area (Pal 1974).

Will it be possible to use induced reproductive incompatibility, perhaps caused by the symbiont *Wolbachia*, to control pests? This prospect has been discussed since Laven (1951) observed the impact of cytoplasmic

**Table 14.7: Some Risk Issues Relevant to Releases of Genetically Modified Insects into the Environment.**

A. Attributes of the unmodified organism
<ul style="list-style-type: none"> <li>• What is the origin of the genetically modified organism (indigenous or nonindigenous) in the accessible environment?</li> <li>• What is the insect's trophic level (predator, parasite, plant feeder) and host range?</li> <li>• What other ecological relationships does it have?</li> <li>• How easy is it to monitor and control it?</li> <li>• How does it survive during periods of environmental stress?</li> <li>• What is the potential for gene exchange with other populations?</li> <li>• Is the insect involved in basic ecosystem processes?</li> </ul>
B. Attributes of the genetic alteration
<ul style="list-style-type: none"> <li>• What is the intent of the genetic alteration?</li> <li>• What is the nature and function of the genetic alteration?</li> <li>• How well characterized is the genetic modification?</li> <li>• How stable is the genetic alteration?</li> </ul>
C. Phenotype of modified organism compared to unmodified organism
<ul style="list-style-type: none"> <li>• What is the host/prey range?</li> <li>• How fit and effective is the genetically modified strain?</li> <li>• What is the expression level of the trait?</li> <li>• Has the alteration changed the organism's susceptibility to control by natural or artificial means?</li> <li>• What are the environmental limits to growth or reproduction (habitat, microhabitat)?</li> <li>• How similar is the genetically modified strain being tested to populations previously evaluated in field tests?</li> </ul>
D. Attributes of the accessible environment
<ul style="list-style-type: none"> <li>• Describe the accessible environment, whether there are alternate hosts or prey, wild relatives within dispersal capability of the organisms, and the relationship of the site to the potential geographic range of the new strain.</li> <li>• Are there endangered/threatened species present that could be affected?</li> <li>• Are there agents that could move the new strain present in the release environment?</li> <li>• Do the test conditions provide a realistic simulation to nature?</li> <li>• How effective are the monitoring and mitigation plans?</li> </ul>

(Modified from Tiedje et al. 1989; USDA 1991; and from a discussion held at a conference on “Risks of Releasing Transgenic Arthropod Natural Enemies,” held November 13–16, 1993 in Gainesville, Florida.)

incompatibility on *Culex pipiens* populations and suggested that it could be used as a means of controlling them (Prout 1994, Sinkins and O’Neill 2000). Yen and Barr (1974) found the cause of the incompatibility in *Cx. pipiens* was due to the presence of *Wolbachia*. Experiments were conducted, but the incompatibility produced was incomplete because transmission was not fully efficient (Pal 1974). However, there is hope that genetic control by this approach could become effective. Turelli and Hoffmann (1991) reported that *Wolbachia* spread rapidly in field populations of *Drosophila simulans* in California in a “natural” experiment. Turelli et al. (1992) concluded that cytoplasmic incompatibility induced by *Wolbachia*, “therefore provides a

mechanism for introducing cytoplasmic factors into natural populations. This may eventually be useful for introducing deleterious factors into pest insect populations." Unfortunately, this is the only example to date in which *Wolbachia* has been shown to "sweep" through a *very large* population in the field, so it is unclear whether such sweeps will occur with other species and with other strains of *Wolbachia*.

Reproductive incompatibility could be transferred to a population of insects lacking *Wolbachia* by microinjection of transgenic *Wolbachia*, mass rearing the infected individuals, and release of the insects into natural populations ([Sinkins and O'Neill 2000](#)). Several questions must be resolved, including whether resistance to the cytoplasmically transmitted organisms could develop in the pest insect populations if the invasion by *Wolbachia* takes a "long" time, whether such incompatibility will be stable, and whether the *Wolbachia* will move horizontally to nontarget insect species. Appropriate release rates of individuals containing the *Wolbachia* are critical or the drive system can fail to function ([Turelli and Hoffmann 1991](#)). A description of releases of the mosquito *Aedes aegypti* containing novel *Wolbachia* strains is described in [Box 14.3](#).

Deployment of genetically manipulated arthropods is complicated if some form of reproductive isolation or drive mechanism cannot be provided when the goal is to obtain population replacement or character replacement. One of the reasons genetically modified predatory mites were successfully used in a pest-management program may be because these natural enemies disperse relatively slowly ([Hoy 2000a,b](#)). Releases of pesticide-resistant strains of natural enemies into pesticide-treated orchards or vineyards provided sufficient isolation that the genetically manipulated strains were able to establish without extensive competition from, or interbreeding with, susceptible native populations ([Caprio et al. 1991](#)). Likewise, releases of a pesticide-resistant strain of the parasitoid *Aphytis melinus* into Israeli citrus groves did not involve competition or interbreeding with susceptible populations because this species was not present in Israel. Selection with pesticides for the resistant population could provide the "drive" mechanism.

Predicting whether, and how, genetically modified pest or beneficial arthropods will establish is difficult ([Hoy 2000a,b](#)). There are at least two models that could be used in the establishment of a genetically modified strain in situations in which a "native" population exists: 1) the released strain displaces the "native" population and replaces it (replacement model). This model assumes relatively little interbreeding occurs between the released and "native" populations. 2) Alternatively, the released strain interbreeds with the "native" population and a hybrid population is produced. By appropriate strong selection, the

desired trait is selected for and the resultant population contains the desired gene (introgression model).

## 14.11 Potential Risks Associated with Releases of Genetically Modified Arthropods

Risk equals the potential for damage *and* the likelihood of its occurrence. Risk estimates may be different for pest and beneficial arthropods and may depend on whether the modified strain is expected to persist in the environment or is unable to reproduce and cannot persist (Table 14.7). Risks also will vary with the specific transgene(s) inserted, or other genetic modification obtained (paratransgenesis or insertion of a novel symbiont). It is easier to suggest *potential* types of harm than to *quantify the likelihood* of its occurrence. Each release should be considered on a case-by-case basis.

### 14.11.1 Could Gene Silencing Reduce Program Effectiveness?

There is always the risk that a genetically modified arthropod population could be released into the field and fail to function as expected due to **gene silencing** or **RNAi**. Transgenic plants and mammals often inactivate multiple copies of transgenes that overexpress proteins or are otherwise abnormal (Dorer and Henikoff 1997, Wolffe 1997, Henikoff 1998, Birchler et al. 2000, Sijen and Kooter 2000). Gene silencing is due to systems that have evolved as a means to prevent high levels of expression of TEs or viruses that can cause genetic damage to their hosts. In fungi and plants, gene silencing is associated with several mechanisms, including methylation of DNA, as well as posttranscriptional and transcriptional processes involving RNAi.

Multiple mechanisms of transgene silencing occur in *D. melanogaster* (Dorer and Henikoff 1994, 1997, Pal-Bhadra et al. 1999, Jensen et al. 1999). Methods may have to be developed to manage transgene silencing in arthropods or this phenomenon could reduce the effectiveness of released strains over time after release into the field. The use of insulators or boundary elements may limit gene silencing (Bell et al. 2001) and genetic elements such as histone deacetylase RPD3, which can counteract gene silencing in both *Drosophila* and yeast, may be useful in counteracting gene silencing (De Rubertis et al. 1996).

Gene silencing could be a positive attribute if specific genes in insects could be turned off. Gene silencing was purposefully induced in *D. melanogaster* by introducing a sequence that codes for an extended hairpin-loop RNA by *P*-mediated transformation (Kennerdell and Carthew 2000). As discussed in Chapter 9, Table 14.4, and Section 14.5.7, gene silencing by RNAi methods could be used to manage pest arthropods.

#### 14.11.2 Relative Risks

The least risky genetically modified arthropods could be the domesticated silkworm (*B. mori*), which is unable to survive on its own in the wild. Transgenic silk moths are able to produce both improved silk and proteins for pharmaceutical use (Zhao et al. 2010, Tatematsu et al. 2010). Transgenic *B. mori* are unlikely to have a negative effect on the environment because they should not be able to persist if they were accidentally released.

Transgenic pest or beneficial arthropods that are sterile and unable to reproduce should pose a lower risk than strains that are able to reproduce and persist in the environment. Transgenic pest or beneficial arthropods that are unable to persist because the environment is unsuitable during a portion of the year also could pose a low risk (McDermott and Hoy 1997, Hoy 2000a,b). Honey bees are only semidomesticated and thus can escape human management to survive in the wild, so transgenic honey bees could pose a greater environmental risk than the domesticated silk moth for this reason.

#### 14.11.3 General Risk Issues

Evaluating potential risks associated with releasing GMAs will likely include, as a minimum, the questions or principles outlined in Tables 14.2 and 14.7, but other issues may become important as we learn more about risk-assessment procedures (Foster et al. 2000, Kapuscinski 2002, Marrelli et al. 2006, 2007, Franz 2009). Current concerns can be summarized as follows:

- Is the genetically modified population stable under a variety of environmental conditions?
- Has its host or prey range been altered?
- Does it have the potential to persist in the environment?
- Will the modified strain will have unintended effects on other species or environmental processes?
- In the case of vector insects, will they cause harm to humans?

The first three questions are relatively easy to answer with a variety of laboratory experiments. The fourth issue is much more difficult to answer. Releases of GMAs in the United States are now evaluated by several regulatory agencies on a case-by-case basis. Permits are issued at present only for short-term releases in controlled situations so that unexpected outcomes might be mitigated more readily (Young et al. 2000).

#### 14.11.4 Horizontal Transfer (HT)

One risk issue that is especially difficult to quantify is the risk of horizontal transfer (HT) of transgenes, TEs, or *Wolbachia* to other organisms (Droge et al. 1998).

Our knowledge of HT and TEs only began in the 1950s when Barbara McClintock discovered TEs in maize. HT could occur from one population to another of the same species, from one species to another, or to other organisms in the environment. It is difficult to quantify this risk because we lack fundamental information on the frequencies and mechanisms of HT. Because HT is rare, effective sampling and statistical methods are especially important. The whole topic of HT has received limited scientific attention until relatively recently.

HT across species of *Drosophila* by *P* and *mariner* elements provides some data. HT of genes does occur between insect species by movement of naturally occurring TEs (Plasterk 1993). HT is thought to be rare, yet we have observed more than one such transfer within historical times in *D. melanogaster* and may have missed other examples. The *P* element invaded *D. melanogaster* populations within the past 60 years, perhaps from a species in the *D. willistoni* group. *P* elements might have been transferred between these *Drosophila* species by a semiparasitic mite (Houck et al. 1991). Another TE, *hobo*, also appears to have invaded natural populations of *D. melanogaster* around the 1960s (Bonnivard et al. 2000), the second invasion of this well-studied insect in the past 60 years.

Transfer of TE vectors from transgenic arthropods to other organisms is potentially feasible, although these transfers should occur very rarely. Recall that risk is determined by frequency of occurrence and the damage that might occur. In this case, the frequency is expected to be very low if the natural invasions represent a realistic estimate of frequency. If active TEs are purposefully released as drive mechanisms and if conversion of inactive TE vectors into active ones can occur, then the frequency could be higher than the natural rate.

It is difficult to estimate the potential damage that invasions of TEs could have on nontarget species. For example, *mariner* is widespread and the data suggest that 1) *mariner* elements have been present in arthropods for a long time, although some lineages have lost them; and 2) HT has occurred between different arthropod families and orders, although some transfers occurred so long ago that the *mariners* are degraded and inactive, probably due to a successful defense against the damage they cause to the genome (Lampe et al. 2000). Lampe et al. (2000) noted the most recent events occurred at least 100,000 years ago. The two *mariners* in the human genome probably invaded in the “past 100 million years” (Lampe et al. 2000).

We are still discovering new aspects of the evolutionary role of TEs (Brookfield 1996, Capy et al. 1996, Kidwell and Lisch 2001), which makes it difficult to predict what would happen if arthropods were released that contained

active TEs or inactive TE vectors (Petrov et al. 1995). The safest course might be to remove any introduced TE vector sequences from a transgenic strain before its permanent release into the environment to reduce the probability that the transgene will move, either within the strain or horizontally between different populations or species.

Elements other than *P* and *mariner* also move horizontally. Jordan et al. (1999) showed that a long terminal repeat retrotransposon (a different class of element than the *P* and *mariner* elements) in the *D. melanogaster* group species moved into *D. willistoni*, perhaps within the past 100–200 years.

Insect viruses could mediate HT of DNA. The *piggyBac* element was discovered embedded within the genome of a baculovirus (Fraser 2000) and another *Tc1*-like transposon was found in the *Cydia pomonella* granulovirus (Jehle et al. 1998). If HT of transgenes by viruses were to occur in the field, there is no guarantee that genes inserted into a particular species would remain within that species.

HT might even occur when DNA is eaten. Although most consumed DNA is degraded over time in the gut, it may not happen for several days (Schubbert et al. 1997, 1998). Bacteriophage DNA fed to mice can persist in fragmented form in the gut, can penetrate the intestinal wall and reach the nuclei of leukocytes, spleen, and liver cells (Schubbert et al. 1998). Fetal and newborn progeny of female mice fed such DNA during pregnancy had the phage DNA in various organs. Furthermore, the phage DNA was located in nuclei and associated with chromosomes, although the DNA had not integrated (Schubbert et al. 1998). Such an association of DNA with the chromosomes could affect normal gene function. Hohlweg and Doerfler (2001) fed mice soybean leaves containing a specific gene and analyzed its fate. This experiment provided a more natural delivery system than feeding naked DNA. The results indicated that DNA in soybean leaves was found less frequently in mice tissues than when naked DNA was fed, in part because the amount of DNA was reduced by ≈5 orders of magnitude (Hohlweg and Doerfler 2001).

Genetic engineering of insect gut symbionts might allow the movement of the inserted genes between the many types of microorganisms found within the insect gut (Watanabe and Sato 1998, Watanabe et al. 1998). *Enterobacter cloacae*, a bacterium found in the guts of insects, and *Erwinia herbicola*, a bacterium that grows on the surface of plants, grow in the guts of silk moth larvae and exchange genetic information via plasmids at very high rates (Watanabe and Sato 1998, Watanabe et al. 1998). The bacteria containing new genetic information were found in insect feces, suggesting that this method of HT is a frequent event in

nature. If gut symbionts of pest insects are transformed with antibiotic resistance genes, these genes might move horizontally to other bacteria within the insect gut. Transfer of antibiotic resistance genes to pathogens is undesirable because HT of antibiotic-resistance genes has led to a serious medical crisis. Some human pathogens are now resistant to almost all available antibiotics.

Whether HT will cause harm would certainly depend on the gene(s) transferred and its destination. The most serious harm might occur if the TE or viral vector inserted into germ-line tissues so it could be transmitted to succeeding generations. However, damage also might occur if the elements damaged somatic tissues; the movement of *mariner* in the soma reduced the lifespan of *Drosophila simulans* males ([Nikitin and Woodruff 1995](#)). The movement of retroelements into human breast, colon and testicular tissues can induce cancer or Duchene muscular dystrophy ([Capy et al. 1996](#)).

Analyses of the role of TEs in the evolution of genomes is undergoing reevaluation and it is clear that naturally occurring HT between species has provided some of the variability upon which evolution has acted, although the initial event can cause harm ([Plasterk 1993](#), [Krishnapillai 1996](#), [Britten 1997](#)). It is unlikely that the presence of a transgene in an arthropod will increase the small probability that the transgene will be transferred to another species by HT *if* the TE or viral-vector sequences used could be removed before release into the field. Even then, however, the probability of HT will not be zero.

Disabled TE vectors probably pose a relatively low risk of HT of the transgene to other organisms. However, it is possible for an inactivated vector (lacking transposase or inverted repeats) to become active through a process called **conversion**. [Peronnet et al. \(2000\)](#) showed that conversion could transform an inactive *P* element into an active one through the interaction of three different *P* elements in the genome in a three-step process. Conversion could make a transgene unstable and could, in theory, pose a risk for HT.

The potential risks of using TEs as drivers for inserting useful genes into arthropod populations should be evaluated carefully on a case-by-case basis. As noted by [Kidwell and Evgen'ev \(1999\)](#), "the transposability of mobile elements, their potential for rapid, and sometimes massive, amplification in copy number, their ability to change genomic locations, as well as their propensity for HT, makes the generalization of results from model organisms far less reliable. Extrapolation of results from one species to another must therefore be made with caution." Models of HT should be examined with care; if inappropriate assumptions or relevant data are unavailable, models can provide misleading results ([Caprio and Hoy 1994, 1995](#)).

## 14.12 Permanent Releases of Genetically Modified Arthropods into the Environment

Currently (2012), there are no guidelines for evaluating the risks of releasing transgenic arthropods for *permanent* establishment in the environment in the United States. Experience suggests that the probability that a “new” organism will become established in a new environment is small. For example, even when we are optimizing conditions to obtain establishment of natural enemies in classical biological control programs, only  $\approx 24\%$  of the introduced species actually establish. Historical examples of biological invasions of pests or of establishment of classical biological control agents demonstrate a lack of predictability in such releases ([Ehler 1990](#)).

Genetically modified arthropods could pose somewhat increased risk of permanent establishment over those posed by invasive species because they are likely to be released in very large numbers and into appropriate environmental conditions; it is likely that most invasive species enter the new environment in low numbers and may not invade optimal environments. Even so, [Williamson \(1992\)](#) speculated that the greater the genetic novelty, the greater the possibility of surprising results of invasive species. Because GMAs are “novel,” they should be considered likely to provide surprises.

Analyses of potential risks associated with GMAs should include evaluation of the survival, reproduction, and dispersal of the population and their effects on other species in the community ([Table 14.7](#)). Questions also should be asked about the inserted DNA, its stability, and its possible effect on other species should the genetic material be transferred ([Tables 14.2, 14.7](#)). In the United States, both state and federal regulatory agencies, including state departments of agriculture and the U.S. Department of Agriculture, Animal and Plant Health Inspection Service (USDA-APHIS), have to be consulted for permission to release genetically modified agricultural pests and natural enemies of agricultural pests. If the release occurs at a university, permission to release is required from the campus biosafety committee. Questions about potential detrimental effects of the genetically modified arthropods on threatened and endangered species will be asked by state and federal agencies, including the U.S. Department of Interior, Fish and Wildlife Service ([Young et al. 2000](#)), ([Figure 14.2](#)).

### 14.12.1 Models to Predict?

Can we use models to predict the outcome of releases of GMAs in pest-management programs? Many types of population and genetic models could be used in attempts to predict what will happen when GMAs are released into the



**Figure 14.2** A) A transgenic strain of the predatory mite *Metaseiulus occidentalis* (Acar: Phytoseiidae) containing a *lacZ* marker gene was released into the field on April 10, 1996, in Gainesville, FL. This was a short-term release designed to evaluate the fitness of the strain, its stability, and our ability to predict its behavior and to contain it. Because this predatory mite lacks wings and tends to stay on the release plants if provided adequate prey, these predators were released into the center row of potted plants with the outside row of pesticide-treated plants serving as “traps” to reduce the likelihood of escape from the plot. The white poles surrounding the release site contain clear sticky panels (not visible) at two heights to monitor any movement of the predators out of the plot; only two or three predators were collected on the sticky panels over the course of the experiment. At the end of the experiment, the plants were placed in plastic garbage bags and autoclaved to preclude the transgenic predators from persisting in the environment. This predator, originally from the western United States, is unable to persist in Florida’s hot, wet summers and so was “climatically contained” and could not permanently establish. Furthermore, no wild type population was present with which it could interbreed. B) Before making the release, approval was obtained from the University of Florida’s biosafety committee, the Florida Department of Agriculture and Consumer Services, the U.S. Department of Agriculture–Animal and Plant Health Inspection Service, the Florida Department of the Environment, and the U.S. Fish and Wildlife Service. Personnel from these agencies were present at the release site to ensure that the requirements of the permit were met.

environment in pest-management programs (Turelli and Hoffmann 1991, Turelli et al. 1992, Kiszewski and Spielman 1998, Tsitrone et al. 1999, Hoy 2000, Dobson 2003, Le Rouzic and Capy 2006, Alphey et al. 2007, Marshall 2009, Alphey et al. 2011, Diaz et al. 2011, Hancock et al. 2011, Matthews et al. 2011). We do not know, however, which model types are most likely to be predictive of the actual outcome of field releases because few models have been validated with empirical data.

Predicting field results from mathematical models can be difficult; for example, three models were developed to predict the success of a biological control program involving applications of fungi for control of grasshoppers and locusts (Wood and Thomas 1999). All three models fit the empirical data; one predicted sustained control at low levels after a single pathogen application, but the other two predicted that repeated pathogen applications would be necessary. Analysis of these divergent expectations demonstrated that two assumptions made by ecologists and modelers were suspect: Wood and Thomas (1999) concluded that quantitatively similar models need not give even qualitatively similar predictions (contrary to expectations and the sensitivity analysis of model predictions to parameter variation is not always sufficient to ensure the accuracy of the predictions).

Population models may lack key ingredients, such as partial reproductive isolation. For example, Caprio and Hoy (1995) developed a stochastic simulation model that varied the degree of mating bias between pesticide-resistant and -susceptible strains of natural enemies, diploidy state (diplo- or haplo-diploid), degree of dominance of the resistance allele, and degree to which mating biases extended to the hybrid progeny. The results were counterintuitive, but demonstrated that models offer insights into the complexities of population genetics and dynamics that might be overlooked. A common assumption made in many models is that all genotypes of a species mate at random, but this assumption may mask important interactions such as mating bias or partial reproductive incompatibilities. The efficacy of GMA releases could be jeopardized if mating biases exist between released and wild populations.

Empirical data generally are lacking to compare the relative usefulness of different model types in predicting population dynamics. Theoretical ecologists usually assume homogeneous and continuous populations. Metapopulation models, by contrast, assume that populations exist in patches varying in area, degree of isolation, and quality. Metapopulation biology increasingly is being recognized as relevant to our understanding of population ecology, genetics, and evolution (Hanski 1998). Recent data, and a variety of metapopulation models, indicate that spatial structure affects populations as much as birth and death rates, competition, and predation (Caprio and Hoy 1994).

### 14.13 Regulatory Issues: Releases of Genetically Modified Arthropods

Regulatory issues surrounding the releases of genetically modified arthropods into the environment have been discussed since 1992. Several issues have been raised, including how such arthropods should be contained in the laboratory until permission(s) have been given to make releases. Discussions about possible risks and how to mitigate them also have been conducted (Table 14.8). At the present time, there are no international or U.S. guidelines that would allow permanent releases of GMAs into the environment. The science is clearly ahead of the regulatory system (Pew Initiative on Food and Biotechnology 2004).

The releases of a transgenic predatory mite in Florida (Box 14.1, Figure 14.2), transgenic sterile pink bollworms in Arizona (Box 14.2), *Aedes aegypti* mosquitoes containing a novel strain of *Wolbachia* in Australia (Box 14.3), and multiple releases of transgenic *Ae. aegypti* containing a RIDL construct (Box 14.4) are described in some detail to outline the diverse scientific and regulatory issues that have been raised.

In addition to these releases, field tests using RNAi were conducted by Hunter et al. (2010) to determine whether Israeli acute paralysis virus (IAPV) disease could be reduced in honey bee colonies by feeding the bees with IAPV ds RNA. The IAPV ds RNA is being developed commercially (called Remebee) and is being evaluated by the U.S. Food and Drug Administration for registration. The results of the field test indicated that Florida colonies had increased adult bee populations if they received the Remebee (plus the virus), while the bees getting only the virus did not increase. The bees in the Pennsylvania trial did not show treatment differences. This appears to be the first field test of an RNAi disease-control strategy directed against an insect disease. Field trials have been conducted using plants expressing RNAi constructs (Ogwok et al. 2012) and discussions of regulatory issues for RNAi-expressing crop plants have occurred (Center for Environmental Risk Assessment 2011).

### 14.14 Conclusions

Genetic manipulation projects involving pest or beneficial arthropods share many problems and issues. Because the potential risks of permanent releases of GMAs into the environment raise concerns about risk, it is appropriate to release relatively low-risk examples first. This might involve the release of a transgenic beneficial arthropod that is carrying a marker gene or sterile pests that cannot reproduce in the environment (Ashburner et al. 1998, Hoy 2000a). A second release might include sterile insects, such as pink bollworms, that contain a fluorescent marker gene.

**Box 14.1 Release of a Transgenic Strain of a Predatory Mite**

The development of a germ-line transformation method in *D. melanogaster* by Rubin and Spradling (1982) instigated this project. Genetic selection of the predatory mite *Metaseiulus occidentalis* had been successful in improving the field performance of this predator in California almond orchards (Hoy 1985). However, a lack of variability within the species limited the traits that could be modified and transgenic methods offered the possibility of introducing exogenous genes into the germ line. Because regulatory issues were perceived to be critical to potential permanent releases of transgenic beneficial arthropods (Hoy 1992a,b, 1995), this project was designed in large part to investigate the regulatory and risk issues associated with such releases.

In 1990, the availability of useful genes for predatory mites was limited, so a marker gene was constructed that contained a *lacZ* construct (*E. coli beta-galactosidase* gene) and a *D. melanogaster* heatshock 70 promoter (Presnail and Hoy 1992). This construct was developed, in part, because we perceived it to present a low risk but could still serve as a model to investigate regulatory and risk issues. No TE sequences were included in the construct out of concerns that TE vectors might be unstable. Efforts to dechorionate and inject eggs (the method that was successful with *D. melanogaster*) with the plasmid failed, so a novel injection method was attempted (maternal microinjection). Gravid adult females were injected and 49% of the 405 females survived and deposited eggs; 48 lines were established, with seven of the lines producing larvae that expressed the *lacZ* construct. Presnail et al. (1997) evaluated the stability of four lines by the PCR, Southern blot analysis, and expression of mRNA. All four colonies were positive by the PCR for the construct, but only one line was positive by the two Southern blot analyses. The Southern blots indicated that two insertions of the plasmid had occurred in Line 18. However, all four lines evaluated after 100 generations produced the appropriate mRNAs, indicating that the *lacZ* construct was being expressed in all four lines.

Transmission of plasmid DNA to eggs by maternal microinjection was found in 70% of all eggs deposited by the injected females (Presnail and Hoy 1994). The persistence and presence of the plasmid DNA in both injected females and their eggs indicated maternal microinjection could be useful for delivering DNA into other phytoseiids. However, Jeyaprakash et al. (1998) found by Southern-blot analysis that, in 12 new putatively transgenic lines, at least eight of the lines had extrachromosomally transmitted plasmid DNA. In five of these eight lines, the plasmid DNA was highly amplified with an estimated 10 million copies per female. Despite the large number of copies, the extrachromosomal plasmid DNA was lost after 30 generations.

Before release of Line 18, which had two insertions of the construct in the nuclear genome, comparisons were made between it and wild-type *M. occidentalis* under laboratory conditions to evaluate fitness and potential risks (Li and Hoy 1996). Egg production, hatchability at three temperatures and four relative humidity conditions, diapause incidence, and proportion of female progeny produced were not different. Li and Hoy (1996) also showed that, at high relative humidities and high temperatures, no eggs would hatch, suggesting that *M. occidentalis* could not survive in Florida's typical summer conditions. (This species is native to western North America, where summers are dry and most agricultural production occurs under irrigation.) Li and Hoy (1996) also confirmed that the transgenic strain remained an obligatory predator and would not feed on pollen. McDermott and Hoy (1997) evaluated the persistence of nontransgenic *M. occidentalis* in field cages during the summer in Florida and discovered that the predator populations repeatedly crashed to zero. In addition, analysis with CLIMEX a population growth model that used climatic factors to determine whether a given poikilothermic species can colonize and persist in new geographic areas (<http://www.macaulay.ac.uk/dynamo/climex.htm>) also indicated this species was unlikely to establish in Florida. Furthermore, this predator was released into Florida by the millions in augmentative biological control efforts directed against spider mites in strawberries in the spring, but has never established. Therefore, we concluded that release of the transgenic strain would be unlikely to establish in Florida and therefore would be climatically contained.

Because this predatory mite lacks wings and tends to stay on plants if provided adequate prey, the transgenic predators were released into the center row of potted plants with the outside row of pesticide-treated plants serving as “traps” to reduce the likelihood of escape from the plot (Figure 14.2A). The white poles surrounding the release site contain clear sticky panels (not visible) at two heights to monitor any movement of the predators out of the plot; only two to three predators were collected on the sticky panels over the course of the experiment. At the end of the experiment, the plants were placed in plastic garbage bags and autoclaved to preclude the transgenic predators from persisting in the environment.

The short-term release was conducted in 1996 after permission was obtained from USDA-APHIS, the University of Florida’s Biosafety Committee, the Florida Department of Agriculture and Consumer Services, the Florida Department of the Environment, and the U.S. Fish and Wildlife Service. Personnel from these agencies were present at the release site to ensure that the requirements of the permit were met (Figure 14.2B). The goals were to determine the stability of the transgenic line, demonstrate that we could contain the predator in the release site, and retrieve the predators at the end of the experiment.

Despite the stability of Line 18 for > 200 generations in the laboratory, we discovered that there was a fitness cost and the *lacZ* construct was unstable under field conditions, for unknown reasons. The construct was lost from the strain within six generations, possibly due to DNA-repair mechanisms. This disappointing result demonstrates why short-term field evaluations are necessary before transgenic arthropods can be deployed in pest-management programs (Hoy 2000).

### Box 14.2 Releases of Sterile Transgenic Pink Bollworms in the United States Containing Fluorescent Marker Genes

The pink bollworm (PBW), *Pectinophora gossypiella*, became a significant pest of cotton in the southwestern United States after its invasion into North America in 1911. Efforts to control this lepidopteran pest have included the use of pesticides, the SIT, planting of *Bacillus thuringiensis* (*Bt*) transgenic cotton, cultural practices (shredding of diapausing larvae during the winter in cotton residues), and mating disruption with the PBW sex pheromone. Area-wide suppression of the PBW using these methods has resulted in population suppression. However, the PBW has been estimated to cost at least US\$32 million per year due to control costs. As a result, the National Cotton Council proposed to eradicate the PBW by using all of these control methods (Grefenstette et al. 2009). Improving the SIT program directed against the PBW is an important component. One of the improvements would be to improve the discrimination between released marked sterile male moths and wild males in the sex-pheromone traps. Marking of released sterile moths has relied on a red dye added to the larval diet, which results in moth tissues becoming red. However, the moths can excrete all of the dye and the dye can be lost from the moths in the field, making it difficult to provide an accurate estimate of the PBW population density. It would be desirable to have a permanent marker that allows program managers to reliably identify released and wild males. “If a fraction of these wild captures are actually misidentified sterile moths, due to marker failure, such releases are a waste of resources” (Walters et al. 2012). Furthermore, because released sterile moths are irradiated with substerilizing doses to minimize damage from irradiation, F<sub>1</sub> progeny are produced, although they are sterile (F<sub>1</sub> sterility). Accurate identification of these sterile F<sub>1</sub> moths is also important to monitoring the progress of the SIT program because these adults are not marked with dye, but these sterile F<sub>1</sub> males will not have been marked.

The first effort to improve the marking system was made by transforming the PBW with an enhanced green fluorescent protein (EGFP) gene regulated by an *actin* promoter using the *piggyBac* vector (Peloquin and Miller 2000, Peloquin et al. 2000a, Miller et al. 2001). A permit was given by the U.S. Department of Agriculture (USDA) Animal and Plant Health Inspection Service (APHIS) to release the transgenic strain in January 2001 and an Environmental Assessment was provided. Releases were made after the strain had been

reared in the USDA-APHIS quarantine facility for 20 generations without evidence of instability. The stability is a bit surprising because endogenous putatively intact *piggyBac*-like elements were found in the PBW (Wang et al. 2010). During laboratory rearing, no differences in length of time spent in larval instars or the pupal stage was found, although the transgenic females produced  $\approx$ 20% fewer eggs and their hatch rate was 26% lower (Miller et al. 2001). Based on the Environmental Assessment provided by USDA-APHIS, these moths were to be contained at the release site in Arizona by sterilization by radiation, having their wings clipped, and by being placed into a double set of cages isolated from other sources of cotton. Pheromone traps were to be deployed to capture any transgenic males that escaped, and any cotton was to be destroyed after the experiment. A release of a high ratio of sterilized moths in the area took place at the end of the experiment, as well. The goal was to compare the fitness of the transgenic and nontransgenic moths. No journal publications appear to be available on the results of these contained field trials of a sterilized transgenic PBW. A transgenic strain containing a red fluorescent protein gene was developed and field releases of live moths were made in 2006 (Simmons et al. 2011). The injected construct contained the marker gene regulated by a promoter fragment from a baculovirus nuclear polyhedrosis virus in the *piggyBac* vector. Four transgenic lines were produced, but only one appeared to be fit. It was reared in a quarantine facility by USDA-APHIS. The USDA-APHIS issued a release permit after conducting an Environmental Assessment and notices of the Environmental Assessment were placed in the Federal Register (Federal Register 2006).

The first test was conducted in field cages under USDA permit number 06-150-01r in 2006 in Arizona, which compared the transgenic strain and the APHIS strain of PBW. Both moth strains were treated with fluorescent powders, and both were irradiated. The two moth strains were compared for size, response to pheromone traps, and longevity. No significant differences were found, except the transgenic strain had a smaller pupal size. In 2007, larger-scale releases into three cotton fields were conducted in Arizona (USDA permit number 07-015-102r). Approximately 1.1 million sterile moths containing the fluorescent marker and 1.1 million nontransgenic sterile moths were released into three fields of conventional cotton (non-*Bt* cotton). Traps containing the sex pheromone were placed within and outside the fields to determine if the released transgenic and nontransgenic males dispersed equally and responded to traps in a similar manner. Mating performance was evaluated by comparing matings between sentinel females tethered at mating stations with transgenic or nontransgenic males and no significant differences were found in mating propensity of the two types of males. The traps captured 20% more transgenic moths than nontransgenic moths, but there was no evidence of a significant difference in “residence” time in the field between the two strains, and the mean dispersal was greater for the transgenic strain than the nontransgenic strain.

Thus, the transgenic line appeared to be sufficiently fit that it could be deployed in an eradication program. In 2008,  $>15$  million transgenic PBWs were released (USDA permit 08-105-102rm). Only transgenic sterile moths were released by air over 2500 acres of cotton in Arizona at a rate of 1 to 2 million moths per week. The goal was to evaluate the transgenic strain using an eradication program protocol. Moths in the traps were evaluated for the presence of the transgene by microscopy and DNA was isolated and evaluated by the PCR to detect the transgene. The transgenic moths were also treated with the red dye so that comparisons could be made of the efficacy of the marker gene and the dye. The transgene was reliable under field conditions (Walters et al. 2012).

#### Box 14.3 Genetic Modification of *Aedes aegypti* (and Other Mosquitoes) by Insertion of a Novel *Wolbachia* Strain (Transinfection)

The mosquito *Aedes aegypti* is a vector, along with the Asian tiger mosquito *Ae. albopictus*, of the virus that causes dengue. No drugs or vaccines are available to manage human cases of dengue, so control of mosquitoes is the primary disease-management tactic available in the tropics and subtropics around

the world. Traditional mosquito-control methods include the use of pesticides and habitat management, including the use of screens on houses and elimination of breeding sites. However, mosquitoes develop resistance to pesticides, and habitat-management methods may be difficult to implement. Dengue potentially can affect 2.5 billion people and causes an estimated 50–100 million cases each year (Bian et al. 2010).

The dengue virus has multiple serotypes and movement of people can introduce new serotypes into an area. If a person is infected with one serotype, the result is a debilitating fever called “break bone fever.” Unfortunately, subsequent infection with a different serotype causes an even more-serious disease, resulting in 1 to 20% mortality in children (Yeap et al. 2011).

Genetic modifications of mosquitoes have been proposed as one method for managing dengue. Approaches include developing transgenic strains of *Ae. albopictus* in which the nuclear genome has been modified in a manner that precludes disease transmission. Another option is to release a strain of mosquito that contains *Wolbachia* as a driver (by means of cytoplasmic incompatibility) to enhance the spread of a transgene inserted into the nuclear genome of the mosquito. A third option is to insert a strain of *Wolbachia* (called “popcorn” or wMelPop-CLA) into *Ae. aegypti* to prevent transmission of dengue (Min and Benzer 1997, McMeniman et al. 2009, Moreira et al. 2009a, Walker et al. 2011, Yeap et al. 2011). The wMelPop strain of *Wolbachia* was found originally in *Drosophila melanogaster*, where it multiplies excessively and reduces the life span of the fruit fly. It was called the “popcorn” strain because it produces large holes in the brain of the flies.

The O’Neill laboratory in Australia has developed *Ae. aegypti* containing the wMelPop strain of *Wolbachia*. Their research resulted in releases into the field in Australia in 2011 and 2012, and plans are underway to conduct trials in other countries. The project progressed because McMeniman et al. (2009) introduced a life-shortening *Wolbachia* strain (wMelPop) into *Ae. aegypti* after rearing the *Wolbachia* in cell lines of the mosquito over a 3-year period to adapt the novel strain to the mosquito. The infection causes cytoplasmic incompatibility between infected and uninfected lines and reduces the longevity of infected females by half; epidemiology suggests that such a reduction in longevity could result in reduced transmission of pathogens. Transmission of the novel *Wolbachia* strain to progeny by infected females is high (>99.7%).

Moreira et al. (2009a) found that the wMelPop strain of *Wolbachia* limited infection with dengue, Chikungunya virus, and *Plasmodium* in *Ae. aegypti*, and speculated that this reduced-infection status and the life-shortening effects shown by McMeniman et al. (2009) could provide a method for controlling disease transmission. Moreira et al. (2009b) showed that aging *Ae. aegypti* females infected with this novel *Wolbachia* strain have difficulty in blood feeding. Turley et al. (2009) quantified the decrease in blood-feeding success in the infected mosquitoes. Again, this was speculated to reduce disease-transmission capability. Evans et al. (2009) showed that mosquitoes infected with this *Wolbachia* strain also had increased locomotor activity and increased metabolism, for unknown reasons. The increased activity levels and metabolic activity of the infected mosquitoes suggests they are fit and could potentially compete with field populations if released. At the cellular level, Frentiu et al. (2010) showed that the wMelPop strain of *Wolbachia* reduced replication of dengue virus in mosquito cells in relation to the density of the *Wolbachia*; these results were comparable to those of Bian et al. (2010) who studied whole mosquitoes. Thus, “although *Wolbachia* infection imposes a metabolic cost that can lead to lowered host fitness, in certain circumstances, such as viral infections and nutritional stress, these bacteria also provide a fitness benefit” (Frentiu et al. 2010). Hussain et al. (2011) discovered that *Wolbachia* uses host microRNAs to manipulate host gene expression.

Yeap et al. (2011) evaluated the effects of wMelPop on *Ae. aegypti* after the infected line was crossed with a population from Cairns, Australia to increase the competitive ability of the infected (laboratory) line. The outcrossed lines were evaluated for transmission of *Wolbachia*, longevity, egg viability, cytoplasmic incompatibility, larval nutrition, and developmental time. Genetic variability was assessed using eight molecular markers. The results indicated that the out-crossed lines had increased genetic variability, infected

females transmitted *Wolbachia* at a high rate, and the hatch rate of eggs in the incompatible crosses was <1%. Even when infected mosquitoes were reared under low food levels and high densities, they still maintained their decreased longevity and CI levels. Infected lines had a reduced egg hatch rate at different temperatures. These laboratory data provided “parameter estimates for predicting invasion by a *Wolbachia* infection for dengue suppression when placed on an outbred background. Outbreeding ensures that release material has a genetic background that matches the target population. Our parameter estimates suggest that wMelPop invasion is possible under humid conditions. However, because of deleterious effects associated with wMelPop-CLA, invasion under dry conditions will be difficult ...” (Yeap et al. 2011). The group subsequently developed another line of mosquitoes infected with a different strain of *Wolbachia*.

Walker et al. (2011) adapted the wMel strain of *Wolbachia* from *D. melanogaster* in *Ae. aegypti* cell culture over 2 years. Stably infected *Ae. aegypti* lines were generated by microinjection into embryos. An outbred line was established by crossing an infected laboratory population with wild-caught mosquitoes. The out-crossed strain induced strong cytoplasmic incompatibility and fitness was not significantly different to that of uninfected mosquitoes under semifield conditions. Walker et al. (2011) next showed that the wMel strain of *Wolbachia* blocked dengue transmission and, under semifield conditions, increased from a starting frequency of 0.65 to near fixation within a few generations, invading *Ae. aegypti* populations at an accelerated rate relative to trials with the wMelPop-CLA strain.

Two towns in Queensland were selected to test the release of mosquitoes infected with *Wolbachia* (Enserink 2010a). In the field trial, mosquitoes were released weekly for 12 weeks, with the goal of determining how well the *Wolbachia*-infected mosquitoes are able to spread the *Wolbachia* into the wild population. Clear regulatory authority to evaluate these releases was not present, but the releases were conducted after the Commonwealth Scientific and Industrial Research Organization (CSIRO 2010) conducted a risk analysis and concluded “negligible risk” that infected mosquitoes would cause harm. The Australian Pesticides and Veterinary Medicines Authority also evaluated the proposed releases. Hoffmann et al. (2011) reported the results of the releases into two sites near Cairns, Australia. Prior to releases, water was removed from breeding containers in the vicinity of the homes, and in early January 2011 adults (both males and females) were released at 184 and 190 locations at each site, followed by nine releases over the following 9–10 weeks for a total of 141,600 and 157,300 adults, respectively. Mosquitoes were monitored for *Wolbachia* infection every two weeks and the infection frequency increased to >15% in both sites within 2 weeks after release. In one site, the infection frequency reached near fixation five weeks after releases were terminated and reached 90% at the same time interval in the second site. Analysis suggested that fitness costs caused by the *Wolbachia* infection could be ≈20% in the field (Hoffmann et al. 2011). The authors did not evaluate the role of these infected mosquitoes in transmission of dengue but, “demonstrated that it is possible to produce *Wolbachia*-infected mosquito populations that can act as ‘nursery’ areas for future human-assisted collection and further dispersal of *Wolbachia*-infected mosquitoes, without the need to rear additional mosquitoes in an insectary. This should provide a strategy for sustainable dengue control at low cost, with a relatively simple deployment system suitable for implementation in developing countries.” Additional releases of *Wolbachia*-infected *Ae. aegypti* in Australia were conducted in 2012 using the wMelPop-CLA strain (Cyranoski 2012). Updates of the project are available at the website [www.eliminatedengue.com](http://www.eliminatedengue.com).

Jeffery et al. (2009) report on efforts to characterize the *Ae. aegypti* population in a Vietnamese village before the release of mosquitoes containing a novel *Wolbachia* strain following the releases in Australia. Such a release will raise regulatory issues regarding movement of genetically modified insects across international boundaries (Lehane and Aksoy 2012).

One concern is that if *Ae. aegypti* is no longer an effective vector, the second vector, *Aedes albopictus*, could continue to transmit dengue. However, *Ae. albopictus* also has been infected with the wMel strain of *Wolbachia*, which causes cytoplasmic incompatibility and blocks dengue transmission. Thus, both of the main vectors of dengue might be amenable to replacing pest populations with genetically modified

mosquitoes incapable of transmitting dengue (Blagrove et al. 2012). Lu et al. (2012) compared the density of three types of *Wolbachia* in *Ae. aegypti* and *Ae. albopictus* and found the *Wolbachia* titer is too low in *Ae. albopictus* to induce resistance to dengue, perhaps because *albopictus* is naturally infected with two strains of *Wolbachia* and has developed mechanisms to suppress replication of *Wolbachia*.

The use of novel *Wolbachia* strains could potentially be useful in control of malaria transmission. Jin et al. (2009) demonstrated that the wMelPop strain of *Wolbachia* establishes in the malaria vector *Anopheles gambiae*, although it was not virulent to the mosquito. Hughes et al. (2011) injected *An. gambiae* with the wMelPop and wAlbB strains of *Wolbachia* and found both infections to be virulent and to inhibit *Plasmodium* infections. The wMelPop strain is unusual in that it is virulent to blood-fed females. Kambris et al. (2009) found that wMelPop up-regulates the immune system of *Ae. aegypti* and inhibits the development of filarial nematodes, so could be of potential use in the effort to control lymphatic filariasis.

Is the use of *Wolbachia* infection to control disease transmission likely to succeed? It is not clear, and Sabesan and Jambulingam (2012) discussed some of the questions and possible limitations to this approach. They noted, “it is not known for how long and to what extent these exercises are required to be implemented to achieve the desired target (infection in wild population) because the vertically inherited parasites such as *Wolbachia* are predicted to evolve towards reduced virulence over time.” They also were concerned that maintaining the *Wolbachia* infection in field populations, “where the environment governs the balance and is beyond human control” could be difficult, especially because, “With the increases in temperature that have been widely experienced in many areas, viral pathogens such as dengue could easily reach its infective titre in a shorter duration in the field by simple multiplication.” They conclude, “At the moment, we do not have much to claim to success in our war against mosquito vectors, except in situations where there has been vector habitat destruction...The mosquito vector host(s) and parasites and/or endosymbionts will try to adapt themselves to maintain a balance in nature.”

Koehncke et al. (2009) note that *Wolbachia* rarely cospeciates with their hosts, indicating that infections are lost in host species over time. They modeled the spread of potential mutants that would repress *Wolbachia* action either by affecting bacterial transmission or the level of cytoplasmic incompatibility and show that host mutations spread, even at a cost to host males, ultimately leading to loss of *Wolbachia* infections. They concluded that *Wolbachia* must move horizontally to survive. Vavre and Charlat (2012) suggest future modeling and research is necessary to resolve some of these important issues. There is a possibility that mosquitoes, or the pathogens, will evolve adaptations so that the *Wolbachia* no longer functions as expected. Enserink (2010a) quoted O'Neill as responding to this concern, “With some luck, that will take at least a couple of decades ... during which time humans may have developed vaccines or another way to thwart their enemies.” Thus, this method of pest control, like the use of chemical pesticides, is perceived to be potentially subject to selection for resistance in the hosts (or to reduced effects on the host by the *Wolbachia*).

What if something goes wrong? Pesticide applications could reduce the mosquito populations. In addition, Dobson (2003) evaluated whether it would be possible to reverse *Wolbachia*-based population replacement in case of unexpected consequences. He concluded, based on models, that it would be possible to slow, stop, or reverse the spread of a released strain of *Wolbachia* if a strain was released that contained a *Wolbachia* type that was bidirectionally incompatible with the *Wolbachia* driver strain. It would also be possible to reverse spread of an undesirable *Wolbachia* strain by unidirectional incompatibility, if a superinfected strain were released. The effectiveness of these approaches depends on the degree of cytoplasmic incompatibility, the rate of transmission of the symbiont to progeny (it is not always 100%), and the release rates. The effect of the *Wolbachia* strain on fecundity of the host females, and the migration rate of the host population are relevant. Rasgon (2008) also evaluated models to optimize the use of *Wolbachia* in managing vector-borne diseases.

This case study demonstrates that regulatory protocols are limited with regard to permanent releases of genetically modified arthropods into the environment. Marshall (2011) discusses the issues related to

the releases of *Wolbachia*-infected mosquitoes under the Cartagena Protocol on Biosafety with regard to concerns about the international movement of genetically modified organisms, noting that this release uncovers a gap in that regulatory machinery. The mosquito naturally occurs in Australia and the mosquito genome itself is not genetically modified (although it has a novel strain of *Wolbachia* artificially inserted into it, so is genetically modified) (De Barro et al. 2011). Existing regulation in Australia was used to evaluate possible risk although the case, “is an example of how science is leading to advances that outstrip existing regulatory frameworks” (De Barro et al. 2011). Lehane and Aksoy (2012) point out that international regulation of transgenic and paratransgenic insects is needed, because insects can disperse from the original release sites (see Table 14.8).

#### Box 14.4 Field Releases of Transgenic Mosquitoes Containing a RIDL Construct

The sterile insect technique (SIT) or sterile insect release method (SIRM) was described in Chapter 10. This genetic-control method involves mass rearing, sterilization by chemicals or irradiation, and release of large numbers of males that mate with wild females (Knippling 1955, Pal and Whitten 1974, Whitten 1985, Tan 2000). Sterile males must compete with wild males and reduce reproduction by wild females; in some cases the goal is *eradication* and in other programs it is to *suppress* the pest population. Ultimately, the goal is to reduce crop damage or transmission of insect-vectored diseases. SIRM programs have been highly effective in the eradication of the screwworm in North and Central America, and of the Mediterranean fruit fly in Florida and other locations (Wyss 2000). However, genetic-control projects are expensive because a large number of sterile males need to be released to compete with wild males (up to 100 sterile males to 1 wild male) (LaChance 1979). This release ratio is needed because sterilization typically is accomplished by whole-body irradiation and sterile males are not as competitive as wild males. It is undesirable to release females (because they can bite, transmit disease, or cause damage to crops), so a method of eliminating females early in the rearing program would reduce costs and eliminate these concerns. Classical genetic methods using mutations and translocations have resulted in sexing strains that produce mostly males (LaChance 1979, Wyss et al. 2000, Marec et al. 2005, Papathanos et al. 2009, Black et al. 2011), but these strains can revert to wild type under factory rearing conditions if strict quality-control methods are not maintained. The colonization and mass rearing of insects, and especially mosquitoes, is difficult, expensive, and must preserve fitness in released males that are essential for competitive mating with wild females (Benedict et al. 2009).

The development of transgenic strains of insects has been proposed to overcome these difficulties: strains that have a conditional female-specific lethality, or strains that have males that are sterile without undergoing irradiation could provide improved cost-effectiveness for SIRM programs. Thomas et al. (2000) proposed using a dominant, repressible, lethal genetic system as a genetic-control method called **RIDL**, or Release of Insects carrying a Dominant Lethal. Oxitec, a company producing and selling RIDL mosquitoes, has called RIDL “biological control,” but biological control commonly is defined as the use of predators, parasitoids or pathogens to control a pest. Oxitec has called the RIDL mosquitoes “sterile insects” (Phuc et al. 2007, Fu et al. 2010, Harris et al. 2011), but sterile insects should be unable to produce viable larvae. In fact, RIDL mosquitoes are transgenic, and the mosquitoes released produce progeny by matings between wild females and transgenic males, with the majority of larvae or pupae dying because there is no tetracycline in the local environment (Phuc et al. 2007). Due to “leakage,” a few (up to 15%) of the progeny do survive. Black et al. (2011) discussed “Why RIDL is not SIT” and pointed out that RIDL actually “refers to a whole suite of different genes and strategies including bisex lethals, flightless females and non-sex specific late-acting lethal systems.” RIDL requires transgenesis and is a genetic-control method.

Transgenic *Aedes aegypti* were transformed using a TE vector (*piggyBac*) and contain a red fluorescent marker gene (DsRed2), as well as a repressible dominant lethal gene (tetracycline-repressible

transcriptional activator [tTAV]) (Phuc et al. 2007). In the absence of tetracycline, the tTAV binds to a minimal promoter and drives expression of more tTAV in a positive feedback loop, resulting in low-level expression. When tetracycline is present, tTAV binds tetracycline, which does not bind to the binding site and does not lead to expression of more tTAV. High-level expression of tTAV is toxic, so its expression provides a tetracycline-repressible lethal system (Phuc et al. 2010). The exact mechanism of toxicity is unknown. The tTAV construct is repressed during mass rearing in the presence of tetracycline, but is active in the field when tetracycline is lacking.

The first release of transgenic *Ae. aegypti* containing a RIDL construct took place in the Cayman Islands in fall 2009. Releases by the Oxitec company were not announced until November of 2010, which created concerns about a lack of transparency (Enserink 2010b). A subsequent release was made the following summer. The Cayman Islands apparently lacked a biosafety law at the time of the releases, but the Mosquito Research and Control Unit of the Cayman approved the releases. However, there were no town hall meetings or public debates prior to the releases and, although information was sent to local newspapers, Enserink (2010) reported there was no mention that the mosquitoes were transgenic, so informed consent by the inhabitants was not obtained.

Harris et al. (2011) reported the results of releasing the RIDL mosquitoes into 10 hectares of the Cayman Islands. They indicated that the 465 male mosquitoes released per hectare per week over a four-week period mated with wild females and fertilized their eggs, based on analysis of eggs that hatched and produced fluorescent larvae. 126 fluorescent larvae were found in the egg traps, which was 9.6% of the 1316 larvae found. PCR analysis of DNA from trapped males indicated there were 20 released males and 105 wild males, or the RIDL males comprised  $\approx$ 16% of the total males trapped during this period, assuming equal trapping efficiency. No information was provided as to the persistence of the released mosquitoes after the trial concluded, or whether dengue transmission was affected. It is known that up to 15% of RIDL mosquito progeny may survive and it is also known that some environments contain tetracycline, an antibiotic that is widely used in medicine, and for veterinary and livestock purposes, and can be found in wastewater and sewage. The Oxitec website claims that environmental levels of tetracycline are too low to allow survival of the transgenic mosquitoes, but it is not clear if that is always true.

Subbaraman (2011), a reporter for *Nature Biotechnology*, reported that some environmentalists are worried, “that transgenic insect releases that reduce wild mosquito numbers might not only create an ‘empty niche’, which other potentially damaging insects might fill, but also affect organisms higher in the food chain that rely on mosquitoes as a dietary source.” Subbaraman (2011) reported, “... the company liberated about 3.3 million sterile male transgenic *Aedes aegypti* mosquitoes into a region spanning about 16 hectares through 80 releases,” results which were not included in the report by Harris et al. (2011). Subbaraman (2011) reported on a presentation at a scientific meeting by Luke Alphey of Oxitec that, “... results from the large release showed up to an 80% reduction in the numbers of wild mosquitoes  $\sim$ 11 weeks after the release. This reduction in the population was sustained for a further  $\sim$ 7 weeks.” It is not clear whether there was a control for this experiment or if dengue transmission was affected. Nor is it clear that informed consent was obtained from the local inhabitants.

Releases of *Ae. aegypti* containing the RIDL construct were made in Malaysia during 2010–2011 and in Brazil in 2011 and 2012. Lacroix et al. (2012) reported on the release of fluorescent powder-marked RIDL and wild-type male mosquitoes into the field near Bentong in Malaysia during December 2010. Males were to be captured in adult traps up to 96 or 328 m away, and eggs were to be evaluated from ovitraps. Lacroix et al. (2012) reported that 50% of the RIDL mosquitoes and 17% of the wild-type mosquitoes were recaptured. The last recaptures were made on days 9 and 12 for the two strains, indicating that the transgenic males survived well in the field. The ovitraps had no fluorescent larvae, indicating that no mating occurred between the RIDL males and wild females, or that the mated females failed to deposit eggs in the ovitraps. According to the Malaysian website, no recaptures were obtained after 3–5 January,

so the trial was stopped and the area was fogged to kill the mosquitoes (Malaysian Institute for Medical Research website [www.imr.gov.my/component/content/article.html?id=1119](http://www.imr.gov.my/component/content/article.html?id=1119)).

The Malaysian website noted that 46,171 dengue cases occurred in Malaysia in 2010 and larvicing, space spraying with insecticides or fogging, public education or legally enforced breeding-site reduction have not stopped the spread of the disease. However, the website calls the RIDL male mosquitoes “sterile,” which is not accurate. RIDL mosquito progeny are produced, although most die at the late larval or pupal stage. Public information was provided to the public before the releases were made, but received a “low turnout of response from the public,” so it is not clear how informed consent was obtained from people in the affected area. Public briefings were given to the local news media and to a variety of government officials.

Luisa Reis de Castro published an M.S. thesis on “Big Issues Around a Tiny Insect: Discussing the Release of Genetically Modified Mosquitoes (GMM) in Brazil and Beyond.” She reported that Margaret Capurro, a professor at the University of Sao Paolo in Brazil, has led the releases of *Ae. aegypti* containing the RIDL construct. The mosquitoes were reared at Moscamed and permission to release the transgenic mosquitoes in five sites was obtained on December 17, 2010. During February 2010, the first releases were made and additional releases were made with plans to continue for 18 months.

James et al. (2011) reviewed both the Oxitec transgenic mosquito releases and the *Wolbachia* mosquito trials in Australia (Box 14.3) and pointed out, “Most vector biologists agree that success is reflected ultimately in reduced morbidity and mortality. Some consider entomological endpoints, such as local elimination of the principal vector species or complete introgression of a gene or symbiotic species that causes pathogen refractoriness, as surrogate markers for impact on infection and disease. However, those familiar with trials of conventional interventions (vaccines, drugs, and insecticides) maintain that sustained epidemiological and clinical impact should be the primary efficacy endpoint. Observations that dengue transmission can sometimes continue even with low mosquito population densities are cited as a reason for vector biologists to conduct trials to measure the incidence of infection and/or disease.” They also note, “Dengue transmission can be endemic, epidemic, multiyear episodic, and unpredictable, so that trials may have to continue for years. Trials also must encompass large geographic areas to ensure that there is sufficient human infection to detect differences between control and treated populations.” Knols et al. (2007) discuss the potential for transgenic mosquitoes and conclude that deployment of genetically modified mosquitoes “requires competencies beyond the field of biology, and the future of transgenic mosquitoes will hinge on the ability to govern the process of their introduction in societies in which perceived risks may outweigh rational and responsible involvement.” Lavery et al. (2008) also noted there are complex ethical, social, and cultural considerations for the site selection for research with genetically modified mosquitoes.

For any pest-management program, it is relevant to consider all alternatives (Gravitz 2012). Potential mosquito-control techniques include indoor residual spraying, insecticide-treated bed nets, odor-baited insecticide traps, pesticides applied to water where mosquitoes breed, fungus (*Beauveria*) applications that kill mosquitoes, infection with bacteria that preclude parasite transmission, and genetically engineered (transgenic) mosquitoes or vaccines. Other options involved the building of more-modern houses with solid roofs and screens on their windows and doors, as well as modifications of the landscape that result in habitat that is less suitable for mosquito production. Each approach has advantages and disadvantages (Gravitz 2012). At this point, it is unclear whether RIDL mosquitoes can provide sufficient control to reduce dengue transmission.

Another possible solution is the development of new vaccines for prevention of dengue (Dengue Vaccine Initiative 2012). The Dengue Vaccine Initiative (2012) reported “very substantial progress has been made in the clinical development of several vaccines and dengue vaccine development has advanced to a point where it now seems likely that at least one vaccine will be available for use in developing countries, perhaps as early as 2014.” Clearly, as in all pest-management programs, reliance on a single tactic is unlikely to be sustainable and multiple tools will be needed to manage the pest unless eradication of the vector can be achieved.

**Table 14.8: Selected References on Potential Risks and Regulatory Issues Relevant to Releasing Genetically Modified Arthropods into the Environment.**

Potential social, ethical, and risk issues and possible mitigations
<ul style="list-style-type: none"> <li>The planned introduction of genetically engineered organisms: ecological considerations and recommendations (Tiedje et al. 1989).</li> <li>Environmental risks from the release of genetically modified organisms (GMOs)—the need for molecular ecology (Williamson 1992).</li> <li>Reducing the risks of nonindigenous species introductions: guilty until proven innocent (Ruesink et al. 1995).</li> <li>Persistence and containment of <i>Metaseiulus occidentalis</i> (Acari: Phytoseiidae in Florida: risk assessment for possible releases of transgenic strains (McDermott and Hoy 1997).</li> <li>North–south research partnerships: the ethics of carrying out research in developing countries. (Edeger 1999).</li> <li>Deliberate introductions of species: research needs. Benefits can be reaped, but risks are high (Ewell et al. 1999).</li> <li>Managing risks of arthropod vector research (Aultman et al. 2000).</li> <li>Ecological and community considerations in engineering arthropods to suppress vector-borne disease (Spielman et al. 2002).</li> <li>Reversing <i>Wolbachia</i>-based population replacement (Dobson 2003).</li> <li>Ethical, legal, and social issues of genetically modified disease vectors in public health (Macer 2003).</li> <li>The first releases of transgenic mosquitoes: an argument for the sterile insect technique (Benedict and Robinson 2003).</li> <li>Biosafety and risk assessment in the use of genetically modified mosquitoes for disease control (Toure et al. 2004).</li> <li>Reversible introduction of transgenes in natural populations of insects (Le Rouzic and Capy 2006).</li> <li>Improving the ecological safety of transgenic insects for field release: new vectors for stability and genomic targeting (Handler et al. 2007).</li> <li>Transgenic mosquitoes and the fight against malaria: managing technology push in a turbulent GMO world (Knols et al. 2007).</li> <li>Guidance for contained field trials of vector mosquitoes engineered to contain a gene drive system: recommendations of a scientific working group (Benedict et al. 2008).</li> <li>Ethical, social, and cultural considerations for site selection for research with genetically modified mosquitoes (Lavery et al. 2008).</li> <li>The effect of gene drive on containment of transgenic mosquitoes (Marshall 2009).</li> <li>The <i>Cartagena Protocol</i> and genetically modified mosquitoes (Marshall 2010).</li> <li>The <i>Cartagena Protocol</i> in the context of recent releases of transgenic and <i>Wolbachia</i>-infected mosquitoes (Marshall 2011).</li> <li>Recombination technologies for enhanced transgene stability in bioengineered insects (Schetelig et al. 2011).</li> <li>Scientists and public involvement: a consultation on the relation between malaria, vector control and transgenic mosquitoes (Boete 2011).</li> </ul>

(Continued)

**Table 14.8: (Continued)**

- Safe and fit genetically modified insects for pest control: from lab to field applications (Scolari et al. 2011).
- *Problem Formulation for the Environmental Risk Assessment of RNAi Plants* (Center for Environmental Risk Assessment 2011).
- Scientists and public involvement: a consultation on the relation between malaria, vector control and transgenic mosquitoes (Boete 2011).
- What ails *Wolbachia* transinfection to control disease vectors? (Sabesan and Jambulingam 2012).

#### **Regulatory issues**

- Proposed guidelines for research involving the planned introduction into the environment of organisms with deliberately modified hereditary traits: Notice, *U.S. Federal Register* Vol. 56(22), Friday, February 1, 1991, pp. 4134–4151.
- Laboratory containment of transgenic arthropods (Hoy et al. 1997).
- Science and the precautionary principle (Foster et al. 2000).
- *Status and Risk Assessment of the Use of Transgenic Arthropods in Plant Protection* (FAO/IAEA 2002).
- *The Cartagena Protocol on Biosafety* (Convention on Biological Diversity 2003).
- Bioconfinement of Animals: fish, shellfish, and insects (National Academy of Sciences 2004).
- *Bugs in the System? Issues in the Science and Regulation of Genetically Modified Insects* (Pew Initiative on Food and Biotechnology 2004).
- This report concludes that the “federal government [of the USA] lacks a coordinated regulatory approach to ensure that all GM insects are reviewed for potential environmental, agricultural, food safety, and public health risks. The regulations issued to date on genetically modified insects only cover plant pests.” “... the issue is not so much the lack of legal authority as whether those authorities will be used in a coordinated way to ensure an adequate and credible regulatory review of all relevant risks. In the absence of such a coordinated policy framework, it is currently impossible to say whether federal regulation adequately protects against possible public health, environmental, agricultural, and food-safety risks.”
- *Status and Risk Assessment of the Use of Transgenic Arthropods in Plant Protection* (2006 FAO/IAEA).
- *Guidelines for Importation and Confined Field Release of Transgenic Arthropods in NAPPO Member Countries* (NAPPO 2007).
- Guidance for contained field trials of vector mosquitoes engineered to contain a gene drive system: recommendations of a scientific working group (Benedict et al. 2008).
- *Defining Environmental Risk Assessment Criteria for Genetically Modified Insects to be Placed on the EU Market* (Benedict et al. 2010).
- The *Cartagena Protocol* and genetically modified mosquitoes (Marshall 2010). This article describes gaps in the Protocol regarding genetically modified mosquitoes.
- Science, regulation, and precedent for genetically modified insects (Mumford 2012).
- Scientific standards and the regulation of genetically modified insects (Reeves et al. 2012).

Releases in the United States likely will involve a two-step process. Releases in the United States of transgenic arthropods were experimental and on a relatively small scale, at least initially ([Boxes 14.1, 14.2](#)). Permanent releases of GMAs in the United States have not been made. Nor is it clear which regulatory agency(ies) in the United States will regulate releases of transgenic arthropods, such as mosquitoes, that vector human diseases. The “Coordinated Framework” that regulates transgenic organisms has gaps, making regulation of transgenic vectors of human diseases problematic ([Pew Initiative on Food and Biotechnology 2004](#)). Uniform regulations regarding appropriate facilities and procedures for containing transgenic arthropods before their release into the environment also are lacking. It is unknown what issues must be resolved before permanent releases of transgenic insects into the environment are possible ([Pew Initiative on Food and Biotechnology 2004](#)). Risk analyses will add a significant cost in both time and resources to pest-management projects involving GMAs.

The issue is not *if* GMAs should be released, but when and how? The debate over evaluation methods and risk issues should include a variety of viewpoints. Much of the debate has been parallel to the debate on the risks of introducing natural enemies for classical biological control programs into new environments ([Ruesink et al. 1995](#)). Most introduced insect natural enemies have provided large benefits, with only a few examples of potential or demonstrated harm to the environment. Despite this, caution is warranted. [Ewell et al. \(1999\)](#) reported the conclusions of a workshop on the risks of deliberate introductions of species into new environments. The participants did not discriminate between the potential risks of genetically modified organisms and unmodified organisms. [Ewell et al. \(1999\)](#) noted that assessing risks is complex and concluded,

“Benefits and costs of introductions [of new organisms] are unevenly distributed among ecosystems, within and across regions, among sectors of society, and across generations. Although an introduction may meet a desired objective in one area, at one time, or for some sectors of society, unwanted and unplanned effects may also occur. Introduced organisms can, therefore, simultaneously have both beneficial and costly effects. Furthermore, the relative magnitudes of costs and benefits vary both in space and over time.”

[Ewell et al. \(1999\)](#) recommended developing a single framework for evaluating all types of introductions, noted there is a need for retrospective analyses of past introductions, and the importance of having a holistic view of the invasion process. [Ewell et al. \(1999\)](#) concluded,

“At the extremes, these views [of risks] range from a handful of advocates of no introductions, or of such rigorous pre-introduction proof of benignness that all introductions are effectively prohibited, to an equally small group

that advocates a freewheeling global eco-mix of species....most proponents of purposeful introductions understand the risks (but believe that technology can deal with them), and most conservation biologists recognize the potential benefits to be derived from carefully controlled introductions. Clearly, there is a need to bring all parties together on common ground that can lead to objective, science-based decisions by policymakers."

Science-based discussions of the potential risks associated with permanent releases of GMAs for pest-management programs are urgently needed. People with diverse viewpoints and vision should conduct such discussions. Inappropriate releases and unintended consequences could detrimentally affect all projects involving GMAs. More funding and effort should be devoted to research on risk-assessment methods.

The potential value of GMAs to practical pest-management problems is often discussed in terms of the social, public health, and economic costs associated with malaria and other arthropod-borne diseases. Traditional pest-management programs for insect vectors of diseases have serious limitations or have failed, especially for malaria. For example, despite enormous efforts, malaria is an ongoing and important health problem, with at least 400 million people falling ill with malaria each year. As many as one million people die each year, especially children younger than 5 years of age in Africa ([Marshall 2000, Marshall and Taylor 2009](#)). Will deployment of GM mosquitoes unable to vector malaria contribute to a solution to malaria? Or will vaccines and drugs be developed that will mitigate the problem? There can be no clear answers without carrying out the appropriate experiments, and different scientists have very divergent opinions as to how scarce research funds should be spent.

[Miller \(1989\)](#) reviewed malaria-control strategies and pointed out that it is unlikely there will be "a magic bullet that will eliminate malaria." He noted, "Even DDT could not be called such a weapon, at least in retrospect" due to the problem of resistance to the pesticide. The strategy of relying on single tactics in pest management, whether it is control of mosquitoes or of agricultural pests, ultimately fails. Relying on genetic modification as a sole tactic is unlikely to be an exception. For example, the complexity of the genetic structure in *An. gambiae* populations in West Africa ([Lanzaro et al. 1998](#)), which may be reflective of the complex genetic architecture of many arthropods, suggests that release programs involving a single GM strain are unlikely to be successful. The integration of several compatible tactics has been more sustainable than relying upon a single management tactic; multitactic management of medically important disease vectors also is more likely to be sustainable ([Miller 1989](#)).

Past experience with natural enemies genetically manipulated by traditional selection methods suggests that the most readily implemented pest-management projects using GM natural enemies will be those where releases can be conducted in relatively small areas such as temporary cropping systems, or where the natural enemy has a low dispersal rate and can be established in individual orchards, or where the natural enemy is released into a geographic region where the wild strain does not occur. The most difficult projects to implement are likely to be those in which the transgenic strain is expected to replace the endemic population. Projects that require the GM strain to replace a wild strain may require very strong selection or "drivers." Teams of experts may have to develop the mass-rearing technology, quality-control methods, and the necessary information on population structure and hidden partial reproductive isolation mechanisms that are likely to occur.

Other questions remain. For example, is informed consent of humans in the release zone required before GM mosquitoes are released that are unable to transmit human diseases? Who bears the burden of liability should environmental, or other, harm occurs after releasing GMAs? Scientists working for nonprofit organizations, such as universities and governmental agencies, are developing most transgenic arthropods. Will the commercial development and sale of transgenic arthropods create new ethical or legal issues?

Significant and rapid advances have been achieved in GM of arthropods and the identification of potentially useful genes to insert. New opportunities should arise over the next few years to identify useful genes now that more arthropod genome projects have been completed (see Chapter 7).

There remains controversy as to the role that transgenic mosquitoes could play in controlling or eradicating diseases such as malaria ([Marshall et al. 2010](#)). [Curtis \(2000\)](#) presents one view:

"There is much excitement about transgenesis as a way to generate strains of mosquito that cannot transmit malaria. If a single dominant gene with these properties could be engineered, this would be an improvement on *Plasmodium* nonsusceptible strains that have already been selected by old-fashioned breeding techniques. However, without extremely reliable systems for driving the transgenes into wild vector populations, possession of a nontransmitter strain would be of no practical use. Even if a totally reliable gene-driving system were produced, there might well be strong political objections to the irrevocable release of genetically manipulated insects that bite people."

Deploying any GMA in a pest-management program is an awesome challenge, requiring risk assessments, detailed knowledge of the population genetics, biology, and behavior of the target species under field conditions, as well as coordinated

efforts between molecular and population geneticists, ecologists, regulatory agencies, pest-management specialists, and sustained efforts to educate the public about the benefits and potential risks of releasing GMAs into the environment. Risks associated with releasing arthropod vectors of human disease require evaluation by vector-borne disease specialists and ethicists, as well as regulatory agency officials (Aultman et al. 2000). When a mosquito, or other vector of human or animal disease, is genetically modified, such risk assessments may be especially controversial.

## References Cited

- Aksoy, S., Maudlin, I., Dale, C., Robinson, A.S., and O'Neill, S.L., (2001). Prospects for control of African trypanosomiasis by tsetse vector manipulation. *Trends Parasitol.* **17:** 29–35.
- Aksoy, S., Weiss, B., and Attardo, G., (2008). Paratransgenesis applied for control of tsetse transmitted sleeping sickness. In: Aksoy, S., Ed., *Transgenesis and the Management of Vector-Borne Disease*, Landes Bioscience, Austin, TX, pp. 36–48.
- Alphey, N., Coleman, P.G., Donnelly, C.A., and Alphey, L., (2007). Managing insecticide resistance by mass release of engineered insects. *J. Econ. Entomol.* **100:** 1642–1649.
- Alphey, N., Bonsall, M.B., and Alphey, L., (2011). Modeling resistance to genetic control of insects. *J. Theor. Biol.* **270:** 42–55.
- Amdam, G.V., Simoes, Z.L., Guidugli, K.R., Norberg, K., and Omholt, S.W., (2003). Disruption of vitellogenin gene function in adult honeybees by intra-abdominal injection of double-stranded RNA. *BMC Biotechnol.* **3:** 1.
- Amenya, D.A., Bonizzoni, M., Isaacs, A.T., Jasinskiene, N., Chen, H., and Marinotti, O., et al. (2010). Comparative fitness assessment of *Anopheles stephensi* transgenic lines receptive to site-specific integration. *Insect Mol. Biol.* **19:** 263–269.
- Araujo, R.N., Santos, A., Pinto, F.S., Gontijo, N.F., Lehane, M.J., and Pereira, M.H., (2006). RNA interference of the salivary gland nitrophorin 2 in the triatomine bug *Rhodnius prolixus* (Hemiptera: Reduviidae) by dsRNA ingestion or injection. *Insect Biochem. Mol. Biol.* **36:** 683–693.
- Ashburner, M., Hoy, M.A., and Peloquin, J., (1998). Transformation of arthropods—research needs and long term prospects. *Insect Mol. Biol.* **7 (3):** 201–213.
- Atkinson, P.W., Hines, E.R., Beaton, S., Matthaei, K.I., Reed, K.C., and Bradley, M.P., (1991). Association of exogenous DNA with cattle and insect spermatozoa in vitro. *Mol. Reprod. Dev.* **23:** 1–5.
- Atkinson, P.W., Pinkerton, A.C., and O'Brochta, D.A., (2001). Genetic transformation systems in insects. *Annu. Rev. Entomol.* **46:** 317–346.
- Atkinson, P.W., Warren, W.D., and O'Brochta, D.B., (1993). The *hobo* transposable element of *Drosophila* can be cross-mobilized in houseflies and excises like the *Ac* element of maize. *Proc. Natl. Acad. Sci. USA* **90:** 9693–9697.
- Auge-Gouillou, C., Bigot, Y., Pollet, N., Hamelin, M.H., Meunier-Rotival, M., and Periquet, G., (1995). Human and other mammalian genomes contain transposons of the *mariner* family. *FEBS Lett.* **368:** 541–546.
- Aultman, K.S., Walker, E.D., Gifford, F., Severson, D.W., Beard, C.B., and Scott, T.W., (2000). Managing risks of arthropod vector research. *Science* **288:** 2321–2322.
- Aung, K.M., Boldbaatar, D., Umemiya-Shirafuji, R., Liao, M., Xuenan, X., and Suzuki, H., et al. (2011). Scavenger receptor mediates systemic RNA interference in ticks. *PLoS One* **6 (12):** e28407.
- Badge, R.M., and Brookfield, J.F.Y., (1997). The role of host factors in the population dynamics of selfish transposable elements. *J. Theor. Biol.* **187:** 171–261.
- Baldarelli, R.M., and Lengyel, J.A., (1990). Transient expression of DNA after ballistic introduction into *Drosophila* embryos. *Nucleic Acids Res.* **18:** 5903–5904.

- Baum, J.A., Bogaert, T., Clinton, W., Heck, G.R., Feldmann, P., and Ilagan, O., et al. (2007). Control of coleopteran insect pests through RNA interference. *Nat. Biotechnol.* **25**: 1322–1326.
- Beard, C.B., Mason, P.W., Aksoy, S., Tesh, R.B., and Richards, F.F., (1992). Transformation of an insect symbiont and expression of a foreign gene in the Chagas' disease vector *Rhodnius prolixus*. *Am. J. Trop. Med. Hyg.* **46**: 195–200.
- Beard, C.B., O'Neill, S.L., Tesh, R.B., Richards, F.F., and Aksoy, F.S., (1993). Modification of arthropod vector competence via symbiotic bacteria. *Parasitol. Today* **9**: 179–183.
- Beard, C.B., Durvasula, R.V., and Richards, F.B., (2000). Bacterial symbiont transformation in Chagas disease vectors. In: Handler, A.M., and James, A.A., Eds., *Insect Transgenesis. Methods and Applications*, CRC Press, Boca Raton, FL, pp. 289–303.
- Beatty, B.J., (2000). Genetic manipulation of vectors: a potential novel approach for control of vector-borne diseases. *Proc. Natl. Acad. Sci. USA* **97**: 10295–10297.
- Beatty, B.J., and Carlson, J.O., (1997). Molecular manipulation of mosquitoes. *Curr. Opin. Infect. Dis.* **10**: 372–376.
- Beckage, N.E., (1998). Parasitoids and polydnaviruses. *Bioscience* **48**: 305–311.
- Beckendorf, S.K., and Hoy, M.A., (1985). Genetic improvement of arthropod natural enemies through selection, hybridization or genetic engineering techniques. In: Hoy, M.A., and Herzog, D.C., Eds., *Biological Control in Agricultural IPM Systems*, Academic Press, Orlando, FL, pp. 167–187.
- Bell, A.C., West, A.G., and Felsenfeld, G., (2001). Insulators and boundaries: versatile regulatory elements in the eukaryotic genome. *Science* **291**: 447–450.
- Belles, X., (2010). Beyond *Drosophila*: RNAi in vivo and functional genomics in insects. *Annu. Rev. Entomol.* **55**: 111–128.
- Benedict, M.Q., and Robinson, A.S., (2003). The first releases of transgenic mosquitoes: an argument for the sterile insect technique. *Trends Parasitol.* **19**: 349–355.
- Benedict, M.Q., D'Abbs, P., Dobson, S., Gottlieb, M., Harrington, L.C., and Higgs, S., et al. (2008). Guidance for contained field trials of vector mosquitoes engineered to contain a gene drive system: Recommendations of a scientific working group. *Vector Borne Zoonotic Dis.* **8**: 127–166.
- Benedict, M.Q., Knols, B.G.J., Bossin, H.C., Howell, P.I., Mialhe, E., Caceres, C., and Robinson, A.S., (2009). Colonisation and mass rearing: learning from others. *Malaria J.* **8** (Suppl. 2): 54.
- Benedict, M., M. Eckerstorfer, G. Franz, H. Gaugitsch, A. Greiter, A. Heissenberger, et al. (2010). Scientific/Technical Report Submitted to EFSA. (European Food Safety Authority). *Defining Environmental Risk Assessment Criteria for Genetically Modified Insects to be Placed on the EU Market*.
- Berghammer, A.J., Klingler, M., and Wimmer, E.A., (1999). A universal marker for transgenic insects. *Nature* **402**: 370–371.
- Beumer, K.J., Trautman, J.K., Bozas, A., Liu, J.L., Rutter, J., Gall, J.G., and Carroll, D., (2008). Efficient gene targeting in *Drosophila* by direct embryo injection with zinc-finger nucleases. *Proc. Natl. Acad. Sci. USA* **105**: 19821–19826.
- Bian, G., Xu, Y., Lu, P., Xie, Y., and Xi, Z., (2010). The endosymbiotic bacterium *Wolbachia* induces resistance to dengue virus in *Aedes aegypti*. *PLoS Pathogens* **6** (4): e1000833.
- Birchler, J.A., Pal Bhadra, M., and Bhadra, U., (2000). Making noise about silence: repression of repeated genes in animals. *Curr. Opin. Genet. Dev* **10**: 211–216.
- Black, W.C., Alphey, L., and James, A.A., (2011). Why RIDL is not SIT. *Trends Parasitol.* **27**: 362–370.
- Blagrove, M.S.C., Arias-Goeta, C., Failloux, A.B., and Sinkins, S.P., (2012). *Wolbachia* strain wMel induces cytoplasmic incompatibility and blocks dengue transmission in *Aedes albopictus*. *Proc. Natl. Acad. Sci. USA* **109**: 255–260.
- Blair, C.D., Adelman, Z.N., and Olson, K.E., (2000). Molecular strategies for interrupting arthropod-borne virus transmission by mosquitoes. *Clin. Microbiol. Rev.* **13**: 651–661.

- Boete, C., (2011). Scientists and public involvement: a consultation on the relation between malaria, vector control and transgenic mosquitoes. *Trans. R. Soc. Trop. Med. Hyg.* **105**: 704–710.
- Bonnivard, E., Bazin, C., Denis, B., and Olson, K.E., (2000). A scenario for the *hobo* transposable element invasion, deduced from the structure of natural populations of *Drosophila melanogaster* using tandem repeats. *Genet. Res. Camb.* **75**: 13–23.
- Borgio, J.F., (2010). RNAi mediated gene knockdown in sucking and chewing insect pests. *J. Biopesticides* **3** (1 Special Issue): 386–393.
- Braig, H.R., and Yan, G., (2002). The spread of genetic constructs in natural insect populations. In: Letourneau, D.K., and Burrows, B.E., Eds., *Genetically Engineered Organisms. Assessing Environmental and Human Health Effects*, CRC Press, Boca Raton, FL, pp. 252–314.
- Britten, R.J., (1997). Mobile elements inserted in the distant past have taken on important functions. *Gene* **205**: 177–182.
- Brookfield, J.F.Y., (1996). Models of the spread of non-autonomous selfish transposable elements when transposition and fitness are coupled. *Genet. Res. Camb.* **67**: 199–209.
- Brookfield, J.F.Y., and Badge, R.M., (1997). Population genetics models of transposable elements. *Genetica* **100**: 281–294.
- Bucher, E., Lohuis, D., and van Poppel, M.M.J.A., et al. (2006). Multiple virus resistance at a high frequency using a single transgene construct. *J. Gen. Virol.* **87**: 3679–3701.
- Bucher, G., Scholten, J., and Klingler, M., (2002). Parental RNAi in *Tribolium* (Coleoptera). *Curr. Biol.* **12**: R85–R86.
- Bucheton, A., (1995). The relationship between the *flamenco* gene and *gypsy* in *Drosophila*: how to tame a retrovirus. *Trends Genet.* **11**: 349–353.
- Burns, J.C., (2000). Pantropic retroviral vectors for insect gene transfer. In: Handler, A.M., and James, A.A., Eds., *Insect Transgenesis. Methods and Applications*, CRC Press, Boca Raton, FL, pp. 125–137.
- Bush, G.L., (1979). Ecological genetics and quality control. In: Hoy, M.A., and McKelvey Jr., J.J., Eds., *Genetics in Relation to Insect Management*, Rockefeller Foundation Press, New York, pp. 145–152.
- Campbell, E.M., Budge, G.E., and Bowman, A.S., (2010). Gene-knockdown in the honeybee mite *Varroa destructor* by a non-invasive approach: studies on a glutathione S-transferase. *Parasit. Vectors* **3**: 73.
- Caprio, M.A., and Hoy, M.A., (1994). Metapopulation dynamics affect resistance development in a predatory mite. *J. Econ. Entomol.* **87**: 525–534.
- Caprio, M.A., and Hoy, M.A., (1995). Premating isolation in a simulation model generates frequency-dependent selection and alters establishment rates of resistant natural enemies. *J. Econ. Entomol.* **88**: 205–212.
- Caprio, M.A., Hoy, M.A., and Tabashnik, B.E., (1991). A model for implementing a genetically improved strain of the parasitoid *Trioxys pallidus* Haliday (Hymenoptera: Aphidiidae). *Am. Entomol.* **34**: 232–239.
- Capy, P., Bazin, C., Anxolabehere, D., and Langin, T., (1996). Horizontal transfers and the evolution of transposable elements. In: Schmidt, E.R., and Hankeln, T., Eds., *Transgenic Organisms and Biosafety*, Springer-Verlag, Berlin, Germany, pp. 15–30.
- Carlson, J., Afanasiev, B., and Suchman, E., (2000). Densonucleosis viruses as transducing vectors for insects. In: Handler, A.M., and James, A.A., Eds., *Insect Transgenesis. Methods and Applications*, CRC Press, Boca Raton, FL, pp. 138–159.
- Carroll, D., (2011). Genome engineering with zinc-finger nucleases. *Genetics* **188**: 773–782.
- Castro, L. (2011). Big issues around a tiny insect: discussing the release of genetically modified mosquitoes (GMM) in Brazil and beyond. M. S. Thesis. 96 pp. Available at: <http://esst.eu/wp-content/uploads/LousiaCastroMasterThesis.pdf>.
- Catteruccia, F., Benton, J.P., and Crisanti, A., (2005). An *Anopheles* transgenic sexing strain for vector control. *Nat. Biotechnol.* **23**: 1414–1417.

- Catteruccia, F., Nolan, T., Loukeris, T.G., Blass, C., Savakis, C., Kafatos, F.C., and Crisanti, A., (2000). Stable germline transformation of the malaria mosquito *Anopheles stephensi*. *Nature* **405**: 959–962.
- Center for Environmental Risk Assessment (2011). *Problem Formulation for the Environmental Risk Assessment of RNAi Plants*. Conference Proc., June 1–3, 2011, Washington, DC.
- Chen, C.H., Huang, H., Ward, C.M., Su, J.T., Schaeffer, L.V., and Guo, M., et al. (2007). A synthetic maternal-effect selfish genetic element drives population replacement in *Drosophila*. *Science* **316**: 597–600.
- Cheng, Q., and Aksoy, S., (1999). Tissue tropism, transmission and expression of foreign genes *in vivo* in midgut symbionts of tsetse flies. *Insect Mol. Biol.* **8**: 125–132.
- Cirimotich, C.M., Dong, Y., Clayton, A.M., Sandiford, S.L., Souza-Neto, J.A., Mulenga, M., and Dimopoulos, G., (2011). Natural microbe-mediated refractoriness to *Plasmodium* infection in *Anopheles gambiae*. *Science* **332**: 855–858.
- Coates, C.J., Turney, C.L., Frommer, M., O'Brochta, D.A., and Atkinson, P.W., (1995). The transposable element *mariner* can excise in nondrosophilid insects. *Mol. Gen. Genet.* **249**: 246–252.
- Collins, F.H., (1994). Prospects for malaria control through the genetic manipulation of its vectors. *Parasitol. Today* **10**: 370–371.
- Collins, F.H., Kamau, L., Ranson, H.A., and Vulule, J.M., (2000). Molecular entomology and prospects for malaria control. *Bull. World Health Org.* **78**: 1412–1423.
- Concha, C., Belikoff, E.J., Carey, B., Li, F., Schliemann, A.J., and Scott, M.J., (2011). Efficient germ-line transformation of the economically important pest species *Lucilia cuprina* and *Lucilia sericata* (Diptera, Calliphoridae). *Insect Biochem. Mol. Biol.* **41**: 70–75.
- Convention on Biological Diversity (2003). The Cartagena Protocol on Biosafety. Available at: <http://bch.cbd.int/protocol/>.
- Cook, P.E., McMeniman, C.J., and O'Neill, S.L., (2008). Modifying insect population age structure to control vector-borne disease. pp. 126–140. In: *Transgenesis and the Management of Vector-borne Disease*. Aksoy, S., Ed., Landes Biosciences, Austin, TX.
- Cornel, A.J., Benedict, M.Q., Rafferty, C.S., Howells, A.J., and Collins, F.H., (1997). Transient expression of the *Drosophila melanogaster cinnabar* gene rescues eye color in the white eye (WE) strain of *Aedes aegypti*. *Insect Biochem. Mol. Biol.* **27**: 993–997.
- Coutinho-Abreu, I.V., Zhu, K.Y., and Ramalho-Ortiago, M., (2010). Transgenesis and paratransgenesis to control insect-borne diseases: current status and future challenges. *Parasitol. Int.* **59**: 1–8.
- Coy, M.R., Sanscrainte, N.D., Chalaire, K.C., Inberg, A., Maayan, I., and Glick, E., et al. (2012). Gene silencing in adult *Aedes aegypti* mosquitoes through oral delivery of double-stranded RNA. *J. Appl. Entomol.* doi: [10.1111/j.1439-0418.2012.01713.x](https://doi.org/10.1111/j.1439-0418.2012.01713.x).
- Crampton, J., Morris, A., Lycett, G., Warren, A., and Eggleston, P., (1990). Transgenic mosquitoes: a future vector control strategy? *Parasitol. Today* **6**: 31–36.
- Cruz, J., Mane-Padros, D., Belles, X., and Martin, D., (2006). Functions of the ecdysone receptor isoform-A in the hemimetabolous insect *Blattella germanica* revealed by systemic RNAi *in vivo*. *Dev. Biol.* **297**: 158–171.
- CSIRO, (2010). *Risk Analysis on the Australian Release of Aedes aegypti (L.) (Diptera: Culicidae) containing Wolbachia*. CSCIRO, Australia.
- Curtis, C.F., (1979). Translocations, hybrid sterility, and the introduction into pest populations of genes favorable to man. In: Hoy, M.A., and McKelvey Jr., J.J., Eds., *Genetics in Relation to Insect Management*, Rockefeller Foundation Press, New York, pp. 19–30.
- Curtis, C.F., (1994). The case for malaria control by genetic manipulation of its vectors. *Parasitol. Today* **10**: 371–374.
- Curtis, C.F., (2000). The case for deemphasizing genomics for malaria control. *Science* **290**: 1508.
- Curtis, C.F., and Townson, H., (1998). Malaria: existing methods of vector control and molecular entomology. *Br. Med. Bull.* **54**: 311–325.

- Cyranoski, D., (2012). Modified mosquitoes set to quash dengue fever. *Nature* doi: 10.1038/nature2023.9752.
- De Barro, P.J., Murphy, B., Jansen, C.C., and Murray, J., (2011). The proposed release of the yellow fever mosquito, *Aedes aegypti* containing a naturally occurring strain of *Wolbachia* *pipientis*, a question of regulatory responsibility. *J. Consumer Prot. Food Safety* 6 (Suppl. 1): 533–540.
- De la Fuente, J., Kocan, K.M., Almazan, C., and Blouin, E.F., (2007). RNA interference for the study and genetic manipulation of ticks. *Trends Parasitol.* 23: 427–433.
- Dengue Vaccine Initiative (2012). Report of the Americas Dengue Prevention Board Meeting. *Points for Consideration for First Introductions of Dengue Vaccines*, Bucaramanga, Colombia, July 16–17, 2012, International Vaccine Institute, World Health Organization, Sabin Vaccine Institute, and IVAC. 17 pp, Available at: <[www.denguevaccines.org/sites/default/files/AmDPB\\_July2012\\_PointsforConsideration.pdf](http://www.denguevaccines.org/sites/default/files/AmDPB_July2012_PointsforConsideration.pdf)>
- De Rubertis, F., Kadosh, D., Henchoz, S., Pauli, D., Reuter, G., Struhl, K., and Spierer, P., (1996). The histone deacetylase RPD3 counteracts genomic silencing in *Drosophila* and yeast. *Nature* 384: 589–591.
- Diaz, H., Ramirez, A.A., Olarte, A., and Clavijo, C., (2011). A model for the control of malaria using genetically modified vectors. *J. Theor. Biol.* 276: 57–66.
- Dobson, S.L., (2003). Reversing *Wolbachia*-based population replacement. *Trends Parasitol.* 19: 128–133.
- Dorer, D.R., and Henikoff, S., (1994). Expansions of transgene repeats cause heterochromatin formation and gene silencing in *Drosophila*. *Cell* 77: 993–1002.
- Dorer, D.R., and Henikoff, S., (1997). Transgene repeat arrays interact with distant heterochromatin and cause silencing in *cis* and *trans*. *Genetics* 147: 1181–1190.
- Droge, M., Puhler, A., and Selbitschka, W., (1998). Horizontal gene transfer as a biosafety issue: a natural phenomenon of public concern. *J. Biotechnol.* 64Z: 79–90.
- Duncker, B.P., Davies, P.L., and Walker, V.K., (1997). Introns boost transgene expression in *Drosophila melanogaster*. *Mol. Gen. Genet.* 254: 291–296.
- Durvasula, R., Gumbi, A., Panackal, A., Kruglov, O., Aksoy, S., and Merrifield, R., et al. (1997). Prevention of insect borne disease: an approach using transgenic symbiotic bacteria. *Proc. Natl. Acad. Sci. USA* 94: 3274–3278.
- Durvasula, R.V., Sundaram, R.K., Kirsch, P., Hurwitz, I., Crawford, C.V., Dotson, E., and Beard, C.B., (2008). Genetic transformation of a Corynebacterial symbiont from the Chagas disease vector *Triatoma infestans*. *Exp. Parasitol.* 119: 94–98.
- Ebinuma, H., Sugita, K., Matsunaga, E., Endo, S., Yamada, K., and Komamine, A., (2001). Systems for the removal of a selection marker and their combination with a positive marker. *Plant Cell. Rep.* 20: 383–392.
- Edeger, T.T., (1999). North-south research partnerships: the ethics of carrying out research in developing countries. *Br. Med. J.* 319: 438–441.
- Eggleston, P., (1991). The control of insect-borne disease through recombinant DNA technology. *Heredity* 66: 161–172.
- Ehler, L. E. (1990). Environmental impact of introduced biological-control agents: implications for agricultural biotechnology. pp. 85–96. In: *Risk Assessment in Agricultural Biotechnology*. Proc. Intern. Conf., Marois, J.J. and Bruyening, G., Eds., Univ. Calif., Div. Agric. Natur. Res. Publ. No. 1928.
- Enserink, M., (2010a). Australia to test 'mosquito vaccine' against human disease. *Science* 330: 1460–1461.
- Enserink, M., (2010b). GM mosquito trial alarms opponents, strains ties in Gates-funded project. *Science* 330: 1030–1031.
- Evans, O., Caragata, E.P., McMeniman, C.J., Woolfit, M., Green, D.C., and Williams, C.R., et al. (2009). Increased locomotor activity and metabolism of *Aedes aegypti* infected with a life-shortening strain of *Wolbachia* *pipientis*. *J. Exp. Biol.* 212: 1436–1441.

- Ewell, J.J., O'Dowd, D.J., Bergelson, J., Daehler, C.C., D'Antonio, C.M., and Gomez, L.D., et al. (1999). Deliberate introductions of species: research needs. Benefits can be reaped, but risks are high. *Bioscience* **49**: 619–630.
- Fadool, J.M., Hartl, D.L., and Dowling, J.E., (1998). Transposition of the *mariner* element from *Drosophila mauritiana* in zebrafish. *Proc. Natl. Acad. Sci. USA* **95**: 5182–5186.
- Fang, J. (2011). 6,000 GM mosquitoes run amok in a Malaysian forest. <[www.smartplanet.com/blog/rethinking-healthcare/6000-gm-mosquitoes-run-amok-in-a-malaysian-forest/2919?tag=content;siu-container](http://www.smartplanet.com/blog/rethinking-healthcare/6000-gm-mosquitoes-run-amok-in-a-malaysian-forest/2919?tag=content;siu-container)>
- FAO/IAEA (2002). *Status and Risk Assessment of the Use of Transgenic Arthropods in Plant Protection*. IAEA-314.D4.02CT01532.
- FAO/IAEA (2006). *Status and Risk Assessment of the Use of Transgenic Arthropods in Plant Protection*. IAEA-TECDOC-1483. Vienna: FAO/IAEA.
- Federal Register (1991). Proposed guidelines for research involving the planned introduction into the environment of organisms with deliberately modified hereditary traits. *Notice. U.S. Federal Register* **56** (22): 4134–4151. Friday, Feb. 1, 1991.
- Federal Register (2006). Availability of Environmental Assessment for a Proposed Field Trial of Genetically Engineered Pink Bollworm. **71(29)**: 7503–7504.
- Federal Register (2006). Availability of an Environmental Assessment and Finding of No Significant Impact for Field Release of Genetically Engineered Pink Bollworm. **71(75)**: 20068–20069.  
Also at: <[www.aphis.usda.gov/brs/aphisdocs/05\\_09801r\\_ea.pdf](http://www.aphis.usda.gov/brs/aphisdocs/05_09801r_ea.pdf)>.
- Federal Register (2006). Availability of an Addendum to an Environmental Assessment for Field Release of Genetically Engineered Pink Bollworm. **71(118)**: 35408–35409.
- Federal Register (2006). Availability of an Addendum to Environmental Assessment and Finding of No Significant Impact for Field Release of Genetically Engineered Pink Bollworm. **71(143)**: 42348–42350.
- Foster, K.R., Vecchia, P., and Repacholi, M.H., (2000). Science and the precautionary principle. *Science* **288**: 979–981.
- Franz, A.W.E., Sanchez-Vargas, I., Piper, J., Smith, M.R., Khoo, C.C.H., and James, A.A., et al. (2009). Stability and loss of a virus resistance phenotype over time in transgenic mosquitoes harbouring an antiviral effector gene. *Insect Mol. Biol.* **18**: 661–672.
- Fraser, M.J., (2000). The TTAA-specific family of transposable elements: identification, functional characterization, and utility for transformation of insects. In: Handler, A.M., and James, A.A., Eds., *Insect Transgenesis. Methods and Applications*, CRC Press, Boca Raton, FL, pp. 249–268.
- Fraser, M.J., (2012). Insect transgenesis: Current applications and future prospects. *Annu. Rev. Entomol.* **57**: 267–289.
- Frentiu, F.D., Robinson, J., Young, P.R., McGraw, E.A., and O'Neill, S.L., (2010). *Wolbachia*-mediated resistance to dengue virus infection and death at the cellular level. *PLoS ONE* **5** (10): e13398.
- Frommer, W.B., and Ninnemann, O., (1995). Heterologous expression of genes in bacterial, fungal, animal, and plant cells. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **46**: 419–444.
- Fu, G., Condon, K.C., Epton, M.J., Gong, P., Jin, L., Condon, G.C., Morrison, N.I., Dafa'alla, T.H., and Alphey, L., (2007). Female-specific insect lethality engineered using alternative splicing. *Nat. Biotechnol.* **25**: 353–357.
- Fu, G., Lees, R.S., Nimmo, D., Aw, D., Jin, L., Gray, P., Berendonk, T.U., White-Cooper, H., Scaife, S., Phuc, H.K., Marinotti, O., Jasinskiene, N., James, A.A., and Alphey, L., (2010). Female-specific flightless phenotype for mosquito control. *Proc. Natl. Acad. Sci. USA* **107**: 4550–4554.
- Garbutt, J.S., Belles, X., Richards, E.H., and Reynolds, S.E., (2012). Persistence of double-stranded RNA in insect hemolymph as a potential determiner of RNA interference success: evidence from *Manduca sexta* and *Blattella germanica*. *J. Insect Physiol.* **59**: 171–178.
- Garcia-Fernandez, J., Bayascas-Ramirez, J.R., Marfany, G., Munoz-Marmol, A.M., Casali, A., Baguna, J., and Slo, E., (1995). High copy number of highly similar *mariner*-like transposons in

- planarian (Platyhelminthes): Evidence for a trans-phyla horizontal transfer. *Mol. Biol. Evol.* **12**: 421–431.
- Gopinathan, K.P., (1992). Biotechnology in sericulture. *Curr. Sci.* **62**: 283–287.
- Goldsmith, M.R., Shimada, T., and Abe, H., (2005). The genetics and genomics of the silkworm. *Bombyx mori. Annu. Rev. Entomol.* **50**: 71–100.
- Golic, M.M., Rong, Y.S., Petersen, R.B., Lindquist, S.L., and Golic, K.G., (1997). FLP-mediated DNA mobilization to specific target sites in *Drosophila* chromosomes. *Nucleic Acids Res.* **25**: 3665–3671.
- Gordon, K.H.J., and Waterhouse, P.M., (2007). RNAi for insect-proof plants. *Nat. Biotechnol.* **25**: 1231–1232.
- Gravitz, L., (2012). The last bite. *Nature* **484**: S26–S27.
- Grefenstette, B., O. El-Lissy and R. T. Staten (2009). Pink bollworm eradication plan in the U.S. Available at: <[www.aphis.usda.gov/plant\\_health/plant\\_pest\\_info/cotton\\_pests/downloads/pbw-erad-plan2-09.pdf](http://www.aphis.usda.gov/plant_health/plant_pest_info/cotton_pests/downloads/pbw-erad-plan2-09.pdf)>
- Grenier, E., Abadon, M., Brunet, F., Capy, P., and Abad, P., (1999). A mariner-like transposable element in the insect parasite nematode *Heterorhabditis bacteriophora*. *J. Mol. Evol.* **48**: 328–336.
- Grenier, A.M., Da Rocha, M., Jalabert, A., Royer, C., Mauchamp, B., and Chavancy, G., (2004). Artificial parthenogenesis and control of voltinism to manage transgenic populations in *Bombyx mori*. *J. Insect Physiol.* **50**: 751–760.
- Gueiros-Filho, F.J., and Beverley, S.M., (1997). Trans-kingdom transposition of the *Drosophila* element *mariner* within the protozoan. *Leishmania*. *Science* **276**: 1716–1719.
- Gwadz, R.W., (1994). Genetic approaches to malaria control: how long the road? *Am. J. Trop. Med. Hyg.* **50** (6 Suppl.): 116–125.
- Halpin, C., Barakate, A., Askari, B.M., Abbott, J.C., and Ryan, M.D., (2001). Enabling technologies for manipulating multiple genes on complex pathways. *Plant Mol. Biol.* **47**: 295–310.
- Hancock, P.A., Sinkins, S.P., and Godfray, H.C.J., (2011). Strategies for introducing *Wolbachia* to reduce transmission of mosquito-borne diseases. *PLOS Negl. Trop. Dis.* **5** (4): e1024.
- Handler, A.M., (2000). An introduction to the history and methodology of insect gene transfer. In: Handler, A.M., and James, A.A., Eds., *Insect Transgenesis. Methods and Applications*, CRC Press, Boca Raton, FL, pp. 3–26.
- Handler, A.M., and Harrell, R.A., (2001). Transformation of the Caribbean fruit fly, *Anastrepha suspensa*, with a piggyBac vector marked with polyubiquitin-regulated FGP. *Insect Biochem. Mol. Biol.* **31**: 199–205.
- Handler, A.M., Zimowska, G.J., and Horn, C., (2007). Improving the ecological safety of transgenic insects for field release: New vectors for stability and genomic targeting. 2007. In: Vreyen, M.J.B., Robinson, A.S., and Hendrichs, J., Eds., *Area-Wide Control of Insect Pests*, Springer, Dordrecht, The Netherlands, pp. 73–83.
- Hanski, I., (1998). Metapopulation dynamics. *Nature* **396**: 41–49.
- Harris, A.F., Nimmo, D., McKemey, A.R., Kelly, N., Scaife, S., Donnelly, C.A., Beech, C., Petrie, W.D., and Alphey, L., (2011). Field performance of engineered male mosquitoes. *Nat. Biotechnol.* **29**: 1034–1037.
- Hay, B.A., Chen, C.H., Ward, C.M., Huang, H., Su, J.T., and Guo, M., (2010). Engineering the genomes of wild insect populations: challenges and opportunities provided by synthetic *Medea* selfish genetic elements. *J. Insect Physiol.* **56**: 1402–1413.
- He, P., Zhang, J., Liu, N.Y., Zhang, Y.N., Yang, K., and Dong, S.L., (2011). Distinct expression profiles and different functions of odorant binding proteins in *Nilaparvata lugens* Stal. *PLoS One* **6** (12): e28921.
- Headley, J.C., and Hoy, M.A., (1987). Benefit/cost analysis of an integrated mite management program for almonds. *J. Econ. Entomol.* **80**: 555–559.

- Heath, B.D., Butcher, R.D.J., Whitfield, W.G.F., and Hubbard, S.F., (1999). Horizontal transfer of *Wolbachia* between phylogenetically distant insect species by a naturally occurring mechanism. *Curr. Biol.* **9**: 313–316.
- Heilmann, L.J., DeVault, J.D., Leopold, R.L., and Narang, S.K., (1994). Improvement of natural enemies for biological control: a genetic engineering approach. In: Narang, S.K., Bartlett, A.C., and Faust, R.M., Eds., *Applications of Genetics to Arthropods of Biological Control Significance*, CRC Press, Boca Raton, FL, pp. 167–189.
- Henikoff, S., (1998). Conspiracy of silence among repeated transgenes. *Bioessays* **20**: 532–535.
- Higgs, S., and Lewis, D.L., (2000). Green fluorescent protein (GFP) as a marker for transgenic insects. In: Handler, A.M., and James, A.A., Eds., *Insect Transgenesis. Methods and Applications*, CRC Press, Boca Raton, FL, pp. 93–108.
- Hoffmann, A.A., Montgomery, B.L., Popovici, J., Iturbe-Ormaetxe, I., and Johnson, P.H., et al. (2011). Successful establishment of *Wolbachia* in *Aedes* populations to suppress dengue transmission. *Nature* **476**: 454–457.
- Hofmann, C., Sandig, V., Jennings, G., Rudolph, M., Schlag, P., and Strauss, M., (1995). Efficient gene transfer into human hepatocytes by baculovirus vectors. *Proc. Natl. Acad. Sci. USA* **92**: 10099–10103.
- Hohlweg, U., and Doerfler, W., (2001). On the fate of plant or other foreign genes upon the uptake in food or after intramuscular injection in mice. *Mol. Genet. Genomics* **265**: 225–233.
- Horard, B., Mange, A., Pelissier, B., and Couble, P., (1994). *Bombyx* gene promoter analysis in transplanted silk gland transformed by particle delivery system. *Insect Mol. Biol.* **3**: 261–265.
- Horn, C., Jaunich, B., and Wimmer, E.A., (2000). Highly sensitive, fluorescent transformation marker for *Drosophila* transgenesis. *Dev. Genes Evol.* **210**: 623–629.
- Horn, C., and Wimmer, E.A., (2003). A transgene-based, embryo-specific lethality system for insect pest management. *Nat. Biotechnol.* **21**: 64–70.
- Houck, M.A., Clark, J.B., Peterson, K.R., and Kidwell, M.G., (1991). Possible horizontal transfer of *Drosophila* genes by the mite *Proctolaelaps regalis*. *Science* **253**: 1125–1129.
- Hoy, M.A., (1976). Genetic improvement of insects: fact or fantasy. *Environ. Entomol.* **5**: 833–839.
- Hoy, M.A., (1985). Almonds: Integrated mite management for California almond orchards In: Helle, W., and Sabelis, M.W., Eds., *Spider Mites, Their Biology, Natural Enemies, and Control*, Vol. **1B**, Elsevier, Amsterdam, The Netherlands, pp. 299–310.
- Hoy, M.A., (1990a). Genetic improvement of arthropod natural enemies: Becoming a conventional tactic? In: Baker, R., and Dunn, P., Eds., *New Directions in Biological Control*, Vol. **112** Alan R. Liss, New York, pp. 405–417. UCLA Symp. Mol. Cell. Biol., New Series.
- Hoy, M.A., (1990b). Pesticide resistance in arthropod natural enemies: Variability and selection responses. In: Roush, R.T., and Tabashnik, B.E., Eds., *Pesticide Resistance in Arthropods*, Chapman & Hall, New York, pp. 203–236.
- Hoy, M.A., (1992a). Commentary: Biological control of arthropods: Genetic engineering and environmental risks. *Biol. Control* **2**: 166–170.
- Hoy, M.A., (1992b). Criteria for release of genetically improved phytoseiids: An examination of the risks associated with release of biological control agents. *Exp. Appl. Acarol.* **14**: 393–416.
- Hoy, M.A., (1993). Transgenic beneficial arthropods for pest management programs: An assessment of their practicality and risks. In: Lumsden, R.D., and Vaughn, J.L., Eds., *Pest Management: Biologically Based Technologies*, American Chemical Society, Washington, DC, pp. 357–369. Amer. Chem. Soc. Conf. Proc. Series.
- Hoy, M.A., (1995). Impact of risk analyses on pest-management programs employing transgenic arthropods. *Parasitol. Today* **11**: 229–232.
- Hoy, M.A., (2000a). Deploying transgenic arthropods in pest management programs: Risks and realities. In: Handler, A.M., James, A.A. (Eds.), *Insect Transgenesis: Methods and Applications* CRC Press, Boca Raton, FL, pp. 335–367.

- Hoy, M.A., (2000b). Transgenic arthropods for pest management programs. risks and realities. *Exp. Appl. Acarol.* **24**: 463–495.
- Hoy, M.A., Gaskalla, R.D., Capinera, J.L., and Keierleber, C.N., (1997). Laboratory containment of transgenic arthropods. *Am. Entomol.* (Winter issue): 206–209. 256.
- Hrycaj, S., Mihajlovic, M., Mahfooz, N., Couso, J.P., and Popadic, A., (2008). RNAi analysis of nubbin embryonic functions in a hemimetabolous insect. *Oncopeltus fasciatus*. *Evol. Dev.* **10**: 705–716.
- Huang, Y., Magori, K., Lloyd, A.L., and Gould, F., (2007). Introducing transgenes into insect populations using combined gene-drive strategies: modeling and analysis. *Insect Biochem. Mol. Biol.* **37**: 1054–1063.
- Hughes, G.L., Koga, R., Xue, P., Fukatsu, T., and Rasgon, J.L., (2011). *Wolbachia* infections are virulent and inhibit the human malaria parasite *Plasmodium falciparum* in *Anopheles gambiae*. *PLOS Pathogens* **7** (5): e1002043.
- Hunt, A.G., and Maiti, I.B., (2001). Strategies for expressing multiple foreign genes in plants as polycistronic constructs. *In Vitro Cell. Dev. Biol. Plant* **37**: 313–320.
- Hunter, W., Ellis, J., vanEngelsdorp, D., Hayes, J., Westervelt, D., and Glick, E., et al. (2010). Large-scale field application of RNAi technology reducing Israeli acute paralysis virus disease in honey bees (*Apis mellifera*, Hymenoptera: Apidae). *PLoS Pathol.* **6** (2): e1001160.
- Hurwitz, I., Fieck, A., Read, A., Hillesland, H., Klein, N., Kang, A., and Durvasula, R., (2011a). Paratransgenic control of vector borne diseases. *Int. J. Biol. Sci.* **7**: 1334–1344.
- Hurwitz, I., Hillesland, H., Fieck, A., Das, P., and Durvasula, R., (2011b). The paratransgenic sand fly: a platform for control of *Leishmania* transmission. *Parasit. Vectors* **4**: 82.
- Hussain, M., Frentiu, F.D., Moreira, L.A., O'Neill, S.L., and Asgari, S., (2011). *Wolbachia* uses host microRNAs to manipulate host gene expression and facilitate colonization of the dengue vector *Aedes aegypti*. *Proc. Natl. Acad. Sci. USA* **108**: 9250–9255.
- Huvenne, H., and Smagghe, G., (2010). Mechanisms of dsRNA uptake in insects and potential of RNAi for pest control: a review. *J. Insect Physiol.* **56**: 227–235.
- Iatrou, K., and Meidinger, R.G., (1990). Tissue-specific expression of silkworm chorion genes *in vivo* using *Bombyx mori* nuclear polyhedrosis virus as a transducing vector. *Proc. Natl. Acad. Sci. USA* **87**: 3650–3654.
- James, A.A., (2000). Control of disease transmission through genetic modification of mosquitoes. In: Handler, and A.M., James, A.A., Eds., *Insect Transgenesis. Methods and Applications*, CRC Press, Boca Raton, FL, pp. 319–333.
- James, A.A., (2001). Present and future control of malaria. *Science* **291**: 435.
- James, A.A., (2004). Gene drive systems in mosquitoes: rules of the road. *Trends Parasitol.* **21**: 64–67.
- James, A.A., (2005). Gene drive systems in mosquitoes: rules of the road. *Trends Parasit.* **21**: 64–67.
- James, S., Simmons, C.P., and James, A.A., (2011). Mosquito trials. *Science* **334**: 771–772.
- Jarosch, A., and Moritz, R.F.A., (2011). Systemic RNA-interference in the honeybee *Apis mellifera*: tissue dependent uptake of fluorescent siRNA after intra-abdominal application observed by laser-scanning microscopy. *J. Insect Physiol.* **57**: 851–857.
- Jaubert-Posamai, S., Le Trionnaire, G., Bonhomme, J., Christophides, G.K., and Rispe, C., et al. (2007). Gene knockdown by RNAi in the pea aphid *Acyrtosiphon pisum*. *BMC Biotechnol.* **7**: 8.
- Jeffery, J.A., Yen, N.T., Nam, V.S., Nghia, T., Hoffmann, A.A., Kay, B.H., and Ryan, P.A., (2009). Characterizing the *Aedes aegypti* population in a Vietnamese village in preparation for a *Wolbachia*-based mosquito control strategy to eliminate dengue. *PLoS Negl. Trop. Dis.* **3** (11): e552.
- Jehle, J.A., Nickel, A., Vlak, J.M., and Backhaus, H., (1998). Horizontal escape of the novel Tc1-like Lepidopteran transposon TCp3.2 into *Cydia pomonella* granulovirus. *J. Mol. Evol.* **46**: 215–224.

- Jensen, S., Gassama, M.P., and Heidmann, T., (1999). Taming of transposable elements by homology-dependent gene silencing. *Nat. Genet.* **21**: 209–213.
- Jeyaprakash, A., Lopez, G., and Hoy, M.A., (1998). Extrachromosomal plasmid DNA transmission and amplification in *Metaseiulus occidentalis* (Acari: Phytoseiidae) transformants generated by maternal microinjection. *Ann. Entomol. Soc. Am.* **91**: 730–736.
- Jin, C., Ren, X., and Rasgon, J.L., (2009). The virulent *Wolbachia* strain wMelPop efficiently establishes somatic infections in the malaria vector *Anopheles gambiae*. *Appl. Environ. Microbiol.* **75**: 3373–3376.
- Jones, I., and Morikawa, Y., (1996). Baculovirus vectors for expression in insect cells. *Curr. Opin. Biotechnol.* **7**: 512–516.
- Jordan, T.V., Shike, H., Boulo, V., Cedeno, V., Fang, Q., and Davis, B.S., et al. (1998). Pantropic retroviral vectors mediate somatic cell transformation and expression of foreign genes in dipteran insects. *Insect Mol. Biol.* **7**: 215–222.
- Jordan, I.K., Matyunina, L.V., and McDonald, J.F., (1999). Evidence for recent horizontal transfer of long terminal repeat retrotransposon. *Proc. Natl. Acad. Sci. USA* **96**: 12621–12625.
- Kambris, Z., Cook, P.E., Phuc, H.K., and Sinkins, S.P., (2009). Immune activation by life-shortening *Wolbachia* and reduced filarial competence in mosquitoes. *Science* **326**: 134–136.
- Kamdar, P., von Allmen, G., and Finnerty, V., (1992). Transient expression of DNA in *Drosophila* via electroporation. *Nucleic Acids Res.* **20**: 3526.
- Kapuscinski, A.R., (2002). Controversies in designing useful ecological assessments of genetically engineered organisms. In: Letourneau, D.K., and Burrows, B.E., Eds., *Genetically Engineered Organisms. Assessing Environmental and Human Health Effects*, CRC Press, Boca Raton, FL, pp. 385–415.
- Karim, S., Troiano, E., and Mather, T.N., (2010). Functional genomics tool: gene silencing in *Ixodes scapularis* eggs and nymphs by electroporated dsRNA. *BMC Biotechnol.* **10**: 1.
- Kawai, S., Matsumoto, Y., Gotoh, T., and Noda, H., (2009). Transinfection of *Wolbachia* in planthoppers: nymphal injection of cultured *Wolbachia* and infection dynamics. *Environ. Entomol.* **38**: 1626–1633.
- Kazuki, Y., Hoshiya, H., Takiguchi, M., Abe, S., Ilda, Y., and Osaki, M., et al. (2011). Refined human artificial chromosome vectors for gene therapy and animal transgenesis. *Gene Ther.* **18**: 384–393.
- Ke, Z., Benedict, M.Q., Cornel, A.J., Besansky, N.J., and Collins, F.H., (1997). The *Anopheles albimanus* white gene: molecular characterization of the gene and a spontaneous white gene mutation. *Genetica* **101**: 87–96.
- Kennerdell, J.R., and Carthew, R.W., (2000). Heritable gene silencing in *Drosophila* using double-stranded RNA. *Nat. Biotechnol.* **18**: 896–898.
- Khila, A., and Grbic, M., (2007). Gene silencing in the spider mite *Tetranychus urticae*: dsRNA and siRNA parental silencing of the *Distal-less* gene. *Dev. Genes Evol.* **217**: 241–251.
- Khoo, C.C.H., Piper, J., Sanchez-Vargas, I., Olson, K.E., and Franz, A.W.E., (2010). The RNA interference pathway affects midgut infection-and escape barriers for Sindbis virus in *Aedes aegypti*. *BMC Microbiol.* **10**: 130.
- Kidwell, M.G., and Ribeiro, J.M.C., (1992). Can transposable elements be used to drive disease refractoriness genes into vector populations. *Parasitol. Today* **8**: 325–329.
- Kidwell, M.G., and Evgen'ev, M.B., (1999). How valuable are model organisms for transposable element studies? *Genetica* **107**: 103–111.
- Kidwell, M.G., and Lisch, D.R., (2001). Transposable elements, parasitic DNA, and genome evolution. *Evolution* **55**: 1–24.
- Kim, A., Terzian, C., Santamaría, P., Pelisson, A., Prud'homme, N., and Bucheton, A., (1994). Retroviruses in invertebrates: the *gypsy* retrotransposon is apparently an infectious retrovirus of *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* **91**: 1285–1289.

- Kiszewski, A.E., and Spielman, A., (1998). Spatially explicit model of transposon-based genetic drive mechanisms for displacing fluctuating populations of anopheline vector mosquitoes. *J. Med. Entomol.* **35**: 584–590.
- Klinakis, A.G., Zagoraiou, L., Vassilatis, D.K., and Savakis, C., (2000). Genome-wide insertional mutagenesis in human cells by the *Drosophila* mobile element. *Minos. EMBO Rep.* **1**: 416–421.
- Knipling, E., (1955). Possibilities of insect control or eradication through use of sexually sterile males. *J. Econ. Entomol.* **48**: 459–462.
- Knols, B.G.J., Hood-Nowotny, R.C., Bossin, H., Franz, G., Robinson, A., Mukabana, W.R., and Kemboi, S.K., (2006). GM sterile mosquitoes—a cautionary note. *Nat. Biotechnol.* **24**: 1067–1068.
- Knols, B.G.J., Bossin, H.C., Mukabana, W.R., and Robinson, A.S., (2007). Transgenic mosquitoes and the fight against malaria: managing technology push in a turbulent GMO world. *Am. J. Trop. Hyg.* **77** (Suppl. 6): 232–242.
- Koehncke, A., Telschow, A., Werren, J.H., and Hammerstein, P., (2009). Life and death of an influential passenger: *Wolbachia* and the evolution of CI-modifiers by their hosts. *PLoS ONE* **4** (2): e4425.
- Kokoza, V.A., and Raikhel, A.S., (2011). Targeted gene expression in the transgenic *Aedes aegypti* using the binary Gal4-UAS system. *Insect Biochem. Mol. Biol.* **41**: 637–644.
- Kokoza, V., Ahmed, A., Shin, S.W., Okafor, N., Zou, Z., and Raikhel, A.S., (2010). Blocking of *Plasmodium* transmission by cooperative action of Cecropin A and Defensin A in transgenic *Aedes aegypti* mosquitoes. *Proc. Natl. Acad. Sci. USA* **107**: 8111–8116.
- Komoto, N., Thibert, C., Boulo, V., Fang, Q., Burns, J.C., and Couble, P., (2000). Gene introduction into silkworm embryos with a pseudotyped retroviral vector. *J. Seric. Sci. Jpn.* **69**: 55–61.
- Kongmee, M., Nimmo, D., Labbe, G., Beech, C., Grieco, J., and Alphey, L., et al. (2010). Irritant and repellent behavioral responses of *Aedes aegypti* male populations developed for RIDL disease control strategies. *J. Med. Entomol.* **47**: 1092–1098.
- Konsolaki, M., Sanicola, M., Kozlova, T., Liu, V., Arca, B., and Savakis, C., et al. (1992). FLP-mediated intermolecular recombination in the cytoplasm of *Drosophila* embryos. *New Biol.* **4**: 551–557.
- Krishnapillai, V., (1996). Horizontal gene transfer. *J. Genet.* **75**: 219–232.
- LaChance, L.E., (1979). Genetic strategies affecting the success and economy of the sterile insect release method. In: Hoy, M.A., and McKelvey Jr., J.J., Eds., *Genetics in Relation to Insect Management*, Rockefeller Foundation Press, New York, pp. 8–18.
- Lacroix, R., McKemey, A.R., Raduan, N., Wee, L.K., Ming, W.H., and Ney, T.G., et al. (2012). Open field release of genetically engineered sterile male *Aedes aegypti* in Malaysia. *PLoS ONE* **7** (8): e42771.
- Lampe, D.J., Akerley, B.J., Rubin, E.J., Mekalanos, J.J., and Robertson, H.M., (1999). Hyperactive transposase mutants of the *Himar1 mariner* transposon. *Proc. Natl. Acad. Sci. USA* **96**: 11428–11433.
- Lampe, D.J., Walden, K.K.O., Sherwood, J.M., and Robertson, H.M., (2000). Genetic engineering of insects with *mariner* transposons. In: Handler, A.M., and James, A.A., Eds., *Insect Transgenesis. Methods and Applications*, CRC Press, Boca Raton, FL, pp. 237–248.
- Lanzaro, G.C., Toure, Y.T., Carnahan, J., Zheng, L., Dolo, G., and Traore, S., et al. (1998). Complexities in the genetic structure of *Anopheles gambiae* populations in west Africa as revealed by microsatellite DNA analysis. *Proc. Natl. Acad. Sci. USA* **95**: 14260–14265.
- Laven, H., (1951). Crossing experiments with *Culex* strains. *Evolution* **5**: 370–375.
- Lavery, J.V., Harrington, L.C., and Scott, T.W., (2008). Ethical, social, and cultural considerations for site selection for research with genetically modified mosquitoes. *Am. J. Trop. Med. Hyg.* **79**: 312–318.
- Lee, D.W., Shrestha, S., Kim, A.Y., Park, S.J., Yang, C.Y., and Kim, Y., et al. (2011). RNA interference of pheromone biosynthesis-activating neuropeptide receptor suppresses mating behavior by

- inhibiting sex pheromone production in *Plutella xylostella* (L.). *Insect Biochem. Mol. Biol.* **41**: 236–243.
- Lehane, M.J., and Aksoy, S., (2012). Control using genetically modified insects poses problems for regulators. *PLoS Negl. Trop. Dis.* **6** (1): e1495.
- Leopold, R.A., Hughes, K.J., and DeVault, J.D., (1996). Using electroporation and a slot cuvette to deliver plasmid DNA to insect embryos. *Gen. Anal. Biomol. Eng.* **12**: 197–200.
- Leroy, H., Leroy, F., Auge-Gouillou, C., Castagnone-Sereno, P., Vanlerberghe-Masutti, F., Bigot, Y., and Abad, P., (2000). Identification of *mariner*-like elements from the root-knot nematode *Meloidogyne* spp. *Mol. Biochem. Parasitol.* **107**: 181–190.
- Le Rouzic, A., and Capy, P., (2006). Reversible introduction of transgenes in natural populations of insects. *Insect Mol. Biol.* **15**: 227–234.
- Li, J., and Hoy, M.A., (1996). Adaptability and efficacy of transgenic and wild-type *Metaseiulus occidentalis* (Acari: Phytoseiidae) compared as part of a risk assessment. *Exp. Appl. Acarol.* **20**: 563–574.
- Li, X., Zhang, M., and Zhang, H., (2011). RNA interference of four genes in adult *Bactrocera dorsalis* by feeding their dsRNAs. *PLoS ONE* **6** (3): 317788.
- Liu, X., Zhang, Y., Yan, X., and Han, R., (2010). Prevention of Chinese sacbrood virus infection in *Apis cerana* using RNA interference. *Curr. Microbiol.* **61**: 422–428.
- Lopez-Martinez, G., Meuti, M., and Denlinger, D.L., (2012). Rehydration driven RNAi: a novel approach for effectively delivering dsRNA to mosquito larvae. *J. Med. Entomol.* **49**: 2115–2218.
- Loukeris, T.G., Livadaras, I., Arca, B., Zabalou, S., and Savakis, C., (1995). Gene transfer into the Medfly, *Ceratitis capitata*, with a *Drosophila hydei* transposable element. *Science* **270**: 2002–2005.
- Lu, P., Bian, G., Pan, X., and Xi, Z., (2012). *Wolbachia* induces density-dependent inhibition to dengue virus in mosquito cells. *PLoS Negl. Trop. Dis.* **6** (7): e1754.
- Lundin, K.E., Simonson, O.E., Moreno, P.M.D., Zaghloul, E.M., Oprea, I.I., and Svahn, M.G., et al. (2009). Nanotechnology approaches for gene transfer. *Genetica* **137**: 47–56.
- Lynch, J.A., and Desplan, C., (2006). A method for parental RNA interference in the wasp *Nasonia vitripennis*. *Nat. Protocols* **1**: 486–494.
- Macer, D. (2003). Ethical, legal and social issues of genetically modified disease vectors in public health. Special Topics No. 1. TDR/STR/SEB/ST/03.1. WHO Special Programme for Research and Training in Tropical Diseases, World Health Organization, Geneva, Switzerland.
- Mao, Y.B., Cai, W.J., Wang, J.W., Hong, G.J., Tao, X.Y., and Wang, L.J., et al. (2007). Silencing a cotton bollworm P450 monooxygenase gene by plant-mediated RNAi impairs larval tolerance of gossypol. *Nat. Biotechnol.* **25**: 1307–1313.
- Maori, E., Paidi, N., Shafir, S., Kalev, H., Tsur, E., Glick, E., and Sela, I., (2009). IAPV, a bee-affecting virus associated with Colony Collapse Disorder can be silenced by dsRNA ingestion. *Insect Mol. Biol.* **18**: 55–60.
- Marec, F., Neven, L.G., Robinson, A.S., Vreysen, M., Goldsmith, M.R., and Nagaraju, J., et al. (2005). Development of genetic sexing strains in Lepidoptera: from traditional to transgenic approaches. *J. Econ. Entomol.* **98**: 248–259.
- Marrelli, M.T., Moreira, C.K., Kelly, D., Alphey, L., and Jacobs-Lorena, M., (2006). Mosquito transgenesis: what is the fitness cost? *Trends Parasitol.* **22**: 197–202.
- Marrelli, M.T., Li, C., Rasgon, J.L., and Jacobs-Lorena, M., (2007). Transgenic malaria-resistant mosquitoes have a fitness advantage when feeding on *Plasmodium*-infected blood. *Proc. Natl. Acad. Sci. USA* **104**: 5580–5583.
- Marshall, E., (2000). A renewed assault on an old and deadly foe. *Science* **290**: 428–430.
- Marshall, J.M., (2008). The impact of dissociation on transposon-mediated disease control strategies. *Genetics* **178**: 1673–1682.
- Marshall, J.M., (2009). The effect of gene drive on containment of transgenic mosquitoes. *J. Theor. Biol.* **258**: 250–265.

- Marshall, J.M., (2010). The Cartagena protocol and genetically modified mosquitoes. *Nat. Biotechnol.* **28**: 896–897.
- Marshall, J.M., (2011). Commentary: The Cartagena Protocol in the context of recent releases of transgenic and *Wolbachia*-infected mosquitoes. *Asia Pac. J. Mol. Biol. Biotechnol.* **19**: 93–100.
- Marshall, J.M., and Taylor, C.E., (2009). Malaria control with transgenic mosquitoes. *PLOS Med.* **6** (2): e100020.
- Marshall, J.M., Toure, M.B., Traore, M.M., Famenini, S., and Taylor, C.E., (2010). Perspectives of people in Mali toward genetically-modified mosquitoes for malaria control. *Malar. J.* **9**: 128.
- Martin, D.O., Maestro, J., Cruz, D., Mane-Padros, and Belles, X., (2006). RNAi studies reveal a conserved role for RXR in molting in the cockroach *Blattella germanica*. *J. Insect Physiol.* **52**: 410–416.
- Matsubara, T., Beeman, R.W., Shike, H., Besansky, N.J., Mukabayire, O., and Higgs, S., et al. (1996). Pantropic retroviral vectors integrate and express in cells of the malaria mosquito, *Anopheles gambiae*. *Proc. Natl. Acad. Sci. USA* **93**: 6181–6185.
- Matthews, P.R., Wang, M.B., Waterhouse, P.M., Thornton, S., Fieg, S.J., and Gubler, F., et al. (2001). Marker gene elimination from transgenic barley, using co-transformation with adjacent 'twin-T-DNAs' on a standard *Agrobacterium* transformation vector. *Mol. Breed.* **7**: 195–202.
- Matthews, S., Rao, S.V., and Durvasula, R.V., (2011). Modeling horizontal gene transfer (HGT) in the gut of the Chagas disease vector. *Rhodnius prolixus*. *Parasit. Vectors* **4**: e77.
- McDermott, G.J., and Hoy, M.A., (1997). Persistence and containment of *Metaseiulus occidentalis* (Acar: Phytoseiidae) in Florida: Risk assessment for possible releases of transgenic strains. *Fla. Entomol.* **80**: 42–53.
- McMeniman, C.J., Lane, R.V., Cass, B.N., Fong, A.W.C., Sidhu, M.A., and Wang, Y.F., et al. (2009). Stable introduction of a life-shortening *Wolbachia* infection into the mosquito *Aedes aegypti*. *Science* **323**: 141–144.
- Merrihew, R.V., Clay, W.C., Condreay, J.P., Witherspoon, S.M., Dallas, W.S., and Kost, T.A., (2001). Chromosomal integration of transduced recombinant baculovirus DNA in mammalian cells. *J. Virol.* **75**: 903–909.
- Meyering-Vos, M., and Muller, A., (2007). RNA interference suggests sulfakinins as satiety effectors in the cricket. *Gryllus bimaculatus*. *J. Insect Physiol.* **53**: 840–848.
- Meza, J.S., Nirmala, X., Zimowska, G.J., Zepeda-Cisneros, C.S., and Handler, A.M., (2011). Development of transgenic strains for the biological control of the Mexican fruit fly, *Anastrepha ludens*. *Genetica* **139**: 53–62.
- Mialhe, E., and Miller, L.H., (1994). Biolistic techniques for transfection of mosquito embryos (*Anopheles gambiae*). *Biotechniques* **16**: 924.
- Miller, L.K., (1988). Baculoviruses as gene expression vectors. *Annu. Rev. Microbiol.* **42**: 177–199.
- Miller, L.H., (1989). Strategies for malaria control: realities, magic, and science. In *Biomedical Science and the Third World: Under the Volcano*. *Ann. N. Y. Acad. Sci* **569**: 118–126.
- Miller, E., State, R.T., Claus, J., Sledge, M., Peloquin, J., and Miller, T., (2001). A multiple generation life history study on rearing a genetically altered (EGFP) strain of pink bollworm (Lepidoptera: Gelechiidae). *Proc. Belt Wide Cotton Conf*, January 9–13, Anaheim, CA.
- Miller, L.H., Sakai, R.K., Romans, P., Gwadz, W., Kantoff, P., and Coon, H.G., (1987). Stable integration and expression of a bacterial gene in the mosquito *Anopheles gambiae*. *Science* **237**: 779–781.
- Miller, T., Durvasula, R., Matthews, S., Lauzon, C., and Lampe, D., (2006). Paratransgenesis applied to control insect-transmitted plant pathogens In: Bourtzis, K., and Miller, T.A., Eds., *Insect Symbiosis*, Vol. 2, CRC Press, Boca Raton, FL, pp. 247–263.
- Miller, J.C., Holmes, M.C., Wang, J., Guschin, D.Y., Lee, Y.L., and Rupniewski, I., et al. (2007). An improved zinc-finger nuclease architecture for highly specific genome editing. *Nat. Biotechnol.* **25**: 778–785.

- Min, K.T., and Benzer, S., (1997). *Wolbachia*, normally a symbiont of *Drosophila*, can be virulent, causing degeneration and early death. *Proc. Natl. Acad. Sci. USA* **94**: 10792–10796.
- Miyawaki, K., Mito, T., Sarashina, I., Zhang, H., Shinmyo, Y., Ohuchi, H., and Noji, S., (2004). Involvement of Wingless/Armadillo signaling in the posterior sequential segmentation in the cricket, *Gryllus bimaculatus* (Orthoptera), as revealed by RNAi analysis. *Mech. Dev.* **121**: 119–130.
- Moreira, L.A., Iturbe-Ormaetse, I., Jeffery, J.A., Lu, G.J., and Pyke, A.T., et al. (2009a). A *Wolbachia* symbiont in *Aedes aegypti* limits infection with dengue, Chikungunya, and *Plasmodium*. *Cell* **139**: 1268–1278.
- Moreira, L.A., Saig, E., Turley, A.P., Ribeiro, J.M.C., O'Neill, S.L., and McGraw, E.A., (2009b). Human probing behavior of *Aedes aegypti* when infected with a life-shortening strain of *Wolbachia*. *PLOS Negl. Trop. Dis.* **3** (12): 3568.
- Mumford, J.D., (2012). Science, regulation, and precedent for genetically modified insects. *PLoS Negl. Trop. Dis.* **6** (1): e1504.
- Nagaraju, J., Kanda, T., Yukihiko, K., Chavancy, G., Tamura, T., and Couble, P., (1996). Attempt at transgenesis of the silkworm (*Bombyx mori* L.) by egg-injection of foreign DNA. *Appl. Entomol. Zool.* **31**: 587–596.
- NAPPO (2007). *RSPM No. 27. Importation and Release into the Environment of Transgenic Arthropods in NAPPO Member Countries*. North American Plant Protection Organization, Ottawa, Canada.
- National Academy of Sciences (2004). Bioconfinement of Animals: fish, shellfish, and insects, pp. 130–158. In: *Biological Confinement of Genetically Engineered Organisms*. National Academies Press, Washington, DC.
- Nikolaev, A.I., Tchkonia, T.T., Kafianieristavi, C.A., and Tarantul, V.X., (1993). Preferential extrachromosomal localization of exogenous DNA in transgenic silkworm *Bombyx mori* L. *Mol. Gen. Genet.* **236**: 326–330.
- Nikitin, A.G., and Woodruff, R.C., (1995). Somatic movement of the *mariner* transposable element and lifespan of *Drosophila* species. *Mutat. Res.* **338**: 43–49.
- Nirmala, X., Olson, S.R., Holler, T.C., Cho, K.H., and Handler, A.M., (2011). A DsRed fluorescent protein marker under polyubiquitin promoter regulation allows visual and amplified gene detection of transgenic Caribbean fruit flies in field traps. *Biocontrol* **56**: 333–340.
- Nolan, T., Petris, E., Muller, H.M., Cronin, A., Catteruccia, F., and Crisanti, A., (2011). Analysis of two novel midgut-specific promoters driving transgene expression in *Anopheles stephensi* mosquitoes. *PLOS One* **6** (2): e16471.
- O'Brochta, D.A., and Atkinson, P.W., (1997). Recent developments in transgenic insect technology. *Parasitol. Today* **13**: 99–104.
- O'Brochta, D.A., and Handler, A.M., (2008). Perspectives on the state of insect transgenics. In: Aksoy, S., Ed., *Transgenesis and the Management of Vector-Borne Disease*, Landes Bioscience, Austin, TX, pp. 1–18.
- Ogwok, E., Odipio, J., Halsey, M., Gaitan-Solis, E., Bua, A., and Taylor, N.J., et al. (2012). Transgenic RNA interference (RNAi)-derived field resistance to cassava brown streak disease. *Mol. Plant Pathol.* **13** (9): 1019–1031.
- Olson, K.E., (2000). Sindbis virus expression systems in mosquitoes: background, methods, and applications. In: Handler, A.M., and James, A.A., Eds., *Insect Transgenesis. Methods and Applications*, CRC Press, Boca Raton, FL, pp. 161–189.
- Olson, K.E., Higgs, S., Gaines, P.J., Powers, A.M., Davis, B.S., and Kamrud, K.I., et al. (1996). Genetically engineered resistance to Dengue-2 virus transmission in mosquitoes. *Science* **272**: 884–886.
- Olson, K.E., Myles, K.M., Seabaugh, R.C., Higgs, S., Carlson, J.O., and Beaty, B.J., (2000). Development of a Sindbis virus expression system that efficiently expresses green fluorescent protein in midguts of *Aedes aegypti* following per os infection. *Insect Mol. Biol.* **9**: 57–65.

- Omholt, S.W., Rishovd, S., Hagen, A., Elmholdt, O., Dalsgard, B., and Fromm, S., (1995). Successful production of chimeric honeybee larvae. *J. Exp. Zool.* **272**: 410–412.
- Oosumi, T., Belknap, W.R., and Garlick, B., (1995). *Mariner* transposons in humans. *Nature* **378**: 672.
- Pal, R., (1974). WHO/ICMR programme of genetic control of mosquitos in India. In: Pal, R., and Whitten, M.J., Eds., *The Use of Genetics In Insect Control*, Elsevier, Amsterdam, The Netherlands, pp. 73–95.
- Pal, R., and Whitten, M.J., Eds., (1974). *The Use of Genetics In Insect Control*, Elsevier, Amsterdam, The Netherlands.
- Pal-Bhadra, M., Bhadra, U., and Birchler, J.A., (1999). Cosuppression of nonhomologous transgenes in *Drosophila* involves mutually related endogenous sequences. *Cell* **99**: 35–46.
- Pan, X., Zhou, G., Wu, J., Bian, G., Lu, P., Raikhel, A.S., and Xi, Z., (2012). *Wolbachia* induces reactive oxygen species (ROS)-dependent activation of the Toll pathway to control dengue virus in the mosquito *Aedes aegypti*. *Proc. Natl. Acad. Sci. USA* **109**: E23–E31.
- Papathanos, P.A., Bossin, H.C., Benedict, M.Q., Catteruccia, F., Malcolm, C.A., Alphey, L., and Crisanti, A., (2009). Sex separation strategies: past experience and new approaches. *Malar. J.* **8 (Suppl. 2)**: 55.
- Pelisson, A., Teysset, L., Chalvet, F., Kim, A., Prud'homme, N., Terzian, C., and Bucheton, A., (1997). About the origin of retroviruses and the co-evolution of the *gypsy* retrovirus with the *Drosophila flamenco* host gene. *Genetica* **100**: 29–37.
- Peloquin, J.J., and Miller, T.A., (2000). Enhanced green fluorescent protein DNA detection in pink bollworm (*Pectinophora gossypiella*) through polymerase chain reaction amplification. *J. Cotton Sci.* **4**: 19–27.
- Peloquin, J.J., Thibault, S.T., Schouest, L.P., and Miller, T.A., (1997). Electromechanical microinjection of pink bollworm *Pectinophora gossypiella* embryos increases survival. *Biotechniques* **22**: 496–499.
- Peloquin, J.J., Thibault, S.T., Staten, R., and Miller, T.A., (2000a). Germ-line transformation of pink bollworm (Lepidoptera: Gelechiidae) mediated by the *piggyBac* transposable element. *Insect Mol. Biol.* **9**: 323–333.
- Peloquin, J.J., Kuzina, L., Lauzon, C.R., and Miller, T.A., (2000b). Transformation of internal extracellular bacteria isolated from *Rhagoletis completa* Cresson gut with enhanced green fluorescent protein. *Curr. Microbiol.* **40**: 367–371.
- Peronnet, F., San Giorgioi, F., Lepesant, J.A., Deutsch, J.S., and Gonzy-Treboul, G., (2000). Three-partner conversion induced by the P-element transposase in *Drosophila melanogaster*. *Mol. Gen. Genet.* **262**: 1123–1131.
- Perrimon, N., Ni, J.Q., and Perkins, L., (2010). In vivo RNAi: Today and tomorrow. *Cold Spring Harb. Perspect. Biol.* **2**: a003640.
- Peterson, K.R., Clegg, C.H., Li, Q., and Stamatoyannopoulos, G., (1997). Production of transgenic mice with yeast artificial chromosomes. *Trends Genet.* **13**: 61–66.
- Petrov, D.A., Schutzman, J.L., Hartl, D.L., and Lozovskaya, E.R., (1995). Diverse transposable elements are mobilized in hybrid dysgenesis in *Drosophila virilis*. *Proc. Natl. Acad. Sci. USA* **93**: 8050–8054.
- Pew Initiative on Food and Biotechnology, (2004). *Bugs in the System? Issues in the Science and Regulation of Genetically Modified Insects*. Pew Trust, Washington, DC.
- Phuc, H.K., Andreasen, M.H., Burton, R.S., Vass, C., Epton, M.J., and Pape, G., et al. (2007). Late-acting dominant lethal genetic systems and mosquito control. *BMC Biol.* **5**: 11. doi: [10.1186/1741-7007-5-11](https://doi.org/10.1186/1741-7007-5-11).
- Pinkerton, A.C., Michael, K., O'Brochta, D.A., and Atkinson, P.W., (2000). Green fluorescent protein as a genetic marker in transgenic *Aedes aegypti*. *Insect Mol. Biol.* **9**: 1–10.

- Pitino, M., Coleman, A.D., Maffei, M.E., Ridout, C.J., and Hogenhout, S.A., (2011). Silencing of aphid genes by dsRNA feeding from plants. *PLoS ONE* **6** (10): e25709.
- Plasterk, R.H.A., (1993). Molecular mechanisms of transposition and its control. *Cell* **74**: 781–786.
- Plasterk, R.H.A., Izsvák, Z., and Ivics, Z., (1999). Resident aliens. The *Tc1/mariner* superfamily of transposable elements. *Trends Genet.* **15**: 326–332.
- Presnail, J.K., and Hoy, M.A., (1992). Stable genetic transformation of a beneficial arthropod by microinjection. *Proc. Natl. Acad. Sci. USA* **89**: 7726–7732.
- Presnail, J.K., and Hoy, M.A., (1994). Transmission of injected DNA sequences to multiple eggs of *Metaseiulus occidentalis* and *Amblyseius finlandicus* (Acari: Phytoseiidae) following maternal microinjection. *Exp. Appl. Acarol.* **18**: 319–330.
- Presnail, J.K., and Hoy, M.A., (1996). Maternal microinjection of the endoparasitoid *Cardiochiles diaphaniae* (Hymenoptera: Braconidae). *Ann. Entomol. Soc. Am.* **89**: 576–580.
- Presnail, J.K., Jeyaprakash, A., Li, J., and Hoy, M.A., (1997). Genetic analysis of four lines of *Metaseiulus occidentalis* (Nesbitt) (Acari: Phytoseiidae) transformed by maternal microinjection. *Ann. Entomol. Soc. Am.* **90**: 237–245.
- Price, D.R.G., and Gatehouse, J.A., (2008). RNAi-mediated crop protection against insects. *Trends Biotechnol.* **26**: 393–400.
- Prout, T., (1994). Some evolutionary possibilities for a microbe that causes incompatibility in its host. *Evolution* **48**: 909–911.
- Raphael, K.A., Shearman, D.C.A., Streamer, K., Morrow, J.L., Handler, A.M., and Frommer, M., (2011). Germ-line transformation of the Queensland fruit fly, *Bactrocera tryoni*, using a *piggyBac* vector in the presence of endogenous *piggyBac* elements. *Genetics* **139**: 91–97.
- Rasgon, J.L., (2008). Using predictive models to optimize *Wolbachia*-based strategies for vector-borne disease control. In: Aksoy, S., Ed., *Transgenesis and the Management of Vector-Borne Disease*, Landes Bioscience, Austin, TX, pp. 114–125.
- Reeves, R.G., Denton, J.A., Santucci, F., Bryk, J., and Reed, F.A., (2012). Scientific standards and the regulation of genetically modified insects. *PLoS Negl. Trop. Dis.* **6** (1): e1502.
- Ren, X., Hoiczyk, E., and Rasgon, J.L., (2008). Viral paratransgenesis in the malaria vector *Anopheles gambiae*. *PLoS Pathogens* **4** (8): e1000135.
- Revuelta, L., Piulahs, M.D., Belles, X., Castanera, P., Ortego, F., and Diaz-Ruiz, J.R., et al. (2009). RNAi of ace1 and ace2 in *Blattella germanica* reveals their differential contribution to acetylcholinesterase activity and sensitivity to insecticides. *Insect Biochem. Mol. Biol.* **39**: 913–919.
- Richards, F.F., (1993). An approach to reducing arthropod vector competence. Dispersion of insect-borne diseases could be modified by genetically altered symbionts. *Amer. Soc. Microbiol. News* **59**: 509–514.
- Rio, R.V.M., Hu, Y., and Aksoy, S., (2004). Strategies of the home-team: Symbioses exploited for vector-borne disease control. *Trends Microbiol.* **12**: 325–336.
- Robertson, H.M., (1993). The *mariner* transposable element is widespread in insects. *Nature* **362**: 241–245.
- Robertson, H.M., (1995). The *Tc1-mariner* superfamily of transposons in animals. *J. Insect Physiol.* **41**: 99–105.
- Robertson, H.M., (1997). Multiple *mariner* transposons in flatworms and hydroids are related to those of insects. *J. Hered.* **88**: 195–201.
- Robertson, H.M., and Lampe, D.J., (1995). Distribution of transposable elements in arthropods. *Annu. Rev. Entomol.* **49**: 333–337.
- Robertson, H.M., and Martos, R., (1997). Molecular evolution of the second ancient human *mariner* transposon, *Hsmar2*, illustrates patterns of neutral evolution in the human genome lineage. *Gene* **205**: 219–228.

- Robinson, A.S., and Franz, G., (2000). The application of transgenic insect technology in the sterile insect technique. In: Handler, A.M., and James, A.A., Eds., *Insect Transgenesis. Methods and Applications*, CRC Press, Boca Raton, FL, pp. 307–318.
- Robinson, K.O., Ferguson, H.J., Cobey, S., Vaessin, H., and Smith, B.H., (2000). Sperm-mediated transformation of the honey bee. *Apis mellifera*. *Insect Mol. Biol.* **9**: 625–634.
- Rothenbuhler, W.C., (1979). Semidomesticated insects: honeybee breeding. In: Hoy, M.A., and McKelvey Jr., J.J., Eds., *Genetics in Relation to Insect Management*, Rockefeller Foundation Press, New York, pp. 84–92.
- Royer, C., Jalabert, A., Da Rocha, M., Grenier, A.M., Mauchamp, B., Couble, P., and Chavancy, G., (2005). Biosynthesis and cocoon-export of a recombinant globular protein in transgenic silkworms. *Transgenic Res.* **14**: 463–472.
- Rubin, G.M., and Spradling, A.C., (1982). Genetic transformation of *Drosophila* with transposable element vectors. *Science* **218**: 348–353.
- Ruesink, J.L., Parker, I.M., Groom, M.J., and Kareiva, P.M., (1995). Reducing the risks of nonindigenous species introductions: Guilty until proven innocent. *Bioscience* **45**: 465–477.
- Sabesan, S., and Jambulingam, P., (2012). What ails *Wolbachia* transinfection to control disease vectors? *Trends Parasitol.* **28**: 1–2.
- Saleh, M.C., Rij, R.P.V., Gillis, A., Foley, E., O'Farrell, P.H., and Andino, R., (2006). The endocytic pathway mediates cell entry of dsRNA to induce RNA silencing. *Nat. Cell Biol.* **8**: 793–802.
- Santamaría, P., (1986). Injecting eggs. In: Roberts, D.B., Ed., *Drosophila A Practical Approach*, IRL Press, Oxford, UK, pp. 159–173.
- Sarkar, A., and Collins, F.H., (2000). Eye color genes for selection of transgenic insects. In: Handler, A.M., and James, A.A., Eds., *Insect Transgenesis. Methods and Applications*, CRC Press, Boca Raton, FL, pp. 79–91.
- Schetelig, M.F., and Handler, A.M., (2012a). Strategy for enhanced transgenic strain development for embryonic conditional lethality in *Anastrepha suspensa*. *Proc. Natl. Acad. Sci. USA* **109**: 9348–9353.
- Schetelig, M.F., and Handler, A.M., (2012b). A transgenic embryonic sexing system for *Anastrepha suspensa* (Diptera: Tephritidae). *Insect Biochem. Mol. Biol.* **42**: 790–795.
- Schetelig, M.F., Caceres, C., Zacharopoulou, A., Franz, G., and Wimmer, E.A., (2009a). Conditional embryonic lethality to improve the sterile insect technique in *Ceratitis capitata* (Diptera: Tephritidae). *BMC Biol.* **7**: 4.
- Schetelig, M.F., Scolari, F., Handler, A.M., Kittelmann, S., Gasperi, G., and Wimmer, E.A., (2009b). Site-specific recombination for the modification of transgenic strains of the Mediterranean fruit fly *Ceratitis capitata*. *Proc. Natl. Acad. Sci. USA* **106**: 18171–188176.
- Schetelig, M.F., Gotschel, F., Viktorinova, I., Handler, A.M., and Wimmer, E.A., (2011). Recombination technologies for enhanced transgene stability in bioengineered insects. *Genetica* **139**: 71–78.
- Schlüns, H., and Crozier, R.H., (2007). Relish regulates expression of antimicrobial peptide genes in the honeybee, *Apis mellifera*, shown by RNA interference. *Insect Mol. Biol.* **16**: 579–753.
- Schubbert, R., Renz, D., Schmitz, B., and Doerfler, W., (1997). Foreign (M13) DNA ingested by mice reaches peripheral leukocytes, spleen, and liver via the intestinal wall mucosa and can be covalently linked to mouse DNA. *Proc. Natl. Acad. Sci. USA* **94**: 961–966.
- Schubbert, R., Hohlweg, U., Renz, D., and Doerfler, W., (1998). On the fate of orally ingested foreign DNA in mice: Chromosomal association and placental transmission to the fetus. *Mol. Gen. Genet.* **259**: 569–576.
- Scolari, F., Siciliano, P., Gabrieli, P., Gomulski, L.M., Bonomi, A., Gasperi, G., and Malacrida, A.R., (2011). Safe and fit genetically modified insects for pest control: From lab to field applications. *Genetica* **139**: 41–52.
- Scott, T.W., Harrington, L.C., Knols, B.G.J., and Takken, W., (2008). Applications of mosquito ecology for successful insect transgenesis-based disease prevention programs. In: Aksoy, S.,

- Ed., *Transgenesis and the Management of Vector-Borne Disease*, Landes Bioscience, Austin, TX, pp. 151–168.
- Shamila, Y., and Mathavan, S., (1998). Gene transfer in silkworm *Bombyx mori* via electroporation. *Biotechnol. Technol.* **12**: 507–510.
- Sherman, A., Dawson, A., Mather, C., Gilhooley, H., Li, Y., and Mitchell, R., et al. (1998). Transposition of the *Drosophila* element *mariner* into the chicken germ line. *Nature Biotechnol.* **16**: 1050–1053.
- Shinmyo, Y., Mito, T., Matsushita, T., Sarashina, I., Miyawaki, K., Ohuchi, H., and Noji, S., (2005). *caudal* is required for gnathal and thoracic patterning and for posterior elongation in the intermediate-germband cricket. *Gryllus bimaculatus. Mech. Dev.* **122**: 231–239.
- Shukla, J.N., and Palli, S.R., (2012). Sex determination in beetles: production of all male progeny by parental RNAi knockdown of *transformer*. *Sci. Rep.* **2**: 602.
- Shuler, M.L., Wood, H.A., Granados, R.R., and Hammer, D.A., Eds., (1994). *Baculovirus Expression Systems and Biopesticides*, Wiley-Liss, New York.
- Sijen, T., and Kooter, J.M., (2000). Post-transcriptional gene-silencing: RNAs on the attack or on the defense? *Bioessays* **22**: 520–531.
- Simmons, G.S., McKemey, A.R., Morrison, N.I., O'Connell, S., Tabashnik, B.E., and Claus, J., et al. (2011). Field performance of a genetically engineered strain of the pink bollworm. *PLoS ONE* **6** (9): e24110.
- Simpson, P., (1993). Flipping fruit-flies: a powerful new technique for generating *Drosophila* mosaics. *Trends Genet.* **9**: 227–228.
- Sinkins, S.P., and O'Neill, S.L., (2000). *Wolbachia* as a vehicle to modify insect populations. In: Handler, A.M., and James, A.J., Eds., *Insect Transgenesis. Methods and Applications*, CRC Press, Boca Raton, FL, pp. 271–287.
- Sinkins, S.P., and Gould, F., (2006). Gene drive systems for insect disease vectors. *Nature Rev. Genet.* **7**: 427–435.
- Spielman, A., (1994). Why entomological antimalaria research should not focus on transgenic mosquitoes. *Parasitol. Today* **10**: 374–376.
- Spielman, A., Beier, J.C., and Kiszewski, A.E., (2002). Ecological and community considerations in engineering arthropods to suppress vector-borne disease. In: Letourneau, D.K., and Burrows, B.E., Eds., *Genetically Engineered Organisms. Assessing Environmental and Human Health Effects*, CRC Press, Boca Raton, FL, pp. 325–329.
- Spradling, A.C., and Rubin, G.M., (1982). Transposition of cloned *P* elements into *Drosophila* germline chromosomes. *Science* **218**: 341–347.
- Stouthamer, R., Luck, R.F., and Werren, J.H., (1992). Genetics of sex determination and the improvement of biological control using parasitoids. *Environ. Entomol.* **21**: 427–435.
- Subbaraman, N., (2011). Science snipes at Oxitec transgenic-mosquito trial. *Nature Biotech.* **29**: 9–11.
- Sundararajan, P., Atkinson, P.W., and O'Brochta, D.A., (1999). Transposable element interactions in insects: cross mobilization of *hobo* and *Hermes*. *Insect Mol. Biol.* **8**: 359–368.
- Takagi, A., Kurita, K., Terasawa, T., Nakamura, T., Bando, T., Moriyama, Y., Mito, T., Noji, S., and Ohuchi, H., (2012). Functional analysis of the role of *eyes absent* and *sine oculis* in the developing eye of the cricket *Gryllus bimaculatus*. *Devel. Growth Differ.* doi: [10.1111/j.1440-169X.2011.01325.x](https://doi.org/10.1111/j.1440-169X.2011.01325.x).
- Takasu, Y., Kobayashi, I., Beumer, K., Uchino, K., Sezutsu, H., Sajwan, S., Carroll, D., Tamura, T., and Zurovec, M., (2010). Targeted mutagenesis in the silkworm *Bombyx mori* using zinc finger nuclease mRNA injection. *Insect Biochem. Mol. Biol.* **40**: 759–765.
- Tan, K.H., Ed., (2000). *Area-Wide Control of Fruit Flies and Other Insect Pests*, Penerbit Universiti Sains Malaysia, Penang.
- Tatematsu, K., Kobayashi, I., Uchino, K., Sezutsu, H., Izuka, T., Yonemura, N., and Tamura, T., (2010). Construction of a binary transgenic gene expression system for recombinant protein

- production in the middle silk gland of the silkworm *Bombyx mori*. *Transgenic Res.* **19**: 473–487.
- Terenius, O., Papanicolaou, A., Garbutt, J.S., Eleftherianos, I., and Huvenne, H., et al. (2011). RNA interference in Lepidoptera: an overview of successful and unsuccessful studies and implications for experimental design. *J. Insect Physiol.* **57**: 231–245.
- Teule, F., Miao, Y.G., Sohn, B.H., Kim, Y.S., Hull, J.J., and Fraser, M.J., et al. (2012). Silkworms transformed with chimeric silkworm/spider silk genes spin composite silk fibers with improved mechanical properties. *Proc. Natl. Acad. Sci. USA* **109**: 923–928.
- Thailayil, J., Magnusson, K., Godfray, H.C.J., Crisanti, A., and Catteruccia, F., (2011). Spermless males elicit large-scale female responses to mating in the malaria mosquito *Anopheles gambiae*. *Proc. Natl. Acad. Sci. USA* **108**: 13677–13681.
- Thomas, D.D., Donnelly, C.A., Wood, R.J., and Alphey, L.S., (2000). Insect population control using a dominant, repressible, lethal genetic system. *Science* **298**: 2474–2476.
- Tiedje, J.M., Colwell, R.K., Grossman, Y.L., Hodson, R.E., Lenski, R.E., Mack, R.M., and Regal, P.J., (1989). The planned introduction of genetically engineered organisms: ecological considerations and recommendations. *Ecology* **70**: 298–315.
- Tomoyasu, Y., and Denell, R.E., (2004). Larval RNAi in *Tribolium* (Coleoptera) for analyzing adult development. *Dev. Genes Evol.* **214**: 575–578.
- Tomoyasu, Y., Miller, S.C., Tomita, S., Schoppmeier, N., Grossmann, D., and Bucher, G., (2008). Exploring systemic RNA interference in insects: a genome-wide survey for RNAi genes in *Tribolium*. *Genome Biol.* **9**: R10.
- Torres, L., Almazan, C., Aylion, N., Galindo, R.C., Rosario-Cruz, R., Quiroz-Romero, H., and de la Fuente, J., (2011). Functional genomics of the horn fly, *Haematobia irritans* (Linnaeus, 1758). *BMC Genomics* **12**: 105.
- Toure, Y.T., Oduola, A.M.J., Sommerfeld, J., and Morel, C.M., (2004). Biosafety and risk assessment in the use of genetically modified mosquitoes for disease control. In: Knols, B.G.J., and Louis, U.C., Eds., *Bridging Laboratory and Field Research for Genetic Control of Disease Vectors*, Springer/Frontis, Wageningen, The Netherlands, pp. 217–222.
- Traver, B.E., Anderson, M.A.E., and Adelman, Z.N., (2009). Homing endonucleases catalyze double-stranded DNA breaks and somatic transgene excision in *Aedes aegypti*. *Insect Mol. Biol.* **18**: 623–633.
- Tsitrone, A., Charles, S., and Biemont, C., (1999). Dynamics of transposable elements under the selection model. *Genet. Res. Camb.* **74**: 159–164.
- Turelli, M., and Hoffmann, A.A., (1991). Rapid spread of an inherited incompatibility factor in California. *Drosophila*. *Nature* **353**: 440–442.
- Turelli, M., Hoffmann, A.A., and McKechnie, S.W., (1992). Dynamics of cytoplasmic incompatibility and mtDNA variation in natural *Drosophila simulans* populations. *Genetics* **132**: 713–723.
- Turley, A.P., Moreira, L.A., O'Neill, S.L., and McGraw, E.A., (2009). *Wolbachia* infection reduces blood-feeding success in the dengue fever mosquito, *Aedes aegypti*. *PLoS Negl. Trop. Dis.* **3** (9): e516.
- Turner, C.T., Davy, M.W., MacDiarmid, R.M., Plummer, K.M., Birch, N.P., and Newcomb, R.D., (2006). RNA interference in the light brown apple moth, *Epiphyas postvittana* (Walker) induced by double-stranded RNA feeding. *Insect Mol. Biol.* **15**: 383–391.
- U. S. Department of Agriculture (1991). Part III. Proposed guidelines for research involving the planned introduction into the environment of organisms with deliberately modified hereditary traits; Notice. Federal register Vol. 56 (22), Friday, February 1, 1991, pp. 4134–4151.
- Upadhyay, S.K., Chandrashekhar, K., Thakur, N., Verma, P.C., Borgio, J.F., Singh, P.K., and Tuli, R., (2011). RNA interference for the control of whiteflies (*Bemisia tabaci*) by oral route. *J. Biosci.* **36**: 153–161.
- Urnov, F.D., Rebar, E.J., Holmes, M.C., Zhang, H.S., and Gregory, P.D., (2010). Genome editing with engineered zinc finger nucleases. *Nat. Rev. Genet.* **11**: 636–646.

- USDA-APHIS (2006). Finding of no significant impact and decision notice. Available at: <[www.aphis.usda.gov/plant\\_health/plant\\_pest\\_info/cotton\\_pests/downloads/pbw-erad-plan2-09.pdf](http://www.aphis.usda.gov/plant_health/plant_pest_info/cotton_pests/downloads/pbw-erad-plan2-09.pdf)>
- de Valdez, M.R.W., Nimmo, D., Betz, J., Gong, H.F., James, A.A., and Alphey, L., et al. (2011). Genetic elimination of dengue vector mosquitoes. *Proc. Natl. Acad. Sci. USA* **108**: 4772–4775.
- Vavre, F., and Charlat, S., (2012). Making (good) use of *Wolbachia*: what the models say. *Curr. Opin. Microbiol.* **15**: 263–268.
- Walker, V.K., (1989). Gene transfer in insects. *Adv. Cell Culture* **7**: 87–124.
- Walker, T., Johnson, P.H., Moreira, L.A., Iturbe-Ormaetxe, I., Frentiu, F.D., McMeniman, C.J., and Leong, Y.S., et al. (2011). The wMel *Wolbachia* strain blocks dengue and invades caged *Aedes aegypti* populations. *Nature* **476**: 450–453.
- Walsh, D.P., Lehane, S.M., Lehane, M.J., and Haines, L.R., (2009). Prolonged gene knockdown in the tsetse fly *Glossina* by feeding double stranded RNA. *Insect Mol. Biol.* **18**: 11–19.
- Walters, M., Morrison, N.I., Claus, J., Tang, G., Phillips, C.E., and Young, R., et al. (2012). Field longevity of a fluorescent protein marker in an engineered strain of the pink bollworm, *Pectinophora gossypiella* (Saunders). *PLoS ONE* **7** (6): e38547.
- Wang, W., Swevers, L., and Iatrou, K., (2000). Mariner (*Mos1*) transposase and genomic integration of foreign gene sequences in *Bombyx mori* cells. *Insect Mol. Biol.* **9**: 145–155.
- Wang, J., Miller, E.D., Simmons, G.S., Miller, T.A., Tabashnik, B.E., and Park, Y., (2010). PiggyBac-like elements in the pink bollworm, *Pectinophora gossypiella*. *Insect Mol. Biol.* **19**: 177–184.
- Watanabe, K., and Sato, M., (1998). Plasmid-mediated gene transfer between insect-resident bacteria, *Enterobacter cloacae*, and plant-epiphytic bacteria, *Erwinia herbicola*, in guts of silkworm larvae. *Curr. Microbiol.* **37**: 352–355.
- Watanabe, K., Hara, W., and Sato, M., (1998). Evidence for growth of strains of the plant epiphytic bacterium *Erwinia herbicola* and transconjugation among the bacterial strains in guts of the silkworm *Bombyx mori*. *J. Invertebr. Pathol.* **72**: 104–111.
- Watanabe, T., Ochiai, H., Sakuma, T., Horch, H.W., Hamaguchi, N., and Nakamura, T., et al. (2012). Non-transgenic genome modifications in a hemimetabolous insect using zinc-finger and TAL effector nucleases. *Nat. Commun.* **3**: 1017.
- Webb, B.A., (2000). Polydnaviruses and insect transgenic research. In: Handler, A.M., and James, A.A., Eds., *Insect Transgenesis. Methods and Applications*, CRC Press, Boca Raton, FL, pp. 203–215.
- Werren, J.H., and Bartos, J.D., (2001). Recombination in *Wolbachia*. *Curr. Biol.* **11**: 431–435.
- Werren, J.H., Loehlin, D.W., and Giebel, J.D., (2009). Larval RNAi in *Nasonia* (parasitoid wasp). *Cold Spring Harb. Protocol* **2009** (10) doi: [10.1101/pdb.prot5311](https://doi.org/10.1101/pdb.prot5311).
- White, L.D., Coates, C.J., Atkinson, P.W., and O'Brochta, D.A., (1996). An eye color gene for the detection of transgenic non-drosophilid insects. *Insect Biochem. Mol. Biol.* **26**: 641–644.
- Whitten, M.J., (1979). The use of genetically selected strains for pest replacement or suppression. In: Hoy, M.A., McKelvey Jr., J.J. (Eds.), *Genetics in Relation to Insect Management* Rockefeller Foundation Press, NY, pp. 31–40.
- Whitten, M.J., (1985). The conceptual basis for genetic control. In: Kerkut, G.A., and Gilbert, L.I., Eds., *Comprehensive Insect Physiology Biochemistry and Pharmacology*, Pergamon Press, Oxford, UK, pp. 465–527.
- Whitten, M.J., and Hoy, M.A., (1999). Genetic improvement and other genetic considerations for improving the efficacy and success rate of biological control. In: Bellows, T.S., and Fisher, T.W., Eds., *Principles and Applications of Biological Control*, Academic, San Diego, CA, pp. 271–296.
- Whyard, S., Singh, A.D., and Wong, S., (2009). Ingested double-stranded RNAs can act as species-specific insecticides. *Insect Biochem. Mol. Biol.* **39**: 824–832.
- Williamson, M., (1992). Environmental risks from the release of genetically modified organisms (GMOs)-The need for molecular ecology. *Mol. Ecol.* **1**: 3–8.

- Windbichler, N., Menichelli, M., Papathanos, P.A., Thyme, S.B., Li, H., and Ulge, U.Y., et al. (2011). A synthetic homing endonuclease-based gene drive system in the human malaria mosquito. *Nature* **473**: 212–217.
- Witte, W., (1998). Medical consequences of antibiotic use in agriculture. *Science* **279**: 996–997.
- Wolffe, A.P., (1997). Transcription control: Repressed repeats express themselves. *Curr. Biol.* **7**: R796–R798.
- Wood, S.N., and Thomas, M.B., (1999). Super-sensitivity to structure in biological models. *Proc. R. Soc. Lond. B* **266**: 565–570.
- Wright, J.W., and Pal, R., Eds., (1967). *Genetics of Insect Vectors of Disease*, Elsevier, Amsterdam, The Netherlands.
- Wright, D.A., Thibodeau-Beganny, S., Sander, J.D., Winfrey, R.J., Hirsh, A.S., Eichtinger, M., and Fu, F., et al. (2006). Standardized reagents and protocols for engineering zinc finger nucleases by modular assembly. *Nature Protocols* **1** (3): 1627–1652.
- Wyss, J.H., (2000). Screw-worm eradication in the Americas—overview. In: Tan, K.H., Ed., *Area-Wide Control of Fruit Flies and Other Insect Pests*, Penerbit Universiti Sains Malaysia, Penang, pp. 79–86.
- Yamamoto, D.S., Nagumo, H., and Yoshida, S., (2010). Flying vaccinator: a transgenic mosquito delivers a *Leishmania* vaccine via blood feeding. *Insect Mol. Biol.* **19**: 391–398.
- Yang, G., You, M., Vasseur, L., Zhao, Y., and Liu, C., (2011). Development of RNAi in insects and RNAi-based pest control. In: Stoytcheva, M., Ed., *Pesticides in the Modern World-Pests Control and Pesticides Exposure and Toxicity Assessment*, pp. 27–38. InTech (<http://www.intechopen.com/books/editor/pesticides-in-the-modern-world-pests-control-and-pesticides-exposure-and-toxicity-assessment>).
- Yeap, H.L., Mee, P., Walker, T., Weeks, A.R., and O'Neill, S.L., et al. (2011). Dynamics of the “Popcorn” *Wolbachia* infection in outbred *Aedes aegypti* informs prospects for mosquito vector control. *Genetics* **187**: 583–595.
- Yen, J.H., and Barr, A.R., (1974). Incompatibility in *Culex pipiens*. In: Pal, R., and Whitten, M.J., Eds., *The Use of Genetics In Insect Control*, Elsevier, Amsterdam, The Netherlands, pp. 97–118.
- Yoder, J.I., and Goldsbrough, A.P., (1994). Transformation systems for generating marker-free transgenic plants. *BioTechnology* **12**: 263–267.
- Yokoyama, T., (1979). Silkworm selection and hybridization. In: Hoy, M.A., and McKelvey Jr., J.J., Eds., *Genetics in Relation to Insect Management*, Rockefeller Foundation Press, New York, pp. 71–83.
- Yoshiyama, M.K., Kimura, and Honda, H., (2000). Technique for microinjection into eggs of the housefly, *Musca domestica* (Diptera: Muscidae). *Appl. Entomol. Zool.* **35**: 27–30.
- Young, O.P., Ingebretsen, S.P., and Foudin, A.S., (2000). Regulation of transgenic arthropods and other invertebrates in the United States. In: Handler, A.M., and James, A.A., Eds., *Insect Transgenesis. Methods and Applications*, CRC Press, Boca Raton, FL, pp. 369–379.
- Yu, Z., Podgwaite, J.D., and Wood, H.A., (1992). Genetic engineering of a *Lymantria dispar* nuclear polyhedrosis virus for expression of foreign genes. *J. Gen. Virol.* **73**: 1509–1514.
- Yu, N., Christiaens, O., Liu, J., Niu, J., Cappelle, K., and Caccia, S., et al. (2012). Delivery of dsRNA for RNAi in insects: an overview and future directions. *Insect Sci.* **00**: 1–11.  
doi: [10.1111/j.1744-7917.2012.01534.x](https://doi.org/10.1111/j.1744-7917.2012.01534.x).
- Zalokar, M., (1981). A method for injection and transplantation of nuclei and cells in *Drosophila* eggs. *Experientia* **37**: 1354–1356.
- Zha, W., Peng, X., Chen, R., Du, B., Zhu, L., and He, G., (2011). Knockdown of midgut genes by dsRNA-transgenic plant-mediated RNA interference in the hemipteran insect *Nilaparvata lugens*. *PLoS ONE* **6**: e20504.
- Zhang, X., Zhang, J., and Zhu, K.Y., (2010). Citosan/double-stranded RNA nanoparticle-mediated RNA interference to silence chitin synthase genes through larval feeding in the African malaria mosquito (*Anopheles gambiae*). *Insect Mol. Biol.* **19**: 683–693.

- Zhao, A., Zhao, T., Zhang, Y., Xia, Q., Lu, C., Zhou, Z., Xiang, Z., and Nakagaki, M., (2010). New and highly efficient expression systems for expressing selectively foreign protein in the silk glands of transgenic silkworm. *Transgenic Res.* **19**: 29–44.
- Zhou, X., Oi, F.M., and Scharf, M.E., (2006). Social exploitation of hexamerin: RNAi reveals a major caste-regulatory factor in termites. *Proc. Natl. Acad. Sci. USA* **103**: 4499–4504.
- Zhou, X., Wheeler, M.M., Oi, F.M., and Scharf, M.E., (2008). RNA interference in the termite *Reticulitermes flavipes* through ingestion of double-stranded RNA. *Insect Biochem. Mol. Biol.* **38**: 805–815.
- Zwiebel, L.J., Saccone, G., Zacharopoulou, A., Besansky, N.J., Favia, G., Collins, F.H., Louis, C., and Kafatos, F.C., (1995). The white gene of *Ceratitis capitata*: a phenotypic marker for germline transformation. *Science* **270**: 2005–2008.

# **Insect Molecular Genetics**

An Introduction to Principles and Applications

# Insect Molecular Genetics

## An Introduction to Principles and Applications

Third Edition

Marjorie A. Hoy



AMSTERDAM • BOSTON • HEIDELBERG • LONDON  
NEW YORK • OXFORD • PARIS • SAN DIEGO  
SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO  
Academic Press is an imprint of Elsevier



Academic Press is an imprint of Elsevier  
32 Jamestown Road, London NW1 7BY, UK  
225 Wyman Street, Waltham, MA 02451, USA  
525 B Street, Suite 1800, San Diego, CA 92101-4495, USA

Copyright © 2013 Elsevier Inc. All rights reserved

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means electronic, mechanical, photocopying, recording or otherwise without the prior written permission of the publisher. Permissions may be sought directly from Elsevier's Science & Technology Rights Department in Oxford, UK: phone (+44) (0) 1865 843830; fax (+44) (0) 1865 853333; email: [permissions@elsevier.com](mailto:permissions@elsevier.com). Alternatively, visit the Science and Technology Books website at [www.elsevierdirect.com/rights](http://www.elsevierdirect.com/rights) for further information.

#### Notice

No responsibility is assumed by the publisher for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions or ideas contained in the material herein. Because of rapid advances in the medical sciences, in particular, independent verification of diagnoses and drug dosages should be made.

#### British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library

#### Library of Congress Cataloguing-in-Publication Data

A catalog record for this book is available from the Library of Congress

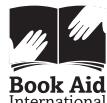
ISBN: 978-0-12-415874-0

For information on all Academic Press publications  
visit our website at [elsevierdirect.com](http://elsevierdirect.com)

Typeset by MPS Limited, Chennai, India  
[www.adimps.com](http://www.adimps.com)

Printed and bound in United States of America

13 14 15 16 10 9 8 7 6 5 4 3



Working together  
to grow libraries in  
developing countries

[www.elsevier.com](http://www.elsevier.com) • [www.bookaid.org](http://www.bookaid.org)

## Preface to the Third Edition

Once again, I must acknowledge the very rapid pace of advances in arthropod molecular genetics. Because ticks and mites are increasingly included in molecular genetic studies, especially now that the genomes of a tick, the twospotted spider mite, *Varroa* mite, and a predatory mite have been sequenced, I use the term arthropod often in this edition. Since the year 2000, the number of arthropod genomes that have been sequenced has increased dramatically, primarily due to the decreasing costs of sequencing using Next-Generation methods. The cover of edition three contains photos of many of the arthropods for which complete genome-sequence data are available. A proposal was made in 2011 to sequence 5000 insect genomes in the next few years and, as a result, the cover of this book would soon be unable to "host" all the photographs of arthropods with completely sequenced genomes. The enormous amount of genomic data being developed will affect all aspects of the study of arthropods, including physiology, phylogenetics, ecology, evolution, behavior, and applied pest-management methods. The release of transgenic insects into the environment is receiving increased focus, now that releases into the environment of transgenic pink bollworms, predatory mites, and mosquitoes have occurred.

A very diverse group of scientists use molecular genetic tools to solve problems that can be answered in no other manner, so it is critical that all entomologists (and acarologists) become familiar with the terminology, concepts, and tools of molecular genetics—even if they do not wish to do such research themselves. Once again, I am writing this book for students and advanced scientists who have a limited background in insect molecular genetics. I hope to provide a portal into the concepts, methods, and applications of these powerful tools. Because the field has expanded so much, I cannot cite all relevant papers and have tended to include reviews and partial lists to give the reader a way to access the literature. I apologize to those scientists who have published excellent papers on diverse topics of relevance, but space considerations limited the examples that can be cited.

After discussing what material should be covered in this edition with several colleagues, I have once again provided simple protocols in the methods chapters (5–9) in an effort to explain what is being done when DNA or RNA are being extracted, genes are being cloned and sequenced, genomic or transcriptomic libraries are developed, and the polymerase chain reaction is being conducted. Although numerous commercial kits are available and widely used, the recommendations were that it is instructive for novices to gain an understanding as to what is being achieved (and how) when carrying out these procedures. The precise details of kit methodologies are often considered proprietary information, but the protocols provided here may make it possible for students to understand the concepts sufficiently that they can learn to “trouble shoot” if problems occur in their own projects.

All chapters have been revised and updated. Three areas have changed dramatically since the second edition: 1) The development of genetically modified insects for potential use in pest-management programs has advanced to the point that several releases have been made. 2) The development of low-cost Next-Generation sequencing methods permit the sequencing of arthropod genomes to be commonplace, although the explosion in genome (and transcriptome) sequencing data threatens to overwhelm our ability to analyze and use the data. Perhaps the next major advance will be the availability of new tools for bioinformatics analyses so that even novices can effectively analyze genomic data. It now costs US\$10,000–\$30,000 to sequence a genome, but the cost of analyzing the data may be ten times that. Even newer sequencing methods just developed will reduce the time and cost to sequence genomes. 3) Finally, the use of molecular methods has allowed immense increases in our knowledge of the diversity and roles that microbial symbionts play in the biology of their arthropod hosts. This expansion in knowledge is reflected in the publication of several books devoted to insect symbiosis. It has been difficult to choose among the many interesting examples of symbiosis, but I have included examples that illustrate the impact that symbionts have on the physiology, behavior, sex determination, reproductive compatibility, and pest status of their hosts.

I thank the students and others who have provided constructive feedback on the earlier editions of this book and Ke Wu, Niklaus Hostettler, Jennifer Gillett-Kaufman, and Aaron Pomerantz, who assisted me in the preparation of this edition. I especially thank Geoffrey Attardo, Susan Brown, Lyle Buss, Chun-che Chang, Michael Clark, Cameron Currie, Marian Goldsmith, Thom Kaufman, Jason Meyer, James Newman, Jeffrey Stuart, and Yoshi Tomoyasu for their excellent photographs of the arthropod species illustrated on the cover of this book for which genome-sequence data are available.

My very best wishes for a productive and interesting read. I believe that molecular-genetic tools will help to resolve many critical issues in arthropod biology, evolution, and pest management.

*Marjorie A. Hoy  
Eminent Scholar and Davies,  
Fischer and Eckes Professor of Biological Control,  
Department of Entomology and Nematology,  
University of Florida, Gainesville  
September 2012*

# Acknowledgments

This book would not have been written without the encouragement of many people. As most authors remark, it was more difficult to write than expected, in part due to a move from the University of California at Berkeley to the University of Florida in Gainesville about half-way through the endeavor. Certainly, this book would not have been written without the support of the Rockefeller Foundation who provided me with 5 weeks of precious time at the Bellagio Study and Conference Center in Italy. I gained a crucial grasp on the scope of the project and gathered my courage there while on sabbatical from the University of California, Berkeley. A fellowship spent at Cold Spring Harbor Laboratories in the summer of 1985 in the "Molecular Cloning of Eukaryotic Genes" course provided my initial training in molecular genetic techniques. I thank the instructors for their patience. Likewise, G. M. Rubin kindly allowed me to participate in his cloning course at the University of California, Berkeley.

Many people have contributed information, advice, and valuable time reviewing this book. I especially thank Mary Bownes, Michael Caprio, Gary Carvalho, Howell Daly, Owain Edwards, Marilyn Houck, James Hoy, A. Jeyaprakash, Srini Kambhampati, Carolyn Kane, James Presnail, Veronica Rodriguez, Mark Tanouye, and Tom Walker, who either reviewed drafts of individual chapters or the entire book. Richard Beeman, Owain Edwards, Glenn Hall, A. Jeyaprakash, Ed Lewis, Jim Presnail, and A. Zacharopoulou provided photographs or illustrations. Lois Caprio and Denise Johanowicz assisted with many of the logistical issues. Finally, I especially thank my husband Jim, who repeatedly motivated me to clarify and elucidate the principles and applications presented here and tolerated my preoccupation with this project over many months, and John Capinera, who patiently waited for me to emerge from my compulsion. Despite the best efforts of the reviewers, errors probably persist and are my responsibility.

Marjorie A. Hoy  
University of California, Berkeley  
University of Florida, Gainesville

# Preface to the First Edition

The development of recombinant DNA techniques during the past 20 years has resulted in exciting advances in the detailed study of specific genes at the molecular level as well as breakthroughs in molecular, cellular, and developmental biology. Of the molecular genetics studies conducted on insects, most have been directed to *Drosophila melanogaster*. Relatively few data have been generated by molecular biological methods from analyses of other insects. Yet, the application of molecular genetics to insects other than *Drosophila* has the potential to revolutionize insect population and organismal biology.

Why have molecular genetic techniques been used so little by entomologists? There may be a number of reasons. Recombinant DNA techniques are most readily carried out by people trained in biochemistry and relatively few entomologists are so trained. The techniques have been, until recently, relatively complex and difficult, so that strong technical skills were required. Also, most entomologists have been slow to ask whether these techniques were appropriate for studies of population or organismal biology because much of the published literature has focused on fundamental issues of *Drosophila* gene structure, regulation and function, developmental regulation, and evolution.

## Goals

My goal is to introduce entomologists to the concepts of molecular genetics without assuming that they have received previous training in molecular biology. This book is not intended to substitute for formal training in biochemistry or molecular genetics. If novice readers wish to develop molecular genetics skills, they must obtain additional training in genetics and biochemistry. However, the book will provide an introduction to terminology, as well as an overview of principles, techniques, and possible applications of molecular genetics to problems of interest to entomologists.

In preference to using examples from the *Drosophila* literature, I have used examples in which other arthropods have been studied. However, without doubt, *Drosophila* is the premier model for insect molecular genetics study. One fond hope is that this book will be a bridge for entomologists seeking to apply the exciting methods developed for *Drosophila* and that it will introduce *Drosophila* workers to some of the problems and issues of interest to entomologists seeking to solve applied problems. Perhaps this book will help to break down the barriers between entomologists and *Drosophila* workers isolated from each other by perspective and technical jargon. If this book helps to achieve these goals, it will have served its purpose.

## Organization

The book was designed for a one-semester course in insect molecular genetics for upper division undergraduates or beginning graduate students. The initial portion of the book reviews basic information about DNA, RNA, and other important molecules (Chapters 1–4). Readers with a recent course in genetics could skip this section. Chapter 5 describes the genetic systems found in insects and an overview of development sufficient to understand subsequent techniques such as *P*-element-mediated transformation and sex determination. Chapters 6–9 provide introductions to useful techniques, including cloning, library construction, sequencing, the polymerase chain reaction, and *P* element-mediated transformation of *Drosophila*. Most molecular biologists reading this book could skip this section as well. Chapters 6–9 are not intended as a laboratory manual, but an outline of laboratory protocols is provided in order to furnish the novice with a sense of the complexity or simplicity of the procedures and some of the issues to consider in problem solving. Throughout the book, references are provided for the reader interested in pursuing specific topics and techniques, although they are not exhaustive. Despite the value of providing an historical overview, I have not always provided references to the first publication on a subject. Rather, review articles or recent publications that include references to earlier work are cited.

Finally, in the third section of the book (Chapters 10–16), I have attempted to demonstrate how molecular genetic techniques can solve a diverse array of basic and applied problems. Part III is intended to introduce readers to the exciting molecular research that is revolutionizing insect biology, ecology, systematics, behavior, physiology, development, sex determination, and pest management. Each chapter in this section could be read by itself, assuming that reader understands the appropriate concepts or information presented in Parts I and II.

Each chapter begins with an overview or brief summary of the material being covered. The overview should be read both before and after reading each chapter to review the concepts covered. The overview is followed by a brief introduction covering the history or rationale for the topic. References at the end of the chapter are provided for further reading. Where possible, books or reviews are cited to provide an entry into the literature. Recent references are provided, but no attempt has been made to review all the literature on a specific topic. Simple protocols may be given to provide the flavor of specific techniques, although these are not intended to be complete. References to handbooks or techniques books are also provided at the end of appropriate chapters. When a term that may be unfamiliar is first introduced, it is underlined and a brief definition or description is given in the Glossary at the end of the book. Finally, in Appendix I, a time line of some significant advances in genetics, molecular biology, and insect molecular genetics provides a perspective of the pace with which dramatic advances have been, and continue to be, made.

Progress is rapid in molecular genetics, and this book can only provide an introduction to the principles of insect molecular genetics and some of its applications. It is impossible to provide a complete review of the insect molecular genetics literature in a book of this size. The literature cited includes references to through 1993 and focuses on genetics. It is not intended to be an introduction to all aspects of "molecular entomology," which has been defined as "... a blend of insect science, molecular biology, and biochemistry." The dividing line between molecular entomology and insect molecular genetics is sometimes difficult to resolve.

Shortly before this book went to the publisher, two related books were published: *Molecular Approaches to Fundamental and Applied Entomology*, edited by J. Oakeshott and M. J. Whitten, and *Insect Molecular Science*, edited by J. M. Crampton and P. Eggleston. Both multiauthored books cover some of the topics included here, but assume the reader is familiar with molecular genetic techniques and terminology; they would be daunting for the novice.

# Preface to the Second Edition

Amazing progress has been made in insect molecular genetics since the first edition appeared in 1994. Transformation of insects other than *Drosophila melanogaster* has become an almost routine project. The *Drosophila* Genome Project was completed in 1999 and produced many surprises and promises a fruitful future for mining genes and developing an understanding of genome structure, function and evolution. The mining of this treasure trove of data will require some years of work, but the possibility exists that we ultimately will be able to understand how this insect develops. Insect biology will become synthetic again with the use of genomics, transcriptomics, and proteomics approaches. The complete sequencing of other complex eukaryotic genomes, including that of *Caenorhabditis elegans* and *Homo sapiens*, opened additional doors to compare genome organization, evolution, and gene function.

Molecular methods and technology have changed rapidly in the past few years, with a plethora of new kits available for extracting and purifying DNA and RNA, for cloning, sequencing, and amplifying DNA and RNA by the polymerase chain reaction (PCR). Gene chip or microarray methods offer new tools for learning about gene function. All the improvements in these molecular toolkits make molecular methods ever more accessible to the entomological community.

The same basic organization with three major sections has been maintained in this edition, but the chapters have been reorganized in Section I, and all have been updated with recent references. References were included that provide an entry into the recent literature; where possible, review articles are cited. I regret I could not include references to all the new molecular studies on insects, but there are just too many! That signals that molecular entomology is maturing.

This book is dedicated to entomologists just beginning their research careers; I hope this book helps you to start exciting and productive projects that employ these valuable molecular tools. For those of you with no background in molecular genetics, the book should be read from start to finish. Key concepts are

highlighted in the Overview and reading it both before and after reading each chapter may be helpful. The diagrams, especially those illustrating molecular methods, should be evaluated while reading the text. In many cases, the concepts involved are most readily obtained if the text and diagrams are read together.

Finally, I thank all the people who so kindly provided feedback on the first edition and those who made thoughtful suggestions on earlier drafts of this one, including A. Jeyaprakash, Lucy Skelley, Juan Alvarez, and Alison Walker. I sincerely thank those who kindly provided new illustrations, Bruce Webb, Carol Lauzon, and Felix Guererro. Thanks to Pam Howell, Mike Sanford, and Pat Hope for their assistance in getting this manuscript and its illustrations completed.

*Marjorie A. Hoy  
Department of Entomology and Nematology  
University of Florida, Gainesville*

# Glossary

## A

- A-DNA:** Right-handed helical form of DNA found in fibers at 75% relative humidity in presence of sodium, potassium or cesium. The bases are tilted with regard to the helical axis and there are more base pairs per turn. The A-form may be very similar to the conformation adopted by DNA-RNA hybrids or by RNA-RNA double-stranded regions.
- acentric:** A chromosome, or chromosome fragment, that lacks a centromere.
- actin:** A protein that is the major constituent of the microfilaments of cells and in muscle. All actins studied are similar in size and amino acid sequence, suggesting they have evolved from a single gene.
- action potential:** A rapid change in the polarity of the membrane of a neuron that facilitates the interaction and transmission of impulses.
- additive genes:** Genes that interact but do not show dominance or epistasis.
- adenine (A):** A purine; one of the nitrogenous bases found in DNA and RNA. A typically pairs with thymine (T).
- adenosine triphosphate:** See ATP.
- Africanized bees:** Honey bees in the Western Hemisphere that are derived from hybridization of African and European subspecies of *Apis mellifera*. The degree of hybridization is unresolved.
- agarose:** A polysaccharide gum obtained from agar, which is obtained from certain seaweeds and used as a gel medium in electrophoresis; used to separate DNA and RNA molecules on the basis of their molecular weight. Agaroses may be of different quality (purity) and have different melting points.

**agarose gel electrophoresis:** The use of an electric field to move DNA or RNA through a gel matrix of agarose. See also electrophoresis.

**alignment:** Lining up two or more DNA or amino-acid sequences to maximize the number of identical bases or amino-acid residues while minimizing the number of mismatches and gaps.

**allele:** One of two, or more, alternative forms of a gene at a particular locus. If more than two alleles exist, the locus is said to exhibit multiple allelism.

**allele frequency:** The proportion of a particular allele at a locus within a population.

**allozyme:** Allozymes are a subset of isozymes. Allozymes are variants of enzymes representing different allelic alternatives of the same locus.

**alternative splicing:** Gene regulation by means of alternative splicing of exons can produce different amounts of protein or, even, different proteins for different tissues or sexes or developmental stages.

**amino acid:** One of the monomeric units that polymerize to make a protein molecule.

**aminoacyl tRNA synthetase:** Enzymes that catalyze the attachment of each amino acid to the appropriate transfer RNA molecule. A tRNA molecule carrying its amino acid is charged.

**amplicon:** The fragment of DNA that is formed as the product of DNA replication, usually referring to the product of the PCR.

**amplification:** The production of additional copies of a DNA sequence; amplification occurs in the PCR and also occurs naturally when amplified DNA is found as either

- intrachromosomal or extrachromosomal DNA within a cell.
- anaphase:** The stage of mitosis during which the two sets of chromosomes separate and are moved apart by the spindle fibers.
- anchored PCR:** A modification of the PCR that allows amplification in situations in which only one sequence is known that is suitable for a primer (rather than two). The procedure involves synthesis of cDNA with the known primer from mRNA. A poly(G) tail is added to the cDNA. The second primer is developed by synthesizing a primer with a poly(C) sequence, which allows amplification of a second DNA strand that is complementary to the cDNA. Subsequent cycles yield amplified DNA from both strands.
- aneuploidy:** A condition in which the chromosome number of an organisms is not an exact multiple of the typical haploid set for the species.
- angstrom:** Abbreviated as Å; one hundred-millionth of a centimeter, or 0.1 nm.
- anneal:** The process by which the complementary base pairs in the strands of DNA combine.
- annotation:** Linking information from the literature to databases of DNA or proteins; also identifying DNA sequences as putative genes by comparing sequences or by statistical methods or models.
- antibody:** A protein that recognizes and binds to a foreign substance introduced into an organism.
- anticodon:** The triplet of nucleotides in a transfer RNA molecule that is complementary to and base pairs with a codon in a messenger RNA.
- antigen:** Any molecule that stimulates the production of specific antibodies or binds specifically to an antibody.
- antiparallel:** The DNA strands are parallel but point in opposite directions, e.g., 5' to 3' and 3' to 5'.
- antisense mRNA:** An mRNA transcribed from a cloned gene that is complementary to the mRNA produced by the normal gene.
- apomorphic:** A character that is derived and not ancestral.
- apoptosis:** Programmed cell death, is a series of steps that cause a cell to die via "self digestion" without rupturing and releasing toxic intracellular contents into the surrounding environment.
- arbitrarily primed PCR (AP-PCR):** Does not require a particular set of primers; rather it uses a single primer chosen without regard to the sequence to be fingerprinted. By using a single primer and two cycles of low stringency PCR followed by many cycles of high-stringency PCR, discrete and reproducible products characteristic of specific genomes are produced. As originally described, the primers used are 20-bp sequencing primers.
- Archaea:** One of the two major divisions or domains of microorganisms (the other is the Eubacteria). Archaea are often found in extreme environments such as hot springs or saline water. Eukaryotes are thought to have genomes that are mosaics, consisting of genes from both the Archaea and the Eubacteria.
- arrhenogenic:** A sex-determining system in which females produce male progeny only in the blowfly *Chrysomya rufifacies* (Calliphoridae).
- arrhenotoky:** A form of parthenogenesis in which an unfertilized egg develops into a male by parthenogenesis and a fertilized egg develops into a female. Arrhenotoky is found in many species in the Hymenoptera.
- artificial selection:** The process of purposefully changing the traits of a population by determining which individuals will reproduce.
- associative learning:** The process of learning through the formation of associations between stimuli and responses.
- asymmetric PCR:** Single-stranded DNA produced by providing an excess of primer for one of the two DNA template strands. Asymmetric primer ratios are typically 50:1 to 100:1. ss DNA produced can be sequenced directly without cloning.
- ATP:** Adenosine triphosphate is the primary molecule for storing chemical energy in a cell.
- autecology:** The ecology of an individual organism or species.
- autoradiography:** A method for detecting radioactively labeled molecules through exposure of an X-ray-sensitive photographic film.

**autoregulatory control:** Regulation of the synthesis of a gene product by the product itself. In some systems, excess gene product behaves as a repressor and binds to the regulatory sequences of its own structural gene.

**autosomes:** All chromosomes except the sex chromosomes. Each diploid cell has two copies of each autosome.

## B

**B chromosomes:** B chromosomes are nonvital supernumerary chromosomes found in many plant and animal species, thought to be derived from one of the normal chromosomes. B chromosomes often are transmitted at higher rates than expected, thus exhibiting “drive.” The PSR condition (paternal sex ratio) of the parasitoid wasp *Nasonia vitripennis* is an example of a B chromosome.

**B DNA:** A helical form of DNA formed by adding water to dehydrated A-DNA. B-DNA is the form of DNA from which Watson and Crick constructed their model in 1953. It is found in very high relative humidity, and is thought to prevail in the living cell.

**BAC:** Bacterial artificial chromosome. Vectors used for cloning large DNA fragments from insects (100s of kilobases); especially useful for whole-genome sequencing.

**back mutation:** Mutations that occur to reverse a point mutation to the original condition.

**Bacteria:** (*Eubacteria*) One of the three main evolutionary domains (Archaea and Eukaryota are the others) of life.

**bacterial conjugation:** A temporary union between two bacteria, in which genetic material is exchanged; DNA from the “male” cell transfers all or part of its chromosomes into the recipient “female.”

**bacteriophage:** A virus whose host is a bacterium. See lambda for description of  $\lambda$  phage.

**baculovirus:** An insect-pathogenic virus, also known as a polydnavirus, with a circular double-stranded DNA genome and rod-shaped enveloped virion, found primarily in lepidopterans. These viruses

have been engineered for two purposes: 1) as expression vectors to produce large quantities of proteins or 2) as biological pesticides for control of pest lepidopterans.

**baculovirus expression vectors:** Vectors used to infect certain types of insect cells to produce particular proteins.

**balancer chromosomes:** Balancer chromosomes initially were developed by H. J. Muller as a method for maintaining lethal mutations in laboratory stocks without continuous selection. *Drosophila* balancer stocks contain several recessive visible mutations, one or more inversions and transpositions on a specific chromosome. These mutations have been induced to suppress crossing over. Balancing chromosomes usually have a clearly visible dominant mutation so that flies heterozygous for it can be identified easily.

**barcoding:** A method of analysis of species using short, standardized genetic marker (usually COI sequences). A problem with barcoding is that it can be uninformative because COI does not discriminate between all valid species. Another problem is that nonfunctional copies of mitochondrial DNA may have been transferred to the nucleus (**numts**), which can create confusion if these are PCR amplified by the COI primers.

**base:** Often used to refer to the purines or pyrimidines in DNA and RNA.

**base excision repair:** One of the mechanisms for repairing damage to DNA in which the damaged (mutated) nucleotide is excised and the correct nucleotide is inserted into the gap.

**base pair (bp):** Two nucleotides that are in different strands of nucleic acid and whose bases pair by hydrogen bonding. In DNA, adenine pairs with thymine (two hydrogen bonds) and guanine pairs with cytosine (three hydrogen bonds).

**BEAST:** A software program that involves Bayesian analysis of phylogenetic trees; several models are included.

**bioinformatics:** Researchers in bioinformatics develop computer software that can store, compare and analyze the very large quantities of DNA sequence data generated by the new genome technologies. See genomics, functional genomics and structural genomics. Bioinformatics tools

can sift through a mass of raw data, finding and extracting relevant information and their relationships.

**biotechnology:** The manipulation of organisms to provide desirable products for human use. It has broader meanings, as well, including all parts of an industry that creates, develops, and markets a variety of products through the molecular manipulation of organisms or using knowledge pertaining to organisms.

**bivalent:** The pair of homologous synapsed chromosomes that are found during the first meiotic division.

**BLAST:** Basic Local Alignment Search Tool.

BLAST is used to search large databases of DNA sequences and will provide sequences similar to the sequences provided by the software user.

**blastoderm:** The layer of cells in an insect embryo that surrounds an internal yolk mass. The cellular blastoderm develops from a syncytium by surrounding the cleavage nuclei with membranes derived from the enfolding of the surrounding membrane.

**blunt end:** An end of a DNA molecule, at which both strands terminate at the same nucleotide position with no extension of one of the strands.

**bootstrapping:** A statistical method based on repeated random sampling with replacement from an original sample to provide an collection of new estimates of a parameter, from which confidence limits can be calculated.

**bottleneck effect:** The loss of certain alleles by chance when a population is reduced greatly in size and only some individuals (and alleles) persist.

**broad-sense heritability:** The proportion of variance that results from genetic differences among individuals.

**buffer:** A weak acid or base that can release or take up protons and thus maintain pH.

## C

**C-banding:** Dark bands on chromosomes produced by strong alkaline treatment at high temperature followed by incubation in sodium citrate, followed by Giemsa

staining. C-bands correspond to regions of constitutive heterochromatin.

**C-value paradox:** C stands for "constant" or "characteristic" and denotes the fact that the DNA content (size) of the haploid genome is fairly constant within a species. C values vary widely among species. Size is usually measured in picograms of DNA. Because many species have large amounts (up to 90%) of their genomes that do not code for proteins, it was thought that this noncoding DNA was "junk." That view of junk DNA has changed now that the noncoding DNA has been shown to be transcribed and much of it appears to be involved in regulating gene function and development.

**capping:** The modification of the 5' end of the pre-mRNA in eukaryotes when a GTP is added to the molecule via an unusual 5'-5' triphosphate bond. Capping is necessary for the ribosome to bind with the mRNA to begin protein synthesis.

**carbohydrate:** A large class of carbon–hydrogen–oxygen compounds, including simple sugars (monosaccharides) such as glucose. Glucose is the major fuel for most organisms and is the basic building block of polysaccharides such as starch and cellulose.

**Cardinium:** A microbial symbiont in insects and mites that can cause incompatibility between strains with and without the symbiont, as well as sex-ratio anomalies. It is typically transmitted transovarially.

**cDNA:** The double-stranded DNA copy of a eukaryotic messenger RNA molecule, produced *in vitro* by enzymatic synthesis and used for producing cDNA libraries or probes for isolating genes in genomic libraries.

**cDNA library:** A collection of clones containing double-stranded DNA that is complementary to mRNAs. Such clones lack introns and regulatory regions of eukaryotic genes. Genes that are inactive will not be represented in a cDNA library, nor will noncoding regions of the genome.

**cell:** The fundamental unit of life; each multicelled organism is composed of differentiated cells; cells may be organized into organs that are relatively autonomous but cooperate in the functioning of the organism.

**cell-autonomous determination:** The establishment of a developmental pathway within a particular cell. Substances diffusing from elsewhere do not influence determination.

**cell cycle:** The sequence of events between one cell mitotic division and another in a eukaryotic cell. Mitosis (M phase) is followed by a growth ( $G_1$ ) phase, then by DNA synthesis (S phase), then by another growth ( $G_2$ ) phase, and then by another mitosis.

**cell division:** The separation of a cell into two daughter cells, ideally with each daughter being (nearly) identical. The process initiates with mitosis (division of the nucleus) and is followed by division of the cytoplasm.

**centimorgan (cM):** A unit of measurement on a genetic map of chromosomes.

**Central Dogma:** The Central Dogma was proposed by F. Crick in 1958. It states that genetic information is contained in DNA, which is transcribed into mRNA, which is translated into polypeptides. The transfer of information was proposed to be unidirectional from DNA to polypeptides; polypeptides are unable to direct synthesis of RNA, and RNA is unable to direct synthesis of DNA. The Central Dogma was modified in 1970 when RNA viruses were found to transfer information from RNA to DNA.

**centromere:** A region of a chromosome to which spindle fibers attach during mitosis and meiosis. The position of the centromere determines whether the chromosome will appear as a rod, a J, or a V during migration of the chromosome to the poles in anaphase of mitosis or meiosis. In some arthropods, the spindle fibers attach throughout the length of the chromosome and such chromosomes are called holocentric. Centromeres are usually bordered by heterochromatin containing repetitive DNA.

#### **chain-terminating method of DNA sequencing:** See dideoxy method.

**channels:** A pore in a lipid membrane through which ions or other molecules can pass. See ion channels.

**chaperones:** Protein molecules that assist with correct protein folding as the protein emerges from the cell's ribosome. Heat-shock protein 70, heat-shock protein 40, and chaperonins are examples.

**chelating agent:** A molecule capable of binding metal atoms; one example is EDTA, which binds  $Mg^{2+}$ .

**chemotaxis:** The movement of a cell or organism toward (positive) or away from (negative) a chemical substance.

**chiasmata (chiasma):** Chiasmata may occur during prophase I of meiosis and represent points where crossing over, or exchange of genetic information, between nonsister chromatids occurred. When the synapsed chromosomes begin to separate in late prophase I, they are held together by these connections between the chromatids of homologous chromosomes.

**chimeric DNA:** Recombinant DNA containing DNA from two different species.

**ChIP-chip microarrays:** This involves chromatin immunoprecipitation (ChIP) and DNA microarray chips to create high-resolution genome-wide maps of the *in vivo* association of DNA sequences with regulatory proteins and histones.

**ChIP-Seq:** Chromatin immunoprecipitation sequencing allows genome-wide mapping of protein-DNA interactions. The sequencing library is made by adding adaptors to ChIP DNAs, followed by size selection and gel purification and sequencing. The resultant data allows greater resolution than ChIP-chip microarrays and better genome coverage.

**chitin:** A water-insoluble polysaccharide that forms the exoskeletons of the Arthropoda and Crustacea.

**chitinase:** An enzyme that degrades chitin.

**chorion:** A complex structure covering the insect egg.

**chromatids:** Chromosome components that have duplicated during interphase become visible during the prophase stage of mitosis. Chromatids are held together at the centromere.

**chromatin:** The complex of DNA and proteins that form the chromosomes of eukaryotes.

**chromomere:** A region on a chromosome of densely packed chromatid fibers that produce a dark band. Chromomeres are readily visualized on polytene chromosomes.

**chromosome:** Units of the genome with many genes, consisting of histone

proteins and a very long DNA molecule; found in the nucleus of every eukaryote. Chromosome numbers, size, and shape can vary; some have centromeres and others are holocentric (fibers attach all along chromosome during cell division).

**chromosome condensation:** When a chromosome becomes densely packed in the cell before the M phase of the cell cycle.

**chromosome imprinting:** The mechanisms involved in chromosomal imprinting, or labeling of DNA, is associated with methylation of DNA in many organisms. Imprinting is a reversible, differential marking of genes or chromosomes that is determined by the sex of the parent from whom the genetic material was inherited.

**chromosome library:** A collection of clones containing the genetic information from a particular chromosome.

**chromosome puffs:** A localized swelling of a region of a polytene chromosome due to synthesis of DNA or RNA. Puffing is readily seen in polytene salivary gland chromosomes of dipteran insects.

**chromosome walking:** A molecular genetic technique that allows a series of overlapping fragments of DNA to be ordered. The technique is used to isolate a gene of interest for which no probe is available but is linked to a gene that has been identified and cloned. The marker gene is used to screen a genomic library. All fragments containing the known cloned gene are selected and sequenced, the fragments are then aligned and those cloned segments farthest from the marker gene in both directions are subcloned for the next step, and so on. The subclones are used as probes to screen the genomic library to identify new clones containing DNA with overlapping sequences. As the process is repeated, the nucleotide sequences of areas farther and farther away from the marker gene are identified, and eventually the gene of interest will be found.

**circadian clock:** Changes in biological or metabolic functions that show periodic peaks or lows of activity based on, or approximating, a 24-hour cycle.

**circadian rhythms:** Biological rhythms with periods of  $\approx$ 24 hours; rhythmicity

is endogenous and self-sustaining, continuing under constant environmental conditions for a period.

**cis:** Cis means on the same side and usually refers to a structure on the same strand of the DNA when discussing transcription of DNA.

**clade:** An evolutionary lineage derived from a single stem species. A branch of a cladogram.

**cladistic systematics:** A method of systematics that uses only shared and derived characters as a basis of constructing classifications. The rate or amount of change subsequent to splitting of phyletic lines is not considered. All taxa must arise by from a common ancestral species.

**cladogenic speciation:** Branching evolution of new species.

**cladogram:** A term used to describe either a dendrogram (tree) produced using the principle of parsimony, or a tree that depicts inferred historical relationships between organisms.

**clock:** The entire circadian system, including the central oscillator and the input and output pathways. A clock can be entrained to the environmental day-night cycle and is used to track the passage of time by an organism.

**clone:** A population of identical cells often containing identical recombinant DNA molecules. Also a group of organisms produced from one individual cell through asexual processes. The offspring are identical. The word may be used either as a noun or a verb.

**cloning vector:** A DNA molecule capable of replicating in a host organism; a gene is often inserted into the vector to construct a recombinant DNA molecule and the vector is then used to amplify (clone) the recombinant DNA.

**ClustalW and ClustalX:** Phylogenetic analysis programs that allow multiple sequence alignments.

**cluster analysis:** A method of hierarchically grouping taxa or DNA sequences on the basis of similarity or minimum distance. UPGMA is an unweighted pair group method of analysis using the arithmetic average. WPGMA is the weighted pair group method of analysis using the arithmetic average.

**coding strand:** The strand of the DNA molecule that carries the genetic information that is transcribed by RNA polymerase into mRNA.

**codominant:** Alleles whose gene products are both manifested in the heterozygote.

**codon:** A triplet of nucleotides that code for a single amino acid. Often the third base in the codon can vary because the genetic code is "degenerate," i.e., 64 codons are possible with four nucleotides but there are only 20 amino acids. As a result, some amino acids may be coded for by two or more codons.

**codon bias:** The propensity for a particular species to use one or a few codons within a codon family (recall that most amino acids are coded for by more than one codon).

**coefficient of gene differentiation:** Interpopulation diversity using allozyme data are usually measured using the *coefficient of gene differentiation* ( $G_{ST}$ ).  $G_{ST}$  is derived by estimating the average similarity within and between populations.  $G_{ST}$  is an extension of Wright's correlation ( $F_{ST}$ ) between two gametes drawn at random from each subpopulation. The coefficient of differentiation is:  $G_{ST} = (H_T - H_S)/H_T$  where  $H_S$  is the average gene diversity within populations, and  $H_T$  is the interpopulation gene diversity.

**colony:** Growth of a group of microorganisms derived from a single cell. After growth on appropriate media in a plate, the population is visible without magnification.

**colony hybridization:** The use of *in situ* hybridization to identify bacterial colonies carrying inserted DNA that is homologous with a particular sequence (the probe).

**comparative genomics:** The study of entire genomes of different species in order to understand the functions and evolution of genomes and genes. One goal is to understand which genome components are evolutionarily conserved and likely to be functional genes.

**competent cells:** Bacterial cells in a physiological state in which exogenous DNA molecules can be internalized, thereby allowing transformation.

**complementary base pairing:** Nucleotide sequences are able to base pair; A and

T are complementary; 5'-ATGC-3' is complementary to 5'-GCAT-3'. Note the DNA molecule consists of two antiparallel strands.

**complementary DNA (cDNA):** A single-stranded DNA that is complementary to a strand of mRNA. The DNA is synthesized by an enzyme called reverse transcriptase. It is a DNA copy of the mRNA.

**complex trait:** A phenotype influenced by a number of genes.

**concatemers:** The linking of multiple subunits into a tandem series or chain results in structures called concatemers.

**concerted evolution:** Maintenance of homogeneity of nucleotide sequences among members of a gene family even though the sequences change over time. Members of a gene family evolving in a nonindependent manner.

**conditional lethal:** A mutation that may be lethal only under certain environmental conditions. Some mutations are lethal at high temperatures (=temperature lethal).

**conditioned stimulus:** A stimulus that evokes a response that was previously elicited by an unconditioned stimulus.

**consensus sequence:** A hypothetical sequence consisting of the most common nucleotide or amino acid at each position in a multiple alignment of DNA or protein sequences, respectively.

**constitutive enzymes:** Enzymes that are part of the basic permanent machinery of the cell. They are formed consistently in constant amounts regardless of the metabolic state of the organism.

**constitutive heterochromatin:** Regions of the chromosome containing mostly highly repeated, noncoding DNA; commonly found near the telomeres and centromeres.

**contig:** Segments of DNA that partially overlap in their sequence are called contigs. DNA sequencing of random DNA segments results in data that can be assembled into a continuous sequence by overlapping contigs.

**convergent evolution:** The evolution of unrelated species resulting in structures with a superficial resemblance.

**conversion of transposable**

**elements:** Conversion of disabled transposable elements into active elements

- sometimes can occur through a DNA repair process.
- copy number:** The number of plasmids in a cell, or the number of genes, transposons, or repetitive elements in a genome.
- core DNA:** The DNA in the core nucleosome that is wrapped around the histone octamer. Linker DNA connects the core nucleosome to others.
- cos sites:** The cos sites are cohesive end sites or nucleotide sequences that are recognized when a phage DNA molecule is being packaged into its protein coat.
- cosmid:** Engineered vectors used to clone segments of exogenous DNA, derived by inserting cos sites from a phage into a plasmid. The resulting hybrid molecule can be packaged in the protein coat of a phage.
- CpG island:** Many genes have CpG islands in or near the promoter and methylation of the cytosines (C) will result in a repression of transcription.
- crossing over:** The reciprocal exchange of polynucleotides between homologous chromosomes during meiosis. This involves breaks at corresponding sites in the homologous chromosomes, which are rejoined to produce two recombined chromosomes; a physical exchange of DNA segments.
- cryptic species:** Species that are valid yet are not morphologically different and must be detected using other criteria (behavior, molecular data, ecology).
- cyclic AMP (cAMP):** A regulatory molecule (adenosine-3',5'-cyclic monophosphate) involved in controlling gene expression.
- cytochrome:** The complex protein respiratory enzymes occurring within plant and animal cells in the mitochondria, where they function as electron carriers in biological oxidation.
- cytokinesis:** Division of the cytoplasm at the end of mitosis or meiosis when the two new nuclei are sorted into separate daughter cells.
- cytological map:** A map of the location of genes or other DNA relative to the banding patterns of the chromosomes.
- cytoplasm:** The components of the cell, NOT including the nucleus.
- cytoplasmic incompatibility:** Reproductive incompatibility between two populations caused by factors that are present in the cytoplasm. Often associated with microorganisms such as *Wolbachia*.
- cytoplasmic inheritance:** See also maternally inherited. Sperm rarely includes much cytoplasm so most mitochondria and endosymbionts are transmitted through the cytoplasm of the ovum.
- cytoplasmic sex ratio distorters:** Cytoplasmic genes that manipulate the sex ratio of their host to promote their own spread. Such microbes (*Wolbachia*, *Cardinium*, spiroplasmas, viruses) often are transovarially (transmitted in the oocytes) and transstadially (transmitted through different life stages, such as larva to nymph, etc.) transmitted.
- cytosine:** A pyrimidine, one of the bases in DNA and RNA. It pairs with guanine, a purine.
- cytoskeleton:** Protein filaments in the cytoplasm of a cell that provide the cell shape and the ability to move. The components include actin filaments and microtubules.
- cytosol:** The fluid portion of the cytoplasm, excluding the organelles, in a cell.

## D

- Dalton:** A unit of mass very nearly equal to that of a hydrogen atom.
- ddNTP:** Dideoxy nucleotide triphosphate; used in Sanger sequencing to terminate a DNA molecule.
- degeneracy:** Refers to the genetic code and the fact that most amino acids are coded for by more than one triplet codon. The set of codons that code for a specific amino acid is called a codon family.
- degenerate codons:** Two or more codons that code for the same amino acid.
- degenerate primers:** Degenerate primers can be used for the PCR when a protein sequence is known for a gene, but the DNA sequence is not known. A pool of primers is synthesized to exploit the degenerate code.
- deletion:** The loss of a portion of the genetic material from a chromosome. The size can vary from one nucleotide to sections containing many genes.

**denatured DNA:** DNA that has been converted from double- to single-stranded form by a process such as heating.

**denaturation:** Breakdown of secondary and higher levels of structure of proteins or nucleic acids by chemical or physical means.

**dendrogram:** A branched diagram that represents the evolutionary history of a group of organisms.

**density gradient centrifugation:** Separation of molecules and particles on the basis of buoyant density, often by centrifugation in a concentrated sucrose or cesium chloride solution.

**deoxyribonuclease (DNase):** An enzyme that breaks a DNA polynucleotide by cleaving phosphodiester bonds.

**deoxyribonucleic acid (DNA):** The genetic information, the gene, the basis of inheritance in most organisms. It is now recognized that nearly all DNA in the nuclear genome has a function, coding for proteins, tRNA, rRNA, as well as regulatory RNAs involved in gene regulation and development.

**deoxyribonucleotide:** The nucleotides in DNA, consisting of deoxyribose, a base, and a phosphate group.

**determination:** The establishment of a cell fate early in development that determines the type of cell it will be.

**deuterotoky:** A form of parthenogenesis in which unfertilized eggs can develop into either males or females.

**dicentric chromosome:** A chromosome or chromatid with two centromeres. Such abnormal chromosomes result in broken fragments that are lost frequently.

**dideoxynucleotide (ddNTP):** A nucleotide that has a 3'-H on the deoxyribose sugar rather than a 3'-OH. If incorporated into a DNA molecule, it terminates the synthesis because no phosphodiester bond can be formed subsequently.

**dideoxy sequencing:** Developed by F. Sanger and A. R. Coulson in 1975, and known as the "plus and minus" or "primed synthesis" method of DNA sequencing. DNA is synthesized *in vitro* so that it is labeled and the reaction terminates specifically at the position corresponding to a specific base. After denaturation, fragments of different

lengths are separated by electrophoresis and identified. The original method used autoradiography, but fluorescent labels are now used.

**diploid:** Having two copies of each chromosome.

**direct repeats:** When some transposable elements insert into a host genome, a small segment, typically 4–12 bp, of the host DNA is duplicated at the insertion site. The duplicated repeats are in the same orientation and are called direct repeats.

**discontinuous gene:** A gene in which the genetic information is separated into two or more different "exons" by an intervening sequence (intron), which typically is noncoding. Most eukaryotic genes are discontinuous.

**discrete character:** A character that is countable (e.g., red vs. white eyes). Cf. continuous characters, which can be measured (length, width).

**distance:** A measure of the difference between two objects.

**distance estimates:** A phrase used to emphasize the fact that evolutionary history is inferred from experimental or sequence data, and distance is thus an estimate. Distance methods are used for analyzing phylogeny; a common distance method is neighbor joining.

**divergence:** Differences (mutations) that occur in DNA sequences that are derived from a common ancestor.

**divergence-time estimate:** The estimation of the time which two different taxa diverged is often done using molecular data, although the date ideally is confirmed using fossil data.

**DNA:** Deoxyribonucleic acid, the genetic molecule, consisting of double-stranded, helical deoxyribonucleotides.

**DNA microarrays:** DNA arrays work by hybridization of labeled RNA or DNA to DNA molecules attached at specific locations on a surface, allowing analysis of thousands of DNA molecules. The hybridization of a sample to an array involves a "search" by each molecule for a matching partner on the matrix, with the eventual pairings of molecules

on the surface determined by base complementarity. Original arrays involved DNA from cDNA, genomic DNA or plasmid libraries spotted on a porous membrane with the hybridized material labeled with a radioactive group. Now, glass is often used as a substrate and fluorescence for detection. New technologies allow synthesizing or depositing nucleic acids on glass slides at very high densities, which means the nucleic-acid arrays have become miniaturized, leading to increased efficiency and information content.

**DNA-binding protein:** Proteins such as histones or RNA polymerase that attach to DNA as part of their function.

**DNA-dependent RNA polymerase:** This enzyme is required for transcription (the process of RNA synthesis using a DNA template).

**DNA–DNA hybridization:** A method for determining the degree of sequence similarity between DNA strands from two different organisms by the formation of heteroduplex molecules.

**DNA fingerprinting:** See fingerprinting.

**DNA ligase:** An enzyme that repairs single-stranded discontinuities in double-stranded DNA. DNA ligases also are used in constructing recombinant DNA molecules.

**DNA methylation:** See methylation.

**DNA polymerase:** An enzyme that catalyzes the formation of DNA from dNTPs, using single-stranded DNA as a template. Three different DNA polymerases (I, II, and III) have been isolated from *E. coli*. Eukaryotes contain different DNA polymerases, found in the nucleus, cytoplasm, or mitochondria that are involved in DNA replication, repair, and recombination.

**DNA polymorphism:** Variability in the sequence or numbers of tandem repeats in a particular locus in the genome. Many polymorphisms are in the noncoding region of the genome. Used in identifying individuals by DNA fingerprinting.

**DNA probe:** Also called a gene probe or genetic probe. Short, specific (complementary to the desired DNA sequence) artificially produced segments of labeled DNA are used to combine with and detect the presence of a specific gene

or DNA sequence within the chromosome. The presence of this labeled probe usually is detected visually.

**DNA repair:** The processes that correct deleterious mutations in DNA molecules. There are many genes involved in DNA (and chromosome) repair.

**DNA replication:** The process of making a copy of a DNA molecule.

**DNA sequencing:** Determining the order of nucleotides in a DNA molecule.

**DNA topoisomerase:** An enzyme that introduces or removes turns from the double helix by transiently breaking one or both of the strands.

**DNA transcription:** See transcription.

**DNA vaccines:** When a strand of DNA that has been extracted from or derived from a pathogen is injected into tissues in the host organism, the tissues may take up the naked DNA and express some of the cell-surface proteins of the pathogen. If the host's immune system mounts an immune response to those proteins (and thus to the pathogen), the injected naked genes are referred to as DNA vaccines.

**DNase:** Deoxyribonuclease, an enzyme that degrades DNA.

**domain:** A protein domain is a region of a polypeptide that is compact and stable.

**domains of life:** The Tree of Life is considered to have three main branches or domains, the Archaea, the Eubacteria (Bacteria), and the Eukaryota.

**dominant:** An allele is dominant when it produces the same phenotype whether it is heterozygous or homozygous. The trait is expressed even if only one copy of the allele is present in the diploid genome.

**dosage compensation:** A mechanism that compensates for the number of genes carried on the X chromosome in XX and XY organisms. In mammals and some insects, one or more of the X chromosomes is inactivated. In *Drosophila* males, the X chromosome is hypertranscribed so that they produce equivalent amounts of product as the two X chromosomes of the female.

**double helix:** The base-paired structure consisting of two polynucleotides in the natural form of DNA with antiparallel strands.

**downstream:** Towards the 3' end of a DNA molecule.

**driver:** Unlabeled DNA used in DNA–DNA hybridization.

**ds DNA:** Double-stranded DNA.

## E

**ecdysone:** A steroid hormone found in insects that initiates and coordinates the molting process and the sequential expression of stage-specific genes.

**Ecdysozoa:** An evolutionary clade of organisms that includes the Arthropoda (Insecta, Crustacea, Myriapoda, and Chelicerata), the Onychophora and Tardigrada and the Nematoda, Nematomorpha, Priapulida, Kinorhyncha, and Loricifera. Members are all segmented and have appendages.

**EDTA:** Ethylene dinitrilotetra-acetic acid, a chelating agent, is able to react with metallic ions, even in minute amounts, and form a stable, inert, water-soluble complex.

**electrophoresis:** The separation of molecules in an electric field. Electrophoresis can be used to separate proteins or DNA molecules, and can use different substrates to separate different sized molecules.

**electroporation:** A process used to introduce DNA into an organism. Electroporation uses a brief electrical pulse to open “micropores” in the surface of cells suspended in water containing DNA sequences. After the DNA enters the cell via the micropores, the electrical pulse ends and the pores close. The cell then may incorporate some of the exogenous DNA into its genome.

**enantiomers:** Compounds showing mirror-image isomerism.

**endonuclease:** An enzyme which degrades nucleic acid molecules by cleaving phosphodiester bonds internally. Many types occur in bacteria; endonucleases resist invading viruses or transposable elements and also have become valuable tools for the genetic engineer. Many are available from commercial sources.

**endoplasmic reticulum (ER):** A system of sacs (cisternae) in the cytoplasm of eukaryotic cells in which the ER is continuous with the plasma membrane and the outer membrane

of the nuclear envelope. If the outer surfaces of the ER membranes are coated with ribosomes, the ER is “rough-surfaced”; otherwise it is called smooth-surfaced.

**endopolyploidy:** The occurrence, in a diploid individual, of cells containing 4-, 8-, 16-, 32-fold, etc., amounts of DNA in their nuclei. Nurse cells of ovaries are often endopolyploid.

**endosymbionts:** Microorganisms, including bacteria, rickettsia, mycoplasmas, viruses and yeasts, live within the cells of many eukaryotic organisms, including arthropods. Symbiosis often is used to mean mutualism, but includes parasitism and mutualism. Intracellular symbionts have been called endocytobionts, with no assumptions being made about whether the relationship is mutualistic or parasitic (=endocytobiosis). Symbionts may be obligatory, facultative, found in specific cells or structures or free in the gut. Some are transovarially transmitted and others are obtained from the environment (frass, eggshells, or the soil). Symbionts may provide services to their arthropod hosts or may cause cytoplasmic incompatibility, male killing, or thelytoky.

**enhancer:** Sequences of DNA that can increase transcription of neighboring genes over long distances upstream or downstream of the gene and in either possible orientation.

**enhancer trap:** A method to identify genes based on their pattern of expression. A reporter gene under the control of a weak constitutive promoter when brought in proximity to a tissue-specific ‘enhancer’ element, will be regulated by that enhancer. The reporter gene should be expressed in a tissue- and stage-specific pattern similar to that of the native gene normally controlled by the enhancer.

**environmental sex determination:** A method of sex determination in which the environment, such as temperature or host, has a significant effect on the developmental processes leading to one or the other sex.

**enzyme:** A protein catalyst that is not itself used up in a reaction. Living cells produce enzymes to catalyze specific biochemical reactions and may contain nonprotein

- components called coenzymes that are essential for catalytic activity.
- epigenetic inheritance:** Inheritance that is caused by changes, such as histone modification or DNA methylation or the effects of noncoding RNAs, that result in a different phenotype although the actual DNA sequences are unchanged. Such changes, caused by environmental influences, may alter the phenotype of the individual or their progeny.
- epistatic:** Epistasis is the nonreciprocal interaction of nonallelic genes. A gene epistatic to another affects the expression of the second gene.
- Escherichia coli*:** A bacterium that commonly inhabits the human intestine. Probably the most studied of all bacteria; its genome has been sequenced and a nonpathogenic strain is used in many genetic experiments as a host for cloning.
- essential gene:** A gene that is required for the organism to survive.
- EST:** See expressed sequence tags.
- ethidium bromide (EtBr):** A dye that binds to double-stranded DNA by intercalating between the stands. DNA stained with EtBr fluoresces a pink color under UV illumination.
- ethomics:** A large-scale approach to the study of behavior. This may involve the use of computerized recording and analysis of multiple individuals at a time.
- Eubacteria:** A large group of microorganisms that are commonly found in the environment, some of which can cause diseases in insects, plants, or other organisms. Some Eubacteria are found in extreme environments, but they are chemically and genetically distinct from Archaea, which are also found in extreme environments.
- euchromatin:** Regions of a eukaryotic chromosome that appear less condensed and stain less well with DNA-specific dyes than other segments of the chromosome during interphase. Euchromatin contains many of the protein-coding genes.
- eukaryote:** An organism with cells that contain a membrane-bound nucleus and other compartments. Eukaryotes reproduce by meiosis, cells divide by mitosis, and oxidative enzymes are within mitochondria, an organelle derived from a symbiotic microorganism.
- evolution:** The changes that take place in organisms over generations through natural selection that result in new species or variations being formed.
- exogenous DNA:** DNA from an outside source. In genetic engineering, DNA from one organism is often inserted into another by a variety of methods.
- exon:** One of the coding regions of a discontinuous gene; exons are produced after introns are spliced out of the pre-messenger RNA.
- exon shuffling:** The hypothesis that new genes are formed by stitching together new combinations of exons that encode different protein domains.
- exonuclease:** A nuclease which degrades a nucleic-acid molecule by progressive cleavage along its length, beginning at the 3' or 5' end.
- expressed sequence tags (EST):** A sequenced piece of cDNA. A full-length cDNA sequence defines a transcript, but an EST is a "tag" that indicates that the particular sequence is part of a transcribed gene.
- expression vector:** Vectors that are designed to promote the expression of gene inserts. Usually an expression vector has the regulatory sequence of a gene ligated into a plasmid that contains the gene of interest. This gene lacks its own regulatory sequence. The plasmid with this new combination (regulatory sequence + gene) is placed into a host cell such as *E. coli*, yeast, or insect cells, where the protein product is produced.
- extrachromosomal gene:** A gene not carried by the cell's nuclear chromosomes, such as mitochondrial or plasmid-borne genes. Some consider the genes of microbial symbionts to be extrachromosomal genes because they provide services to the organism.

**F**

**F pili:** The presence of an F (fertility) factor determines the sex of a bacterium. Cells with F factors (circular DNA molecules that are ≈2.5% of the length of the bacterial

chromosome) are able to function as males, by producing an F pilus. The F pilus is a hollow tube through which chromosomal DNA is transferred during bacterial conjugation or "mating."

**F statistics:** A set of coefficients that describe how genetic variation is partitioned within and among populations and individuals, such as  $F_{ST}$  and inbreeding coefficient.

**$F_1$  generation:** The first filial generation that result from a cross between two different strains.

**$F_2$  generation:** The progeny that result from crossing  $F_1$  individuals (second filial generation).

**facultative heterochromatin:** Chromosomal material that, unlike euchromatin, shows maximal condensation in nuclei during interphase. Constitutive heterochromatin is composed of repetitive DNA, is late to replicate, and is transcriptionally inactive. Portions of the chromosome that are normally euchromatic may become heterochromatic at a particular developmental stage (=facultative heterochromatin). An example of facultative heterochromatin is the inactivated X chromosomes in the diploid somatic cells of mammalian females.

**fate maps:** A technique used to analyze behavior in *Drosophila*. Using a ring X chromosome that is usually lost, individuals can be produced that are partly male and partly female. Because sex determination in *D. melanogaster* is cell-specific (not regulated by circulating sex hormones), the pattern of genetic markers can be used to construct a fate map. This map correlates precise anatomical sites on the embryonic blastoderm with abnormalities affecting behavior.

**FB transposons:** A family of transposons in *Drosophila* that are associated with chromosomal abnormalities.

**fertilization:** The union of haploid male and female gametes produces a diploid zygote, marking the start of the development of a new individual and the beginning of cell differentiation.

**fibroin:** See silk.

**fingerprinting:** DNA fingerprinting relies on the presence of simple tandem-repetitive

sequences that are present throughout the genome. The regions show length polymorphisms, but share common sequences. DNA from different individuals is cut and separated by size on a gel. A probe containing the core sequence is used to label those fragments that contain the complementary DNA sequences. The pattern on each gel is specific for a given individual, and can be used to establish parentage.

**fitness:** The relative reproductive ability of individuals that have different genes.

**flanking sequence:** A segment of DNA that precedes or follows the region of interest on the DNA molecule.

**FLP recombinase:** Yeast FLP recombinase is able to catalyze recombination in which a DNA segment that is flanked by direct repeats of FLP target sites (FRTs) can be excised from the chromosome. If two homologous chromosomes each bear an FRT site, mitotic recombination can occur in *Drosophila*, leading to the introduction of DNA into known, and specific, sites. FRT sites can be introduced into *Drosophila* chromosomes by *P*-element-mediated transformation.

**foldback DNA:** DNA that contains palindromic sequences that can form hairpin double-stranded structures when denatured DNA is allowed to renature.

**forward genetics:** Analysis of the phenotype or function leads to identification of interesting mutants, which might be used to analyze a particular process or clone the genes responsible for regulating this process. This is a classic genetic approach.

**forward mutation:** A mutation from the wild type to the mutant. A **back mutation** restores the wild-type phenotype.

**founder effect:** Changes in allele frequency due to the fact that only a small number of individuals may initiate a population so that some alleles are lost.

**frameshift mutation:** A mutation resulting from inserting or deleting a group of nucleotides that is not a multiple of three so that the polypeptide produced will probably have a new set of amino acids downstream of the frameshift.

**$F_{ST}$ :** Coancestry coefficient; a measure of the relatedness of individuals.

**functional genomics:** Study of what traits/functions are conferred on an organism by specific DNA sequences, including the coding and noncoding regions. Typically, functional genomics studies occur after the genome has been sequenced.

**fusion protein:** A hybrid protein molecule produced when a gene of interest is inserted into a vector and displaces the stop codon for a gene already present in the vector. The fusion protein may begin at the amino end with a portion of the vector protein sequence and end with the protein of interest.

## G

**G-banding:** Dark bands on chromosomes produced by Giemsa staining; G-bands occur in A-T rich regions of the chromosome.

**gamete:** A reproductive cell: the sperm and ovum, or egg.

**gap genes:** Gap-gene mutants lack large areas of the normal cuticular pattern in *D. melanogaster*. Three wild-type gap genes, *Krüppel*, *hunchback*, and *krüppel-like factor*, regionalize the embryo by delimiting domains of homeotic gene expression and effect position-specific regulation of the pair-rule genes.

**gating:** The process of shutting off a function when the value of a specific parameter attains a critical level.

**gel electrophoresis:** Separation of molecules on the basis of their net electrical charge and size in agarose or polyacrylamide gel substrates.

**gene:** A segment of DNA that codes for an RNA (tRNA, rRNA) and/or a polypeptide molecule. A protein-coding gene includes regulatory regions preceding and following the coding region, as well as the introns, if the organism is a eukaryote.

**gene amplification:** The production of multiple copies of a DNA segment in order to increase the rate of expression of a gene carried by the segment. The chorion genes of *Drosophila* are amplified in the ovary.

**gene boundaries:** Boundaries occur along the chromosomes between active and inactive chromatin. Insulators help to establish barriers by acting as a neutral barrier to the influence of neighboring elements.

**gene cloning:** Insertion of a fragment of DNA containing a gene into a cloning vector and subsequent propagation of the recombinant DNA molecule in a host organism (*E. coli*, yeast). Recently, cloning of a DNA fragment by the polymerase chain reaction has simplified the technology if primers can be developed.

**gene conversion:** A genetic process by which one sequence replaces another at an orthologous or paralogous locus, resulting in concerted evolution. May result from mismatch-repair processes.

**gene duplication:** The duplication of a DNA segment coding for a gene. Gene duplication produces two identical copies, which may retain their original function, allowing the organism to produce larger amounts of a specific protein. Alternatively, one of the gene copies may be lost by mutation and become a pseudogene. Or, a duplicated gene can evolve to perform a different task.

**gene expression:** The process by which the information carried by a gene is made available to the organism through transcription and translation.

**gene gun:** A method for propelling microscopic particles coated with DNA into cells, tissues, and organelles to produce stable or transient transformation.

**gene library:** A collection of recombinant clones derived from genomic DNA or from the cDNA transcript of an mRNA preparation. A complete genomic library is sufficiently large to have a high probability of containing every gene in the genome.

**gene ontology (GO):** Gene ontology provides a standardized vocabulary to describe cellular phenomena in terms of biological processes, molecular function, and subcellular localization. The GO term describes what the gene may do. GO analysis of nonmodel organisms assumes that genes in the well-studied model organism and the nonmodel species are evolutionarily related and conserved in function. This may not always be true and gene function may need to be confirmed by functional analyses.

**gene regulation:** The mechanisms that determine the level and timing of gene expression.

**gene replacement:** The replacement of a normal gene in an organism with one that is mutated in order to investigate gene function.

**gene silencing:** A situation in which a gene is transcriptionally silent due to its location, or due to the action of a repressor. It is an example of epigenesis. Gene silencing can occur because a gene is located near a telomere or other regions with heterochromatin. It can also occur through DNA methylation. See also RNA interference (RNAi).

**gene targeting:** A technique for inserting changes into a specific genetic locus in a desired manner.

**gene transfer:** The movement of a gene or group of genes from a donor to a recipient organism.

**genetic bottlenecks:** Genetic drift can occur if there are reductions in the population size over a number of generations so that some alleles are lost.

**genetic code:** The rules that determine which triplet of nucleotides code for which amino acid during translation. There are more than 20 different amino acids and four bases (adenine, thymine, cytosine and guanine). There are 64 potential combinations of the four bases in triplets ( $4 \times 4 \times 4$ ). A doublet code would only be able to code for 16 ( $4 \times 4$ ) amino acids. Because only 20 amino acids exist, the system is redundant, so that some amino acids are coded for by two or three different triplets (codons) (=degenerate code).

**genetic distance:** A measure of the evolutionary divergence of different populations of a species, as indicated by the number of allelic substitutions that have occurred per locus in the two populations. The most widely used measure of genetic distance is that of Nei (1972).

**genetic diversity ( $G_{ST}$ ):** Variation in populations averaged over different loci.

**genetic drift:** If populations are not infinitely large, changes in allele frequencies in populations may occur due to chance effects (not due to selection).

**genetic engineering:** The deliberate modification of genes or organisms by man. Also called gene splicing,

gene manipulation, recombinant DNA technology.

**genetic linkage:** Genes are located on the same chromosome.

**genetic map:** A map of the order of an distance between genes in a genome based on the frequency of recombination between marker genes.

**genetic marker:** An allele whose phenotype is recognized that can be used to monitor its inheritance during genetic crosses between organisms with different alleles.

**genetic sex-determination system:** The mechanism in a species by which sex is determined. In most organisms, sex is genetically, rather than environmentally, determined.

**genic-balance model of sex determination**

**in Hymenoptera:** Sex is determined by a balance between non-additive male-determining genes and additive female-determining genes scattered throughout the genome of some Hymenoptera.

Maleness genes (m) have noncumulative effects but femaleness genes (f) are cumulative. In haploid individuals  $m > f$ , which results in a male.

**genome:** The total complement of DNA in an organism. A eukaryote contains nuclear and mitochondrial genomes, as well as any symbiont genomes.

**genomic footprinting:** A technique for identifying a segment of a DNA molecule in a living cell that is bound to some protein of interest. The phosphodiester bonds in the region covered by the protein are protected from attack by endonucleases. A control sample of pure DNA and one of protein-bound DNA are subjected to endonuclease attack. The resulting fragments are electrophoresed on a gel to separate them according to size. For every bond that is susceptible to restriction, a band is found on the control gel. The gel prepared from the protein-bound DNA will lack these bands and these missing bands identify where the protein is protecting the DNA from being cut. The goal of genomic footprinting is to determine the contacts between DNA bases and specific proteins in a living

cell. DNA footprinting determines these interactions *in vitro*.

**genomic imprinting:** The process by which some genes are found to function differently when they are transmitted by the mother rather than the father, or vice versa. Mechanisms of imprinting may include methylation of the DNA or histone modification.

**genomic library:** A random collection of DNA fragments from a given species inserted into a vector (plasmids, phages, cosmids, BACs). The collection must be large enough to include all the unique nucleotide sequences of the genome.

**genomics:** The study of genome data. The complete DNA sequences of organisms such as the human, mouse, rat, zebrafish, *D. melanogaster*, *C. elegans*, and *Arabidopsis thaliana* can provide a plethora of information on entire families of genes and whole pathways of interacting proteins. See also proteomics, transcriptomics, structural genomics, and functional genomics.

**genotype:** The genetic constitution of an organism. The **phenotype** of the organism is its appearance or suite of observable characters.

**geotaxis:** The movement of an animal in response to gravity and is either positive or negative.

**germ cell:** The cell in diploid organisms that carries only one set of chromosomes and is specialized for sexual reproduction (e.g., egg or sperm).

**germ line:** The population of cells in a eukaryote that will become the reproductive cells and produce sperm or eggs.

**gigabase pairs (Gbp):** 1,000,000,000 bp.

**glycosylation:** A process in which a sugar or starch is linked to a protein molecule. Most secreted proteins and most proteins on the outer surface of the plasma membrane are glycosylated.

**GMA:** Genetically modified arthropod.

**GMO:** Genetically modified organism.

**Golgi apparatus:** An organelle that is enclosed in membranes where proteins and lipids made in the endoplasmic reticulum are modified and sorted for transport to other sites.

**green fluorescent protein (GFP):** A fluorescent protein that was originally isolated from a jelly fish. The gene encoding this protein is often inserted into cells to monitor processes in living cells.

**guanine:** A purine, one of the nucleotides in DNA and RNA.

## H

**haploid:** Cells or organisms that contain a single copy of each chromosome.

**haplotype:** A group of alleles (which may include SNPs or short tandem repeats) of different genes on the chromosomes that are transmitted together.

**Hardy–Weinberg equilibrium:** An equilibrium of genotypes achieved in populations of infinite size in which there is no migration, selection, or mutation after at least one generation of panmictic (random) mating. With two alleles, *A* and *a*, of frequency *p* and *q*, the Hardy–Weinberg equilibrium frequencies of the genotypes *AA*, *Aa*, and *aa* are *p*<sup>2</sup>, 2*pq*, and *q*<sup>2</sup>, respectively.

**helicase:** The enzyme responsible for breaking the hydrogen bonds that hold the DNA double helix together so that replication can occur.

**helix:** A spiral, staircase-like structure with a repeating pattern. DNA is a double helix consisting of two antiparallel DNA molecules.

**helper plasmid:** A plasmid that is able to supply something to a defective plasmid, thus enabling the defective plasmid to function. For example, the “wings clipped” helper plasmid supplies transposase to the *P*-element vector (which lacks transposase) so that it can integrate into the chromosome. The “wings clipped” plasmid cannot insert into the chromosome because it lacks terminal inverted repeats.

**heritability:** In the **broad sense** ( $h_B^2 = V_G/V_P$ ), the fraction of the total phenotypic variance that remains after exclusion of the variance due to environmental effects. In the **narrow sense**, the ratio of the additive genetic variance to the total phenotypic variance ( $V_A/V_P$ ).

**Hermes:** A transposable element that has been engineered as a vector for transforming

insects other than *Drosophila*. *Hermes* was discovered in the house fly, *Musca domestica*.

**heterochromatin:** The regions of the chromosome that appear relatively condensed and stain deeply with DNA-specific stains. Heterochromatin contains DNA that does not code for proteins and also contains transposable elements.

**heteroduplex DNA:** A hybrid DNA molecule formed from tracer DNA and driver DNA from different individuals or species.

**heterogametic sex:** The sex that produces gametes containing unlike sex chromosomes. Many arthropod males are XY and thus heterogametic, but lepidopteran females are the heterogametic sex (ZW). Crossing over on the sex chromosomes is often suppressed in the heterogametic sex.

**heterogeneous nuclear ribonucleoproteins (hnRNPs):** Pre-mRNAs and mRNAs are associated with a set of at least 20 proteins throughout their processing in the nucleus and transport to the cytoplasm. Some of these hnRNPs contain nuclear export signals.

**heterologous DNA:** DNA from a species other than that being examined.

**heterologous recombination:** Recombination between two DNA molecules that apparently lack regions of homology.

**heteroplasmy:** The coexistence of more than one type of mitochondrial DNA within a cell or individual.

**heterosis:** Also known as hybrid vigor.

**heterozygosity:** Having a pair of dissimilar alleles at a locus; a measure of genetic variation in a population estimated by a single locus or an average over several loci.

**heterozygous:** A diploid cell or organism that contains two different alleles of a particular gene.

**Hexapoda:** One of the names applied to the Class that includes the orders of insects.

**highly repetitive DNA:** DNA made up of short sequences, from a few to hundreds of nucleotides long, which are repeated an average of 500,000 times.

**histone gene family:** See histones.

**histones:** Basic proteins that make up nucleosomes and have a fundamental role in chromosome structure.

**hobo:** A transposable element that has been engineered as a vector for transforming insects.

**Hogness box:** A DNA sequence 19–27 bp upstream from the start of a eukaryotic structural gene to which RNA polymerase II binds. The sequence is usually 7 bp long (TATAAAA); named in honor of D. Hogness. Often called TATA box and pronounced “tah-tah.”

**holocentric:** Chromosomes that have diffuse centromeres.

**homeo domain:** See homeobox.

**homeobox:** A conserved DNA sequence ≈180 bp in size found in a number of homeotic genes involved in eukaryotic development. Homeobox genes (genes in which the homeobox is present) are responsible for fundamental aspects of embryonic development.

**homeotic:** The replacement of one serial body part by a serially homologous body part.

**homeotic gene:** Genes that determine the identification and sequence of segments during embryonic development in insects and other eukaryotes. Although most genes with a homeo domain are in the homeotic class, a few are found among the segmentation genes. Homeotic genes have been described in a variety of insects other than *Drosophila*, including *Musca*, *Aedes*, *Anopheles*, *Blattella*, and *Tribolium*.

**homeotic mutations:** Mutations in which one developmental pattern is replaced by a different, but homologous one. Homeotic mutations of *Drosophila* and other insects cause an organ to differentiated abnormally and form a homologous organ that is characteristic of an adjacent segment. Examples in *Drosophila* include *aristapedia* in which the antenna becomes leg-like, and *bithorax* in which halteres are changed into wing-like appendages.

**homoduplex DNA molecules:** A double-stranded DNA molecule in which the two strands come from the same source in DNA-DNA hybridization. Heteroduplex DNA will denature or melt into single strands at lower temperatures than homoduplex DNA from a single source; the differences between hetero- and homoduplex DNA melting points allow scientists to estimate

the degree of sequence similarity in two populations or species.

**homogametic sex:** The sex that produces gametes with only one kind of sex chromosome. The females of many insects are XX and thus homogametic. Lepidopteran males are WW and homogametic.

**homolog:** Any structure that has a close similarity to another as a result of common ancestry.

**homologous chromosomes:** Two or more chromosomes with similar DNA sequences.

**homologous genes:** Genes from different organisms that may have different sequences, but code for related gene products; the genes are similar to each other due to descent from a common ancestor.

**homology:** Homology has been defined as having a common evolutionary origin.

**homoplasy:** Phenomena that lead to similarities in character states for reasons other than inheritance from a common ancestor, including convergence, parallelism, and reversal.

**homozygous:** Diploid cells or organisms that contain two identical alleles of a particular gene.

**horizontal gene transfer:** The transfer of genetic information from one species to another. Also known as lateral transfer or horizontal transfer (HT).

**hot-start PCR:** Hot start is a method to optimize the yield of desired PCR product and to suppress nonspecific amplification. Withholding an essential component, such as DNA polymerase, until the reaction mixture has been heated to a temperature that inhibits nonspecific priming and primer extension, results in less nonspecific amplification. See also PCR.

**housekeeping genes:** Genes whose products are required by the cell for normal maintenance.

**hybrid dysgenesis:** A syndrome of genetic abnormalities that occurs when hybrids are formed between strains of *Drosophila melanogaster*, one carrying (P) and the other lacking (M) the transposable *P* element. The abnormalities induced include chromosomal damage, lethal and visible mutations, and sterility. Crosses between P males and M females is dysgenic, but the

reciprocal cross is not. Hybrid dysgenesis can occur in other insects and be caused by other transposable elements.

**hybridization probe:** A labeled nucleic-acid molecule used to identify complementary or homologous molecules through the formation of stable base pairs.

**hydrogen bond:** A hydrogen bond is a weak electrostatic attraction between an electronegative atom (such as oxygen or nitrogen) and a hydrogen atom attached to a second electronegative atom. In effect, the hydrogen atom is shared between the two electronegative atoms.

**hypertranscription:** Transcription of DNA at a rate higher than normal. For some insect species with an XY sex-determination system, the male compensates for his single X chromosome by hypertranscribing the protein-coding genes on it. He produces a nearly equal amount of gene product compared with what is produced by females with two X chromosomes.

## I

**imaginal discs:** Cells set off during embryonic development that will give rise, during the pupal stage, to adult organs.

**in silico biology:** In silico biology refers to the use of computers to perform biological studies.

**inbreeding coefficient:** The correlation of genes within individuals ( $F_{IT}$ ), or the correlation of genes within individuals within populations ( $F_{IS}$ ). Both  $F_{IS}$  and  $F_{IT}$  are measures for deviation from expected Hardy-Weinberg proportions.

**indel:** An insertion or deletion in a genome found in some, but not all, individuals. Indels are one to a few bases to several kilobases. Large indels are often due to the insertion of transposable elements.

**independent assortment:** See Law of Independent Assortment.

**inducible enzymes:** Enzymes whose rate of production is increased by the presence of certain molecules.

**initiation codon:** AUG serves as an initiation codon when it occurs at the start of a gene; it marks the site where translations should

begin. AUG also codes for methionine so most newly synthesized polypeptides will have this amino acid at the amino terminus, although it may later be removed by post-translational processing of the protein.

AUG is the only codon for methionine, so AUGs that are not initiation codons also are found in the middle of a gene.

**insertion:** mutation Alteration of a DNA sequence by inserting one or more nucleotides.

**insertion sequences:** Insertion sequences are the simplest transposable elements, carrying no genetic information except what is needed to transpose (i.e., transposase). Usually 700–2500 bp long, denoted by the prefix IS and followed by the type number.

**insertion vectors:** Vectors that have a single target site at which foreign DNA is inserted.

**in situ hybridization:** The pairing of complementary DNA and RNA strands, or the pairing of complementary DNA single strands to produce a DNA-DNA hybrid in intact chromosomes. Pairing is detected by some form of label. It can also be used to identify DNA sequences in DNAs released from lysed bacterial colonies onto nitrocellulose filters.

**isozyme:** An isomer of an enzyme. Various structurally related forms of the same enzyme having the same mechanism but differing from each other in chemical or immunological characteristics.

**insulators:** Novel sequence elements found recently in *Drosophila* that are associated with boundaries between active and inactive genes. Insulators act as a neutral barrier against both positive and negative effects of the chromosomal environment.

**interactome:** The networks of interacting proteins, which are dynamic, changing from cell to cell or at different development times. Interactomes may involve thousands of proteins and understanding them may allow scientists to understand their function.

**intercalating agent:** A chemical compound which is able to invade the space between adjacent base pairs of a double-stranded DNA molecule; an example is ethidium bromide.

**intergenic region:** The noncoding region between segments of DNA that code for genes.

**interphase:** The stage of the cell cycle when chromosomes are not visible by light microscopy. During interphase, DNA synthesis occurs.

**introgression:** The incorporation of genes of one species into the gene pool of another. If the ranges of two species overlap and fertile hybrids are produced, they will tend to backcross with the more abundant species.

**intron:** A region of eukaryotic DNA coding for RNA that is later removed during splicing of the pre-messenger RNA; it does not contribute to the final mRNA product.

**inverse PCR:** Inverse PCR allows amplification of an unknown DNA sequence that flanks a ‘core’ region with a known sequence. The basic method for inverse PCR

involves digesting template DNA with an endonuclease, circularizing the digested DNA, and amplifying the flanking DNA outside the “core” region with the primers oriented in the opposite direction of the usual orientation. Primers for inverse PCR are synthesized in the opposite orientation and are homologous to the ends of the core region so that DNA synthesis proceeds across the *uncharacterized* region of the circle rather than across the characterized core region.

**inversion:** Alteration of the sequence of a DNA molecule by removal of a segment followed by its reinsertion in the opposite orientation.

**inverted repeat:** Two identical nucleotide sequences repeated in opposite orientation in a DNA molecule, either adjacent to one another or some distance apart.

**in vitro packaging:** The production of infectious particles by enclosing naked DNA in lambda ( $\lambda$ ) phage packaging proteins and preheads.

**in vivo:** Within an intact cell or organism.

**ion channels:** The membrane passages that allow certain ions to cross the membrane.

**ionic selectivity:** The ability of ion channels to permit certain ions to cross the membrane, but not others.

**isozymes (isoenzymes):** Multiple forms of an enzyme that differ from each other in their substrate affinity, in their activity, or in their regulatory properties.

**J**

- jumping genes:** Genes that move within the genome, usually because they are associated with transposable elements.
- junk DNA:** The proportion of DNA in a genome that *apparently* has no function. Also called parasitic or selfish DNA. We are learning, however, that this DNA may contain multiple types of RNAs that are involved in gene regulation or regulation of development.

**K**

- karyotype:** The set of chromosomes of a cell arranged with regard to size, shape, and number.
- kilobase:** A kilobase (kb) of DNA = 1000 nucleotides.
- kilodalton (kDa):** A unit of mass equal to 1000 Daltons. One Dalton is nearly equal to the mass of a hydrogen atom.
- kin selection:** A theory put forth by W. D. Hamilton (1964) that states that an altruistic act is favored because it increases the inclusive fitness of the individual performing the social act. Inclusive fitness is the fitness of the individual as well as his effects on the fitness of any genetically related neighbors. The theory is that alleles change in frequency in a population due to effects on the reproduction of relatives of the individual in which the trait is expressed rather than on the reproductive success of the individual. A mutation that affects the behavior of a sterile worker bee, even though detrimental to her, could increase the fitness of the worker if her behavior increased the likelihood that a close relative would reproduce. Kin selection could explain the evolution of sociality, which appears to have developed as many as eleven times in the Hymenoptera.
- Klenow fragment:** A portion of bacterial DNA polymerase I derived by proteolytic cleavage. It lacks the 5' to 3' exonuclease activity of the intact enzyme.

**L**

- lagging strand:** The DNA strand in the double helix which is copied in a discontinuous manner during DNA replication; the short segments of DNA produced during replication are called Okazaki fragments and might be ligated together to produce the final product.
- lambda or  $\lambda$ :** A double-stranded DNA virus (bacteriophage) that can invade *E. coli*. Once inside the cell lamb can enter a **lysogenic cycle** or a **lytic cycle** of replication, which results in death of the host cell. This bacteriophage has been genetically engineered as a vector for cloning.
- lateral gene transfer:** Also known as horizontal gene transfer (HGT) or horizontal transfer (HT).
- Law of Independent Assortment:** One of Mendel's laws. Genes located on different chromosomes are randomly distributed to gametes. An individual of genotype Aa Bb will produce equal numbers of four types of gametes (AB Ab, aB, and ab), if the A/a locus is on a different chromosome from the B/b locus.
- Law of Segregation:** One of Mendel's laws. The factors of a pair of characters segregate. Separation into different gametes, and thus into different progeny, of the two members of each pair of alleles possessed by the diploid parent.
- leader sequence:** An untranslated segment of mRNA from its 5' end to the start codon.
- leading strand:** The DNA strand in the double helix that is copied in a continuous manner during DNA replication.
- lethal mutation:** Mutation of a gene to yield no product, or a defective gene product, resulting in the death of the organism because the gene product is essential to life.
- leucine zipper:** DNA-binding proteins that contain four to five leucine residues separated from each other by six amino acids. The leucines on two protein molecules interdigitate and dimerize in a specific interaction with a DNA recognition sequence. Leucine zippers are involved in regulating gene expression.
- library:** A set of cloned DNA fragments that represent the entire genome.

**ligase:** DNA ligases catalyze the formation of a phosphodiester bond between adjacent 3'-OH and 5'-P termini in single-stranded DNA. DNA ligases function in DNA repair to seal single-stranded nicks between adjacent nucleotides in a double-stranded DNA molecule.

**ligation:** Enzymatic joining together of nucleic-acid molecules through their ends.

**likelihood methods:** Likelihood methods of analyzing DNA sequence data rely on genetic models and provide a basis for statistical inference. Maximum-likelihood methods of tree construction assume the form of the tree and then choose the branch length to maximize the likelihood of the data given that tree. These likelihoods are then compared over different possible trees and the tree with the greatest likelihood is considered to be the best estimate. Best used with relatively small data sets.

**linkage:** A linkage group is a group of genes located on a single chromosome.

**linkage map:** A diagram of the order and relative distances between genes on chromosomes, based on the frequency of recombination of the linked genes in the genomes of progeny obtained from crossing parents with different genetic markers.

**linker DNA:** The DNA that links nucleosomes; the function of linker DNA is unresolved.

**lipid:** An organic molecule that is insoluble in water but dissolves in organic solvents. Phospholipids are part of the structure of biological membranes.

**locus:** The position of a gene on a chromosome. Plural: loci

**long-branch attraction:** The erroneous grouping of two or more long branches as sister groups during phylogenetic analysis due to artifacts in the analysis method used.

**long germ-band development:** A pattern of embryonic development in insects, such as *D. melanogaster*, in which the pattern of segmentation is established by the end of blastoderm.

**long intergenic noncoding RNAs**

(lincRNAs): These RNAs are at least 200 nt long and appear to regulate protein-DNA interactions, including development. Some

may interact with microRNAs and histones to affect protein-coding genes. These RNAs are involved in gene regulation and development.

**long noncoding RNAs (lncRNA):** A class of noncoding RNAs found in genomes. These RNAs are transcribed but do not code for proteins. They are being investigated for their role in insect development.

**long period interspersion genome organization:** This organization of the DNA in the genome involves long (>5600 bp) repeats alternating with very long (>12 kb) of uninterrupted stretches of unique DNA sequences. Long period interspersion is characteristic of species with small genomes. Short period interspersion involves a pattern of single copy DNA, 1000–2000 bp long, alternating with short (200–600 bp) and moderately long (1000–4000 bp) repetitive sequences, which is characteristic of the DNA in most animal species.

**lysis:** The process of disintegrating a cell, which involves rupturing the membrane, breaking up the cell wall.

**lysogenic:** During the lysogenic phase of a bacteriophage, the DNA of a virus is integrated into the chromosome of its bacterial host.

**lysosome:** An organelle within a cell that contains digestive enzymes.

**lytic:** A virus in a lytic phase undergoes intracellular multiplication and lyses the bacterial host cell, releasing progeny viruses.

## M

**major groove:** The larger of the two grooves that spiral around the surface of the double helix of the DNA molecule.

**map unit:** In linkage maps, a 1% recombination frequency is defined as a map unit or one centimorgan. It is a number proportional to the frequency of recombination between two genes.

**mariner:** A transposable element that has been engineered as a vector for transforming insects other than *Drosophila*. *mariner* elements are widely found in arthropods

and in insect-parasitic nematodes, other nematodes, flatworms, hydras, humans, mouse, rat, Chinese hamster, sheep, and cow. *mariner* has been used to transform chicken, zebrafish, and a protozoan.

**marker (DNA marker):** A DNA fragment of known size used to calibrate an electrophoretic gel.

**marker (genetic):** A trait that can be observed to occur (or not) in an organism. Marker genes include genes conferring resistance to antibiotics, expression of green fluorescent protein, eye color, etc.

**maternal-effect gene:** Genes with a maternal effect are genes in the mother that have an effect on the phenotype of her progeny; usually maternal effects are the result of depositing products or maternally derived mRNAs in the egg that are used or transcribed by the embryo.

**maternal inheritance:** Characters that are transmitted primarily by cytoplasmic genetic factors, including mitochondria, viruses and some mRNAs, that are derived solely from the maternal parent. Also known as cytoplasmic inheritance or extranuclear heredity.

**Maxam and Gilbert sequencing method:** A “chemical” method to sequence DNA developed in 1977 by A. M. Maxam and W. Gilbert. Single-stranded DNA derived from double-stranded DNA and labeled at the 5' end with  $^{32}\text{P}$  is subjected to several chemical cleavage protocols to selectively make breaks on one side of a particular base. The fragments are separated by size by electrophoresis on acrylamide gels and identified by autoradiography.

**maximum parsimony methods:** Taxonomic methods that focus on the character values observed and minimizing the number of changes in character state between species over the tree, making the assumption that there have been approximately constant rates of change. The changes at each node in the tree are inferred to be those that require the least number of changes to give each of the two character states of the immediate descendants.

**median melting temperature:** The temperature at which 50% of the DNA double helices

have denatured; the midpoint of the temperature range over which DNA is denatured.

**MEGA:** A software package for molecular evolutionary genetic analysis and sequence alignment.

**megabase pairs (Mbp):** 1,000,000 bp

**meiosis:** The sequence of events occurring during two cell divisions to convert diploid cells into haploid cells that will serve as gametes during sexual reproduction.

**meiotic drive:** Any mechanism that results in the unequal recovery of the two types of gametes produced by a heterozygote.

**melting of DNA:** Melting DNA means to denature it by heat, resulting in broken hydrogen bonds that hold the two strands together.

**membrane:** The thin sheet consisting of lipid molecules and proteins that enclose cells and form the boundaries of many organelles. Membranes typically consist of a lipid bilayer containing proteins that serve as carriers of ions or small molecules from one side of the membrane to the other.

**Mendelian genetics:** Gregor Johann Mendel founded modern genetics by publishing his studies on inheritance in garden peas in 1866. He discovered hereditary traits were transmitted from generation to generation and proposed the Principle of Segregation and the Principle of Independent Assortment. His work was not “discovered” until 1900, when other scientists concluded the same laws of inheritance.

**Mendelian population:** A group of individuals that can interbreed and thus share a common set of genes.

**messenger RNA (mRNA):** RNA molecules which code for proteins and which are translated on the ribosomes.

**metabolic pathway:** The series of enzymatic reactions in which the product of one reaction is the substrate of the next.

**metabolomics:** Metabolomics is an analytical tool that involves rapid, high-throughput characterization of the small molecule metabolites found in an organism and the products of the biochemical pathways.

**metagenomics:** The analysis of the DNA sequences of microorganisms in

communities. DNA is extracted from soil or the gut of an insect, for example, and is sequenced without culturing the microbes individually. Using bioinformatics tools to compare the sequences obtained to sequences in databases, the genomes of resident organisms can be sampled and functions of the organisms can be inferred.

**metaphase:** The stage of mitosis in which chromosomes are present at the equator of the mitotic spindle before separating and moving to the opposite poles.

**methylation:** In bacteria, enzymes (modification methylases) that bind to the DNA attach methyl groups to specific bases. This methylation pattern is unique to and protects the species from its own restriction endonucleases. Methylation also occurs in eukaryotes and may be involved in genomic imprinting. Genes that are methylated are less likely to be active.

**M13 bacteriophage:** A single-stranded bacteriophage cloning vehicle, with a closed circular DNA genome of  $\approx 6.5\text{ kb}$ . M13 produces particles that contain ss DNA that is homologous to only one of the two complementary strands of the cloned DNA and therefore is particularly useful as a template for DNA sequencing.

**M13 universal primer:** A primer derived from the M13 bacteriophage is used for sequencing reactions, and has been used to identify satellite DNA sequences in many organisms.

**$\mu\text{g}$ :** A microgram ( $\mu\text{g}$ ) is one-millionth ( $10^{-6}$ ) of a gram.

**micron:** one-thousandth ( $10^{-3}$ ) of a millimeter. The Greek  $\mu$  is used as its symbol.

**microRNAs (miRNAs):** Small noncoding RNAs (19–25 nt in length) that control gene expression by base-pairing with specific mRNAs to regulate their stability and translation.

**microsatellite DNA:** Pieces of the same small segment that are repeated many times, most commonly di- or trinucleotides such as (TC) $n$  or (GAC) $n$ .

**minor groove:** The smaller of the two grooves that spiral around the surface of the DNA double helix.

**Minos:** A transposable element that has been engineered as a vector for transforming

insects other than *Drosophila*. *Minos* has a wide host range and can transform human cell lines, making it potentially useful for mutagenesis and analysis of the human genome.

**mirtron:** Mirtrons are a type of microRNAs that are found in introns and are processed as introns before being processed into microRNAs.

**mitochondrion:** An organelle that occurs in the cytoplasm of all eukaryotes. Each mitochondrion is surrounded by a double membrane. The inner membrane is highly invaginated, with projections called cristae that are tubular or lamellar. Mitochondria are the sites of oxidative phosphorylation that results in the formation of ATP. Mitochondria contain distinctive ribosomes, tRNAs, and aminoacyl-tRNA synthetases. Mitochondria depend upon genes located within the nucleus of the cells in which they inhabit for many essential mRNAs. Proteins translated from mRNAs in the cytoplasm are imported into the mitochondrion. Mitochondria are thought to be endosymbionts derived from aerobic bacteria that associated with primitive eukaryotes and have their own circular DNA molecules. The genetic code of mitochondria differs slightly from the universal genetic code. Mitochondria are transferred primarily through the egg, and thus are maternally inherited.

**mitosis:** The sequence of events that occur during the division of a single cell into two daughter cells.

**mitotic spindle:** An array of microtubules that forms between the opposite poles of a cell during mitosis. The spindle moves the two sets of chromosomes apart.

**mobile genetic element:** See transposable element.

**model organism:** A species that is studied intensively as a representative of a large group of species. The human, the mouse, *Drosophila melanogaster*, *Arabidopsis*, yeast, and *Caenorhabditis elegans* served as model organisms at the start of whole-genome sequencing efforts.

**moderately repetitive DNA:** Nucleotide sequences that occur repeatedly in

chromosomal DNA. Repetitive DNA is moderately repetitive or highly repetitive. Highly repetitive DNA contains sequences of several nucleotides repeated millions of times. It is a component of constitutive heterochromatin. Middle-repetitive DNA consists of segments 100–500 bp long repeated 100–10,000 times each. This class also includes the genes transcribed into tRNAs and rRNAs.

**molecular biology:** A term broadly used to describe biology devoted to the molecular nature of the gene and its biochemical reactions such as transcription and translation.

**molecular clock:** The hypothesis that molecules evolve in direct proportion to time so that differences between molecules in two different species can be used to estimate the time elapsed since the two species last shared a common ancestor. The molecular clock must be calibrated, often using the fossil record.

**molecular evolution:** That subdivision of the study of evolution that studies the structure and functioning of DNA at the molecular level.

**molecular genetics:** Genetic studies that focus on the molecular nature of genes and gene expression.

**molecular phylogeny:** An analysis of the relationships of groups of organisms as reflected by the evolutionary history detected in molecules (proteins, DNA).

**molecular systematics:** The detection, description, and explanation of molecular diversity within and among species.

**monoclonal antibody:** A single antibody produced in quantity by cultured hybridoma cell lines.

**morphogen:** Molecules whose local concentration directly determines the local pattern of differentiation during development.

**mRNA:** Messenger RNA.

**mtDNA:** Mitochondrial DNA.

**Muller's ratchet:** The accumulation of deleterious mutations that can lead to extinction of a population that does not reproduce sexually.

**multigene family:** A group of genes that are related either in nucleotide sequence or in terms of function; they are often clustered together.

**multiple displacement amplification (MDA):** A procedure in which genomic DNA is amplified from purified or unpurified lysates in a few hours without a thermal cycler. MDA uses exonuclease-resistant thiophosphate-modified degenerate hexamers as primers and bacteriophage Phi29 DNA polymerase to amplify genomic DNA. It is assumed that the hexamers bind at random over the genome allowing the Phi29 DNA polymerase to synthesize DNA strands up to 10 kb in length and should enrich each amplified DNA strand up to 10,000-fold.

**multiple-locus, multiple-allele model:** A model for sex determination in Hymenoptera.

**multiplex PCR:** When more than one pair of primers is used in a PCR, multiple segments of target DNA can be amplified simultaneously and thus conserve template, save time, and minimize expense. See also PCR.

**mutagen:** A chemical or physical agent able to induce a mutation in a DNA molecule.

**mutant:** An organism expressing the effects of a mutated gene in its phenotype.

**mutation:** A change in the nucleotide sequence of a DNA molecule. Mutations can involve duplications, deletions, inversions, translocations, and substitutions.

## N

**nanogram (ng):** A nanogram = one billionth of a gram.

**nanometer (nm):** A nanometer = one billionth of a meter ( $10^{-9}$  m).

**nearly neutral theory of evolution:** See neutral theory of molecular evolution.

**negative heterosis:** The inferiority of a heterozygote over that of the homozygote with respect to one or more traits such as growth, survival, or fertility.

**neuropeptides:** Small molecules functioning within and without the nervous system of arthropod to modify behavior.

**neutral theory of molecular evolution:** A theory that the majority of the nucleotide substitutions in the course of evolution are the result of the random fixation of neutral or nearly neutral mutations, rather than the

result of positive Darwinian selection. Many protein mutations are selectively neutral and are maintained in the population by the balance between new mutations and their random extinction. Neutral mutations have a function, but they are equally effective in comparison to the ancestral alleles in the survival and reproduction of the organisms carrying them. Neutral mutations spread within populations by chance because only a relatively small number of gametes are sampled each generation and thus are transmitted to the next generation.

**next-generation sequencing or Second Generation Sequencing:**

High-throughput sequencing. Three sequencing platforms include the Roche 454 system, the ABI SOLiD system, and the Illumina system. All involve template preparation, massively parallel clonal amplification, and sequencing and alignment of short reads (<200 nt). In these systems, emulsion PCR amplification of DNA fragments is done to make a light signal strong enough for reliable base detection by cameras. This can introduce base sequence errors. Even newer sequencing platforms (Third Generation) are now developed that do not involve PCR amplification of the DNA because sequencing of single molecules can be achieved, allowing longer reads to be obtained.

**nick:** A break in a single strand of a double-stranded DNA molecule.

**nick translation:** A commonly used method of labeling DNA molecules with radioactive isotopes. DNA polymerase I is used to incorporate radiolabeled nucleotides in an *in vitro* reaction.

**nitrogenous base:** A purine or pyrimidine compound that forms part of the structure of a nucleotide.

**noncoding RNAs (ncRNAs):** Large amounts of animal genomes consist of sequences that are transcribed but do not code for proteins. The sizes and roles of these noncoding RNAs seem to be varied, ranging from regulatory functions, gene silencing, genomic imprinting, and development. ncRNAs include microRNAs, small interfering RNAs, and long noncoding RNAs.

**noncoding strand:** The polynucleotide of the DNA double helix that does not carry the genetic information, but which is the complement of the coding strand.

**nonsense mutation:** A mutation in a nucleotide sequence that changes a triplet coding for an amino acid into a termination codon so that a truncated polypeptide is produced which can alter the protein's activity.

**Northern blotting:** A technique for transferring mRNAs from an agarose gel to a nitrocellulose filter paper sheet via capillary action. The RNA segment of interest is probed with a labeled DNA fragment or gene.

**nuclear envelope:** The double membrane that surrounds the nucleus, consisting of outer and inner membranes that contain nuclear pores.

**nuclear genome:** The portion of the genome contained in the nucleus of eukaryotes on chromosomes.

**nuclear pore complex:** A large structure forming a transport channel through the nuclear envelope.

**nucleic acid:** Either DNA or RNA.

**nucleic-acid hybridization:** The bonding of two complementary DNA strands, or one DNA and one RNA strand, to identify nucleic sequences of interest. Southern blot, Northern blot, and plaque, or colony hybridization techniques are all based on nucleic-acid hybridization. All employ labeled probes to identify DNA or RNA of interest.

**nucleolus:** A nucleolus is an RNA-rich spherical body associated with a specific chromosomal segment, the nucleolus organizer. The nucleolus organizer contains ribosomal RNA genes and the nucleolus is composed of the primary products of these genes, their associated proteins, and a variety of enzymes.

**nucleoside:** A chemical comprising a purine or pyrimidine base attached to a five-carbon sugar.

**nucleosome:** The basic structure in which eukaryotic chromosomes are organized and compacted. Nucleosomes are comprised of an octamer of histone proteins with DNA coiled around them and are connected to other nucleosomes by linker DNA.

**nucleotide:** A chemical consisting a purine or pyrimidine base attached to a five-carbon sugar, to which a mono-, di-, or triphosphate is attached. A monomeric unit of DNA or RNA.

**nucleus:** The membrane-bound structure of a eukaryotic cell containing the DNA organized into chromosomes.

**null allele:** An allele that produces no functional product and therefore usually behaves as a recessive.

**numts:** Numts are nuclear copies of mitochondrial pseudogenes. These pseudogenes can be amplified by COI primers and may provide an erroneous estimate of species diversity if amplified instead of or in addition to the mitochondrial COI sequences in DNA barcoding.

## O

**odorant binding protein:** A protein that enhances the ability to smell odors in small quantities, quantities lower than those needed to activate olfactory nerves.

**Okazaki fragments:** Short fragments of DNA that are synthesized during replication of the lagging strand of the DNA molecule.

**oligo:** See oligonucleotide.

**oligonucleotide:** Short chains of single-stranded DNA or RNA nucleotides that have been synthesized by linking together a number of specific nucleotides. Used as synthetic genes or DNA probes.

**oocytes:** Cells produced by the ovaries that eventually become an ovum (egg cell) after meiosis.

**open reading frame (ORF):** A series of codons with an initiation codon at the 5' end. Often considered synonymous with "gene" but used to describe a DNA sequence that looks like a protein-coding gene, but to which no function has been assigned.

**organelle:** A structure or subcompartment of a cell that can carry out a particular function, including the mitochondria and the Golgi apparatus.

**origin of replication (ORI):** A sequence in DNA that is recognized as the position at which the replication of DNA should begin. In eukaryotes, multiple origins of

replication occur on each chromosome so that replication of entire chromosomes can occur quickly.

**orphan genes:** These are genes that are unique to a species, having no apparent homologs in related species.

**orthology:** Sequences that were separated by a speciation event.

## P

**P element:** *P* elements are transposable DNA elements first found in *Drosophila melanogaster*, where they can cause hybrid dysgenesis if *P*-containing males are crossed with M strain females lacking *P* elements. *P* elements have been engineered to serve as vectors to insert DNA into the germ line of *Drosophila* embryos.

**PAGE:** Polyacrylamide gel electrophoresis.

**pair-rule genes:** Mutated pair-rule genes result in repetitive aberrations throughout the germ band, with the removal of alternate segment-width areas. The pair-rule genes (including *runt*, *hairy*, *fushi tarazu*, *even skipped*, *paired*, *odd-paired*, *odd-skipped*, *sloppy-paired*) are transiently expressed in seven or eight stripes during cellularization of the blastoderm.

**palindrome:** A DNA sequence that reads the same in both directions taking into account the antiparallel nature of the two strands, i.e., 5'- AAAAATTTT-3' 3'- TTTTTAAAAAA-5'

**PAML:** A software package that allows phylogenetic analysis by maximum likelihood.

**paralogy:** Homology that arises via gene duplication.

**parasegment:** The visible cuticular patterns of sclerites and sutures in an insect do not represent the embryonically determined true segments. Rather, the visible "segments" are parasegments.

**paratransgenesis:** The genetic modification of a symbiont in order to genetically modify an arthropod. One example involves the modification of a gut symbiont with a gene that reduces transmission of the Chagas' disease-causing agent by the insect vector. The insect vector is treated with antibiotics

to eliminate the wild-type microbial symbiont and the genetically modified symbiont is inserted into the insect.

**parental imprinting (also genomic imprinting):**

The degree to which a gene expresses itself depends on the parent transmitting the trait to the progeny. Imprinting may result from different patterns of DNA methylation, which occur during gametogenesis in the two sexes. For such a system to maintain itself generation after generation, it would have to be reversible.

**parsimony:** Parsimony dictates that the minimal number of assumptions are made in a phylogenetic analysis.

Parsimony is a common phylogenetic analysis method.

**paternal sex ratio (PSR):** The PSR condition is only carried by males of the parasitic wasp *Nasonia vitripennis* and is transmitted via sperm to fertilized eggs. The PSR chromosome itself survives, disrupting normal sex determination by changing fertilized diploid (female) eggs into haploid PSR males. PSR is unusual in its ability to destroy the complete genome of its carrier each generation.

**pathogen:** A virus, bacterium, parasitic protozoan, or other organism that causes disease by invading the body of a host; infection does not always cause disease because infection does not always lead to injury of the host.

**PCR:** See polymerase chain reaction.

**PCR-RFLP:** A technique that combines the PCR and RFLP analysis. Genomic DNA is amplified by the PCR. Once the DNA is amplified, it is cut with restriction enzymes, electrophoresed, and visualized by ethidium bromide staining. Because the DNA was amplified by the PCR, the DNA fragments can be visualized without having to blot and probe with a labeled probe, thus making PCR-RFLP more efficient and inexpensive than traditional RFLP analysis.

**peptide bond:** The chemical bond that links adjacent amino acids into a polypeptide.

**pH:** A value used to measure the acidity of a solution. It indicates the concentration of hydrogen ions in the solution in moles per liter.

**phage (bacteriophage):** A virus that attacks bacteria. Frequently used as vectors for carrying foreign DNA into cells by genetic engineers.

**phagemid:** A phagemid is a hybrid vector molecule engineered from plasmid and M13 vectors. Phagemids provide a method for obtaining single-stranded DNA because they contain two replication origins, one a standard plasmid origin that allows production of ds DNA, and the other from M13, which allows the synthesis of ss DNA if the host cell is superinfected with a helper phage.

**phenetic systematics:** Classification based on overall similarities among living organisms. All possible characters are examined and average similarities are calculated, with all characters assumed to be of equal importance.

**phenogram:** A branching diagram that links different taxa by estimating overall similarity based on data from characters. Characters are not evaluated as to whether they are primitive or derived.

**phenomics:** The study of phenotypes with knowledge of the genotypes.

**phenotype:** The observable characteristics of an organism determined by both genotype and environment.

**pheromone-binding protein:** Two soluble proteins are found in the olfactory sensilla lymph, a pheromone-degrading esterase and a pheromone-binding protein. The pheromone-binding proteins bind species-specific pheromones and are present in very high concentrations. Volatile hydrophobic odorant molecules have to enter an aqueous compartment and traverse a hydrophilic barrier before reaching olfactory neurons. The function of the pheromone binding proteins are thought to carry the hydrophobic odorant through the sensillum lymph towards the receptor proteins located in the dendrite membranes.

**phosphodiester bond:** The chemical bond that links adjacent nucleotides in a polynucleotide.

**phosphorylation:** The combination of phosphoric acid with a compound. Many proteins in eukaryotes are phosphorylated.

- phototaxis:** The movement of a cell or organisms toward or away from light.
- phyletic speciation:** The gradual transformation of one species into another without an increase in species number at any time within the lineage. Also called vertical evolution or speciation.
- phylogenetic tree:** A graphic representation of the evolutionary history of a group of taxa or genes.
- phylogenetics:** The reconstruction of the evolutionary history of a group of organisms or genes.
- phylogeny:** The evolutionary history of a group of taxa or genes, and their ancestors.
- physical map:** A map of the order of genes on a chromosome. The gene locations are determined by DNA sequencing, producing overlapping deletions in polytene chromosomes, or electron micrographs of heteroduplex DNAs.
- picogram:** A picogram is  $10^{-12}$  gram. A pg of DNA is approximately  $0.98 \times 10^9$  base pairs.
- piggyBac:** A transposable element that has been engineered as a vector for transforming insects other than *Drosophila*.
- piRNAs:** Piwi-interacting RNAs repress expression of transposons and other repetitive genomic sequences. These small RNAs are 23–31 nucleotides long and are found in the germ line and bind to PIWI proteins that use small RNA guides to silence gene expression. In *Drosophila*, piRNAs are required for fertility in both males and females.
- plaque:** A clear spot on an opaque bacterial lawn in a petri dish. A plaque results after a single phage adsorbs to a bacterial cell, infects it, and lyses, releasing progeny phage. The progeny phages infect nearby bacteria and produce more phage until a clear area becomes visible to the naked eye. Each clear area contains many copies of a single phage and, if the phage is a vector containing exogenous DNA, it contains many copies of the foreign DNA.
- plaque hybridization:** See plaque screening.
- plaque screening:** Plaque screening is employed to identify, by nucleic-acid hybridization with labeled probes, those plaques containing specific DNA sequences.
- plasmid:** Circular, ds DNA molecules found in bacteria that are often genetically engineered and used in cloning. Plasmids are independent, stable, self-replicating, and may confer resistance to antibiotics. Often used in recombinant DNA work as vectors of exogenous DNA.
- pleiotropic:** Term used to describe a gene that affects more than one, apparently unrelated, trait.
- point mutation:** A mutation that results from changes in a single base pair in a DNA molecule.
- pole cells:** The precursors of germ cells become separated early in embryonic development in *D. melanogaster* into distinctive cells (pole cells) in the posterior of the egg.
- poly A tail:** The processing of the 3' end of the pre-mRNA molecule by the addition of as many as 200 adenine (A) nucleotides, which may determine mRNA stability.
- polyacrylamide gel:** Polyacrylamide gels result from the polymerization of acrylamide monomers into linear chains and the linking of these chains with *N,N'*-methylenebisacrylamide (bis). The concentration of acrylamide and the ratio of acrylamide to bis determine the pore size of the three-dimensional network and its sieving effect on nucleic acids and proteins of different size.
- polyacrylamide gel electrophoresis (PAGE):** Process by which molecules are separated based on the size and charge using a polyacrylamide gel and electrical current.
- polydnaviruses:** The polydnaviruses are a group of viruses with double-stranded, circular DNA genomes, found only within certain groups of parasitoid Hymenoptera. Virus particles replicate only in the wasp ovary and are secreted into the oviducts from where, during oviposition, virus is injected into host larvae. It is believed that one or more genes in the virus contribute to immunosuppression of the host, thus allowing the parasitoid eggs and larvae to survive. Some polydnaviruses seem to integrate into parasitoid

chromosomal DNA, but some are present in extrachromosomal molecules.

**polylinker:** A genetically engineered segment in a vector that allows exogenous DNA to be cloned into that region because it contains two or more unique restriction sites.

**polymer:** A chemical compound constructed from a long chain of identical or similar units.

**polymerase:** A general term for an enzyme that catalyzes the addition of subunits to a polymer. DNA polymerase synthesizes DNA and RNA polymerase produces RNA.

**polymerase chain reaction (PCR):** A method for amplifying DNA using DNA polymerases such as *Taq* DNA polymerase. PCR fundamentally involves denaturing double-stranded template DNA by heat. The ss template DNA is amplified by the DNA polymerase, using dNTPs and two primers that flank the target DNA. DNA synthesis initially results in a near doubling of the number of DNA strands defined by the ends of the two primers. Additional rounds of denaturation and synthesis occur, resulting in a nearly geometric increase in DNA product molecules (or amplicons) because, after the first few cycles, each newly synthesized molecule serves as the template for subsequent DNA amplification. Modifications of the PCR reaction have been developed for special purposes. PCR is used to clone genes, produce probes, produce ss DNA for sequencing, and carry out site-directed mutagenesis. DNA sequence differences are used to identify individuals, populations, and species.

**polymorphism:** Two or more genetically different classes in the same interbreeding population.

**polynucleotide:** A polymer consisting of nucleotide units.

**polypeptide (protein):** A chain of amino acids linked by peptide bonds; each protein is a gene product.

**polyploidy:** An increase in the number of copies of the haploid genome. Most arthropods are 2n, but species are known that are polyploid (3n, 4n, 5n, 6n) and such species are parthenogenetic because it is difficult to maintain normal meiosis

otherwise. Many arthropod species are diploid but have tissues that are polyploid, such as the salivary glands, nurse cells of the ovary, and fat body. The germ line tissues remain 2n, however.

**polyribosome (polysome):** An mRNA molecule in the process of being translated by multiple ribosomes.

**polytene chromosomes:** Chromosomes in which the chromatid has duplicated up to 1000-fold without separating. Salivary gland chromosomes in *Drosophila* and other Diptera are polytene. The discrete bands of polytene chromosomes allow a physical map of genes to be constructed using light microscopy. "Puffed" bands in the polytene salivary gland chromosomes of *D. melanogaster* indicate the genes in that region are being transcribed.

**polyteny:** See polytene chromosomes.

**position-effect variegation:** The change in the expression of a gene when it is moved to a different region of the genome. The change in expression can be stable or variegated. Variegated position effects usually involve the suppression of activity of a wild-type gene when it is placed in contact with heterochromatin because of a chromosomal mutation. Under some conditions the gene may escape suppression and the final phenotype of the organism may be variegated, with patches of normal and mutant tissues. Transgenes may be affected by their chromosomal locations.

**positive and negative selection:** A method for detecting and obtaining, from among many cells or organisms, those few with the desired genetic changes induced by genetic engineering (transgenesis) or selection. Marker genes can be inserted into the organism along with the desired transgenes; such marker genes may confer resistance to antibiotics or other chemicals and allow researchers to identify those cells/individuals that contain the newly inserted genes.

**posttranscriptional control:** Regulation of gene expression that occurs after transcription of the gene has occurred; it could occur through RNA splicing or other

modifications of RNA. It also includes control of translation by microRNAs.

**posttranscriptional processing:** Changes made to mRNAs, rRNAs, and tRNAs before they are finished products or changes to polypeptide chains after they have been synthesized, such as cleavage of specific regions to convert proenzymes to enzymes or phosphorylation.

**pre-mRNA:** The unprocessed transcript of a protein-coding gene.

**primary transcript:** The immediate product of transcription of a gene (pre-mRNA) or group of genes that will be processed to give the mature transcript(s) (mRNA).

**primase:** The RNA polymerase that synthesizes the primer needed to initiate replication of a DNA polynucleotide during DNA replication.

**primer:** A short oligonucleotide that is attached to a ss DNA molecule to provide a site at which DNA replication can begin.

**primer-dimer artifacts:** Low molecular weight DNA products produced during the PCR as artifacts when the reaction is carried out with high primer concentrations, too much DNA polymerase in early cycles, and small amounts of template DNA. The primer-dimer is made when the DNA polymerase makes a product by reading from the 3' end of one primer across to the 5' end of the other primer. This process results in a sequence being produced that is complementary to each primer and can serve as a template for additional primer binding and extension.

**Principle of Independent Assortment:** One of Mendel's laws in which different pairs of traits assort independently of each other, e.g., pairs of alleles on different chromosomes segregate independently in the formation of gametes during meiosis.

**Principle of Segregation:** Formulated by Gregor Mendel in which recessive traits, which are masked in the F<sub>1</sub> generation from a cross between two pure lines, reappear in specific ratios in the F<sub>2</sub> generation. This means that two members of a gene pair (alleles) segregate (separate) from each other during meiosis. As a result, half the gametes carry one allele and the other

half carry the other allele. The progeny are produced by random combinations of gametes from the two parents.

**prion:** Proteinaceous molecules found in the membrane of cells in the brains of vertebrates. In 1982, Stanley Prusiner discovered that mutated versions could cause a neurodegenerative disease called Bovine Spongiform Encephalopathy (BSE or "mad cow disease") in cattle and Creutzfeldt-Jakob Disease and kuru in humans. These "proteinaceous infected particles" do NOT contain DNA but are able to transmit the disease by modifying normal proteins.

**probe:** A probe is a molecule labeled with radioactive isotopes or another tag (biotin or fluorescent dye) that is used to identify or isolate a gene, gene product, or protein.

**prokaryote:** An organism whose cells lack a distinct nucleus bounded by a nuclear membrane.

**promoter:** A region of DNA crucial to the accuracy and rate of transcription initiation that usually is immediately upstream of the gene itself.

**proof reading:** A mechanism by which errors in DNA synthesis are corrected.

**prophase:** The chromosomes are condensed during this first stage of mitosis.

**protease:** An enzyme that degrades proteins.

**protein:** Polymeric compounds made of amino acids. Proteins are folded into three-dimensional shapes determined by the amino-acid sequence. The shape is unique to the protein and determines its function.

**protein phosphorylation:** The addition of a phosphate group to a protein, that can alter the activity of the protein in some manner.

**proteoglycan:** A protein that is glycosylated to a variety of polysaccharide chains.

**proteome:** The protein complement of a cell at a specific time.

**proteomics:** The science and process of analyzing and cataloging all the proteins encoded by a genome (a proteome). Currently the majority of all known and predicted proteins have no known cellular function. Determining protein function on a genome-wide scale can provide clues to the metabolism of cells and organisms. Proteomics involves

understanding the biochemistry of proteins, processes and pathways. Two-dimensional gel analyses were used in the late 1970s to identify proteins active (expressed) in different tissues at different times. Now, biological mass spectrometry is a powerful method for protein analysis, involving identification or localization of proteins and interactions of proteins.

**proteosome:** A large protein complex in the cytoplasm of eukaryotic cells that contains proteolytic enzymes. Proteosomes break down proteins that have been tagged for destruction by the addition of ubiquitin.

**pseudogene:** A nucleotide sequence that is similar to a functional gene, but without accurate information so it is not functional. Pseudogenes often lack introns because they were produced by reverse transcription. Some have mutated over time and become nonfunctional “junk.”

**PSR:** See paternal sex ratio.

**puffing:** A swelling in the giant polytene chromosomes in the salivary glands of many dipterans; puffing indicates genes are being transcribed.

**pulsed field gel electrophoresis:** A technique for separating DNA molecules by subjecting them to alternately pulsed, perpendicularly oriented electrical fields. The technique allows separation of the yeast genome into a series of intact chromosomes on a gel. Chromosomes larger than yeast chromosomes (which includes most arthropod chromosomes) are digested with a restriction enzyme before electrophoresis.

**purine:** One of the two types of nitrogenous bases (adenine and guanine) that are components of nucleotides.

**pyrimidine:** One of the two types of nitrogenous bases (cytosine and thymine) that are components of nucleotides.

**pyrosequencing:** One of the Next-Generation sequencing methods that uses single-stranded DNA templates and a sequencing primer. The sequencing machine detects the incorporation of nucleotides into the growing strand after a pyrophosphate molecule is released when a dNTP is used to extend a new DNA strand. As the pyrophosphate molecule is released, an

enzyme in the mix uses it to produce ATP and a third enzyme uses the ATP to produce light. The pyrosequencer detects and quantifies the amount of light released and correlates it to which dNTP was present. Sequencing of many DNA templates is possible in the large number of wells. However, the length of the reads produced is short, ranging from 20–200 nt in length.

## Q

**Q-banding:** Bands on chromosomes produced by quinacrine staining. The staining can only be seen under UV light and is brightest in AT-rich regions.

**qPCR:** Quantitative PCR. A method that can estimate the number of template molecules in the sample.

**QTL analysis:** Quantitative trait loci (QTL) analysis is a way to study quantitative inheritance. The loci that contribute to a quantitative trait are QTLs.

**quantitative genetics:** Analysis of the genetic influence of many genes and substantial environmental variation. It is assumed that Mendel's laws apply to complex characteristics, but that many genes, each with small effect, combine to produce observable differences among individuals in a population. Quantitative genetics determines the sum of heritable genetic influence on behavior, regardless of the complexity of genetic modes of action or the number of genes involved. It does not tell us which genes are responsible for the trait.

**quantitative trait loci (QTL):** Specific DNA sequences that are related to (located near to) known traits, that may be determined by multiple loci.

## R

**radiolabeling:** The attachment of a radioactive atom to a molecule, such as incorporation of  $^{32}\text{P}$ -dNTPs into DNA.

**RAPD-PCR:** RAPD is derived from the term Random Amplified Polymorphic DNA. PCR using single primers of arbitrary nucleotide sequence consisting of nine or 10 nucleotides with a 50–80% G + C content, and no palindromic sequences. These 10-mers can

- act as a primer in PCR and yield reproducible polymorphisms from random segments of genomic DNA.**
- reading frame:** A nucleotide sequence (a triplet) that codes for an amino acid.
- real-time PCR:** Real-time PCR is used to quantify gene expression using a fluorescence-detecting thermocycler to amplify specific sequences and measure their concentration simultaneously. See also PCR.
- recessive:** A trait or gene is recessive if it is expressed in homozygous, but not heterozygous, cells.
- reciprocal cross:** Crosses between individuals from two different strains (A, B), e.g., AxB and BxA.
- recombinant DNA molecule:** A DNA molecule created by combining DNA fragments that are not normally contiguous.
- recombinant DNA technology:** All the techniques involved in the construction, study and use of recombinant DNA molecules. Often abbreviated rDNA, which can be confused with ribosomal DNA (rDNA).
- recombination:** A physical process that can lead to the exchange of segments of two DNA molecules and that can result in progeny from a cross between two different parents with combinations of alleles not displayed by either parent.
- redundancy:** Some amino acids have more than one codon. There are 64 possible combinations of four bases arranged in a triplet codon, but only about 20 amino acids.
- regulatory gene:** A gene that codes for a protein that is involved in the regulation of the expression of other genes.
- regulatory sequence:** A DNA sequence involved in regulating the expression of a gene (a promoter or operator).
- regulatory mutation:** Mutations that affect the ability to control gene expression.
- reinforcement:** An event (reward or punishment) that follows a response and increases or decreases the likelihood that it will recur.
- repetitive DNA:** DNA sequences that are repeated a number of times in a DNA molecule or in a genome. Some repetitive DNA is associated with heterochromatin, centromeres, and telomeres. Some middle-repetitive DNA may code for ribosomal RNAs and transfer RNAs.
- replacement vectors:** Vectors that have a pair of insertion sites that span a DNA segment that can be exchanged with an exogenous DNA fragment.
- replica plating:** A technique to produce identical patterns of bacterial colonies on a series of petri plates. A plate containing colonies is inverted and its surface is pressed against a block covered with velveteen. The block can then be used to inoculate up to about eight additional petri plates. By marking the patterns of the colonies on the different plates with different selective properties, it is possible to identify which colonies differ in their responses to these agents.
- replication fork:** The region of a ds DNA molecule that is being unwound so that DNA replication can occur.
- replication origin:** The site(s) on a DNA molecule where unwinding of the double helix occurs so that replication can occur. There are multiple origins of replication (ORIs) on eukaryotic chromosomes.
- reporter gene:** A gene used to identify or locate another gene.
- repression of gene transcription:** The inhibition of transcription by the binding of a repressor protein to a specific site on the DNA molecule. A repressor protein is the product of a repressor gene.
- response to selection (R):** The difference in mean phenotypic value between the offspring of the selected parents and the mean phenotypic value of the entire parental generation before selection.
- restriction endonuclease:** An enzyme that cuts DNA only at a limited number of specific nucleotide sequences. Also called restriction enzyme.
- restriction fragment length polymorphism (RFLP):** A polymorphism in an individual, population or species defined by restriction fragments of a distinctive length. Usually caused by gain or loss of a restriction site in DNA, but could result from an insertion or

deletion of DNA between two conserved restriction sites. Differences in RFLPs usually are visualized by gel electrophoresis.

**restriction site:** A specific sequence of nucleotides in a piece of ds DNA that is recognized by a restriction enzyme and that signals its cleavage.

**restriction-site mapping:** DNA is digested with a series of different restriction endonucleases, the DNA fragments are electrophoresed, and the DNA fragments are ordered to produce a linear physical map of the locations of specific DNA sequences.

**retroelement:** DNA or RNA sequences that contain a gene for reverse transcriptase. There are different classes of retroelements, including retroviruses and retrons.

**retroposition:** The transfer of genetic information through an RNA intermediate. The genetic information carried by the DNA is transcribed into RNA, and is then reverse-transcribed into cDNA. The result is that the element is duplicated and the copy of the element is transposed.

**retrosequences:** Retrosequences or retrotranscripts are sequences derived through the reverse transcription of RNA and their subsequent integration into the genome. They lack the ability to produce reverse transcriptase.

**retrotransposon:** A type of transposable element that transposes by means of an RNA intermediate. At least 10 families of retrotransposons are known in *Drosophila*.

**retrovirus:** RNA viruses that use reverse transcriptase during their life cycle. This enzyme allows the viral genome to be transcribed into DNA. The transcribed viral DNA is integrated into the genome of the host cell where it replicates in unison with the genes of the host. The cell suffers no damage from this relationship unless the virus carries an oncogene. If so, it could be transformed into a cancer cell. Retroviruses violate the Central Dogma during their replication. The HIV virus responsible for the AIDS epidemic is a retrovirus.

**reverse genetics:** A particular gene is targeted for inactivation or expression in an unusual

environment to investigate gene function. See also forward genetics.

**reverse transcription:** DNA synthesis from an RNA template, mediated by reverse transcriptase (also known as RNA-directed DNA polymerases).

**reverse transcriptase:** An enzyme that synthesizes a DNA copy from an RNA template.

**reversions:** Reverse mutation.

**RFLP:** See restriction fragment length polymorphism.

**ribonuclease:** An enzyme that degrades RNA.

**ribosome:** A self-assembling cellular organelle comprised of proteins and RNA in which translation of mRNA occurs and polypeptides are produced. Ribosomes consist of two subunits, each composed of RNA and proteins. In eukaryotes, ribosome subunits sediment as 40S and 60S particles.

**ribosomal RNA:** Ribosomal RNA genes (rRNA genes) are found as tandem repeating units in the nucleolus organizer regions of eukaryotic chromosomes. Each unit is separated from the next by a nontranscribed spacer. Each unit contains three regions coding for the 28S, 18S, and 5.8S rRNAs. The RNA that is the structural component of ribosomes.

**ribozyme:** An RNA molecule with catalytic activity. One ribozyme is the RNA in the large subunit of the ribosome.

**ring chromosome:** An aberrant chromosome with no ends.

**RNA:** Ribonucleic acid, one of two forms of nucleic acids, with ribose as the sugar and uracil as a base.

**RNA editing:** RNA editing involves altering mRNA after transcription. This results in different proteins being produced from a single gene. The molecular mechanisms include single or multiple base insertions or deletions, as well as base substitutions. RNA editing occurs in both prokaryotes and eukaryotes.

**RNA interference (RNAi) or RNA silencing:** A cellular response initiated by the presence of double-stranded RNA molecules (including viruses or transposable elements) that results in the degradation of RNAs containing a similar nucleotide sequence.

It is a natural process, but is now being studied as a method to reduce expression of selected genes (gene silencing). When double-stranded RNA (ds RNA) is injected into a cell, the RNA is cut up into smaller chunks ( $\approx 22$  nt long) and the fragments are then degraded.

**RNA polymerase:** An enzyme capable of synthesizing an RNA complementary to a DNA template.

**RNA processing:** The modifications to RNA that occur as it reaches its mature form. This usually includes capping, splicing, and polyadenylation.

**RNA-Seq analysis:** Large-scale sequencing, using Next-Generation methods, of RNA allows measurement of transcript abundance and splicing. It provides more information than microarray analyses. A cDNA library is made from RNA, the cDNAs have adaptor sequences added, then the long samples are fragmented. Small fragments (150–300 bp) are separated by electrophoresis, isolated, purified and sequenced. The sequences are then aligned to a reference genome or used for *de novo* assembly if no genome information is available.

**RNA splicing:** The process that results in the removal of introns from the pre-messenger RNA.

**RNA surveillance:** A system in eukaryotic cells to degrade aberrant mRNAs.

**RNA transcript:** An RNA copy complementary to a gene, also known as pre-messenger RNA.

**rough endoplasmic reticulum:** The region of the endoplasmic reticulum that contains ribosomes and is important in synthesizing proteins.

**RT-PCR:** Reverse transcription-PCR in which RNA is first converted to cDNA using reverse transcriptase and the cDNA is amplified by the PCR. It can detect and quantify RNA.

## S

**S phase:** The portion of interphase in the cell cycle in which DNA replication occurs. The S phase occurs between the G<sub>1</sub> and G<sub>2</sub> phases of the interphase. Mitosis occurs after the G<sub>2</sub> phase.

**S1 nuclease:** An enzyme that specifically degrades single-stranded DNAs or splits short ss segments in DNA but does not attack any ds molecules. Used to convert sticky ends of duplex DNA to form blunt ends or to trim off ss ends after conversion of ss cDNA to the ds form.

**satellite DNA:** Highly repeated DNA sequences with such a uniform nucleotide composition that, upon fractionalization of the genomic DNA and separation by density gradient centrifugation, form one or more bands that are clearly different from the main band of DNA and from the smear created by other fragments of a more heterogeneous composition. The base composition of satellite DNA differs from that of the majority of DNA in a eukaryotic species; it is either A + T rich or G + C rich and is usually highly repetitive in sequence.

**secondary structure:** Proteins fold into alpha helices and beta sheets.

**secondary transposition:** Movement of an transposable element after its initial insertion into the chromosome. Secondary transposition can be induced with P elements in *Drosophila* if an appropriate transposase is added.

**segment-polarity genes:** Segment-polarity genes appear to determine a linear sequence of repeated positional values within each segment. Segment-polarity mutants have repetitive deletions of pattern, but the deletions occur within each segment and are followed by a partial mirror-image duplication of the part that remains. Segment-polarity genes (including *engrailed*, *naked*, *patched*, *wingless*, *gooseberry*, *patched hedgehog*, *porcupine*, *armadillo*, *fused*) are required either continuously or over extensive periods to maintain the segmental pattern. Most or all are required to maintain patterns in the imaginal tissues.

**segmentation genes:** Genes, including the gap, pair-rule and segment-polarity classes of genes that determine the number and polarity of the body segments during embryonic development in arthropods.

**selectable marker:** A gene that allows identification of specific cells with a desirable

new genotype. Many vectors used for genetic engineering carry antibiotic resistance genes, or other genes, that allow identification of cells containing exogenous DNA.

**selection differential:** In artificial selection, the difference in mean phenotypic value between individuals selected as parents of the following generation and the whole population.

**selfish DNA:** DNA that may not provide any advantage to its carrier or host but ensures its own survival. Transposable elements are considered to be selfish DNA, although they may sometimes assume a useful role in the evolution of genomes.

**semi-conservative replication:** DNA replication in which each daughter double helix consists of one strand from the parent and one newly synthesized and complementary strand. Complementary base pairing is essential to this process and ensures accurate replication of DNA.

**sensory transduction:** Sensory cells transform and amplify the energy provided by a stimulus into an electrical signal. Sensory transduction is probably due to a change in the ionic permeability of the sensory cell membrane, which causes a depolarization of the membrane.

**sericin:** See silk.

**sequencing:** The process used to obtain the sequential arrangement of nucleotides in a DNA molecule.

**sex chromosome:** A chromosome that is involved in sex determination, while autosomes are not.

**short germ-band development:** A pattern of development found in some insects in which all or most of the metamerized pattern is completed after the blastoderm stage by the sequential addition of segments during elongation of the caudal region of the embryo.

**short period interspersion pattern of genome organization:** This form of genome organization has single-copy DNA, 1000–2000 bp long, alternating with short (200–600 bp) and moderately long (1000–4000 bp) repetitive sequences. This pattern is found in *Musca domestica*, the *Lucilia cuprina*, and *Antheraea pernyi*.

**short tandem repeats:** Also called microsatellites, short tandem repeats are 2–6 bp in length and are tandemly repeated in the genome, in some cases thousands of times. Many are polymorphic in a population so they can be used as genetic markers.

**shotgun cloning:** Cloning of genes from random fragments of DNA from an organism.

**shotgun libraries:** Genomic libraries in which a random collection of a sufficiently large sample of cloned fragments of the DNA are present so that all the genes are represented.

**shotgun method:** A method for introducing exogenous DNA into cells in which tiny bullets made of tungsten or other metal are coated with DNA and shot into the cell.

**silent mutation:** Changes in DNA that do not influence the expression or function of a gene or gene product.

**silk:** The cocoon filament spun by the fifth instar larva of *Bombyx mori* and other silk moths; silk also is produced by other insects, spiders, and spider mites. Each cocoon filament of *B. mori* contains two cylinders of **fibroin**, each surrounded by three layers of sericin. Fibroin is secreted by the cells of the posterior portion of the silk gland. The fibroin gene is present in only one copy per haploid genome, but these silk-gland cells undergo 18 to 19 cycles of endomitotic DNA replication before they begin transcribing fibroin mRNAs. The **sericin** proteins are named because they contain abundant serines (>30% of the total amino acids). Sericins are secreted by the cells from the middle region of the silk gland. Silk is very strong and could be commercially useful for many purposes if it could be produced synthetically.

**similarity:** A measure of the resemblance between two objects, usually on a scale of 0 to 1.

**single-locus, multiple-allele model:** A model for a sex determination mechanism in the Hymenoptera.

**single-stranded binding proteins:** One of the proteins that attaches to ss DNA in the replication fork to prevent reannealing of the DNA during DNA replication.

**snRNA:** Small nuclear RNAs are molecules of  $\approx 200$  nt that are involved in RNA splicing.

**site-directed mutagenesis:** Mutagenesis to produce a predetermined change at a specific site in a DNA molecule.

**slot blot:** A hybridization technique that allows multiple samples of DNA to be applied to nitrocellulose filters in specific sites (slots) using a vacuum.

**small interfering RNAs (siRNAs):** Small interfering RNAs are 19–29 nt in length and are involved in cleavage of mRNAs. siRNAs base-pair with identical sequences in other RNAs and result in the inactivation or degradation of the target RNA.

**small nuclear RNAs (snRNAs):** Small nuclear RNAs.

**small nucleolar RNAs (snoRNAs):** Small nucleolar RNAs are involved in modification of rRNAs.

**smooth endoplasmic reticulum:** The region of the endoplasmic reticulum not associated with ribosomes; this region is involved in lipid synthesis.

**SNP:** Single nucleotide polymorphism. Pronounced “snip,” it is a single base-pair change found in some individuals at a specific site in the genome. Individuals may have a SNP about once every 1000 bp, so SNPs can be used as DNA markers in population genetics and population ecology studies. SNPs also may be used in genetic selection projects.

**somatic cells:** All of the eukaryotic cells in the body except the germ-line cells and the gametes they produce.

**Southern blotting:** A technique developed by E. M. Southern for transferring DNA fragments isolated electrophoretically in an agarose gel to a nitrocellulose filter paper sheet by capillary action. The DNA fragment of interest is then probed with a labeled nucleic-acid probe that is complementary to the fragment of interest. The position on the filter is determined by autoradiography if the probe was radiolabeled. Southern blots allowed molecular biologists to identify specific DNA fragments for the first time. The related techniques for RNA and proteins have been dubbed “Northern” and

“Western” blots, respectively. There is no “Eastern” blot.

**speciation genes:** The genes that contribute to the evolution of reproductive isolation.

**species:** There are multiple definitions of “species,” with the biological species one of the most common. Ultimately, a species is defined by the descriptor and is based on judgments that are based on a variety of criteria. Some taxonomists are considered “splitters” and others are “lumpers.” It can be difficult to identify a “species” if the process is recent.

**specific activity:** The ratio of radioactive to non-radioactive molecules of the same kind. Probes with a high specific activity can produce a more intense signal than a probe with a low specific activity.

**spliceosome:** The RNA and protein particles in the nucleus that remove introns from pre-messenger RNA molecules.

**ss DNA:** Single-stranded DNA.

**stable transformation:** Transformation that alters the germ line of an organism so that the progeny transmit the trait of interest through subsequent generations.

**start codon:** The mRNA codon, usually AUG, at which synthesis of a polypeptide begins.

**sterile insect release method (SIRM):** A genetic control method used to suppress or eradicate pest insects. Large numbers of mass-produced males are given nonlethal, but sterilizing, doses of radiation or chemical mutagens and then released in nature. The natural populations are so overwhelmed by the excess sterile males that they almost always fertilize females. The resultant matings produce unviable progeny and a new generation is not produced. Used to eradicate the screwworm and Medflies from the United States. Sometimes called the sterile male technique or the sterile insect release method.

**sterile male technique:** See sterile insect release method (SIRM). Sometimes also called sterile insect technique (SIT).

**sticky end:** Single-stranded ends of DNA fragments produced by restriction enzymes; sticky ends are able to reanneal or to anneal to a vector with appropriate sticky ends by hydrogen bonding.

**stop codon:** One of the three mRNA codons (UAG, UAA, and UGA) that prevents further polypeptide synthesis.

**stress proteins:** Also called heat-shock proteins. Proteins (chaperones) made when the cells are stressed by environmental conditions (chemicals, pathogens, heat) that preserve critical proteins.

**stringency:** Stringency, as used in hybridization reactions such as the Southern blot, refers to the conditions that can be altered, such as temperature, to influence the ease with which a probe hybridizes to template nucleic acids.

**structural gene:** A gene that codes for protein (enzyme or component of the cell structure) other than a regulatory gene.

**structural genomics:** The study of the three-dimensional protein structure determined using modeling and experimental analyses.

**subclones:** A DNA fragment that has been cloned into one vector may be moved, or subcloned, into a second type of vector in order to perform a different procedure.

**supercoiled:** The coiling of a covalently closed circular duplex DNA molecule upon itself so that it crosses its own axis. A supercoil is also called a superhelix. The B form of DNA is a right-handed double helix. If the DNA duplex is wound in the same direction as that of the turns of the double helix, it is positively supercoiled. Twisting of the DNA molecule in a direction opposite to the turns of the strands in the double helix is called negative supercoiling.

**symbiont:** An organism living in or on another organism of a different species. The effects can range from neutrality to parasitism to mutualism. Symbionts can be obligatory or facultative. Some may reside in specialized cells or organs, while others reside in the gut or externally. Some may be transmitted transovarially and others can be obtained from the environment from frass or egg shells or the soil. Symbionts can contribute to the biology of their hosts in multiple ways.

**sympatry:** Living in the same geographic location. Sympatric species have overlapping or coinciding distributions and often have well-developed reproductive isolation mechanisms.

**synapsis:** The pairing of homologous chromosomes during the zygotene stage of meiosis.

**syncytium:** A mass of protoplasm containing many nuclei that are not separated by cell membranes.

**synecology:** The study of relationships among communities of organisms and their environment.

**syngamy:** The fusion of sperm and egg to form a zygote.

**synonymous DNA changes:** Changes in DNA sequence that do not change the amino acid coded for.

**synteny:** Synteny refers to the fact that many genes remain grouped together in the same relative positions in the genome across taxa.

**synthetic biology:** The development of artificial parts of life, including the synthesis of chromosomes or entire genomes. Goals include both the creation of artificial life or the assembly of interchangeable parts into systems that function in novel ways. One goal is to understand how organisms develop and function. Another is to develop organisms that can produce novel products of value to humans such as drugs and fuel.

**systematics:** The study of classification, based on evolutionary change.

## T

**TALEN:** TALENs are DNA-binding protein sequences known as transcription activator-like effectors (TALEs) that have been attached to a nuclease (*Fok I*), and subsequently labeled TALENs. They are engineered to be gene-editing modules that appear more effective and easier to use than zinc-finger nucleases (ZFNs). The TALEs were identified in plant pathogens (*Xanthomonas*) and different TALENs recognize different nucleotides in the genome. The nuclease cleaves a desired sequence in the genome, creating a double-stranded break. The DNA is then repaired using the natural ability of cells and, if homologous DNA containing a desired change is introduced into the cell

with the TALEN, the repair yields a genome modification at that specific location. Such genome modifications have occurred in a variety of organisms, including *Drosophila*.

**tandem repeat:** Direct repeats in DNA codons adjacent to each other.

**Taq DNA polymerase:** A DNA polymerase that isolated from the bacterium *Thermus aquaticus* that is tolerant of high temperatures. Used in the polymerase chain reaction.

**TaqMan PCR:** A realtime type of PCR which uses an oligonucleotide that anneals to an internal sequence within the amplified DNA fragment. This oligo, usually 20–24 bases long, is labeled with a fluorescent group at its 5' end and a quenching group at its 3' end. When both the fluorescent and quenching groups are in close proximity on the intact probe, any emission from the reporter dye is absorbed by the quenching dye and fluorescence emission is low. As the reaction progresses and the amount of target DNA increases, progressively greater quantities of oligo probe hybridize to the denatured target DNA. During the extension phase of the PCR, the fluorophore is cleaved from the probe by the 5' to 3' exonuclease activity of the polymerase. Because the fluorophore now is no longer close to the quencher, it begins to fluoresce. The intensity of the fluorescence is directly proportional to the amount of target DNA synthesized and allows the researcher to quantify the reaction in "real time" without running the product on a gel.

**targeted gene replacement:** The ability to replace or modify genes in their normal chromosomal locations has not been possible with *Drosophila* until recently. The cut-and-paste model of *P* element transposition provided a model for inserting a gene into the double-stranded gap left behind by a *P* element. The gap can be repaired, using a template provided by an extrachromosomal element introduced by the investigator. Other methods show promise, such as zinc finger nucleases and TALENS, for precise insertion of DNA into the chromosome.

**targeted gene transfer:** See targeted gene replacement.

**targeted mutagenesis:** The ability to replace or modify DNA sequences in their normal chromosomal location.

**TATA box:** See Hogness box.

**taxa:** The general term for taxonomic groups whatever their rank. taxon = singular.

**taxonomy:** The principles and procedures according to which species are named and assigned to taxonomic groups.

**tDNA-PCR:** Universal primers for transfer RNA can be used to generate tDNA by the PCR. The resulting fragments are visualized by gel electrophoresis and produce characteristic fingerprints for different species.

**telomerase:** An enzyme that adds specific nucleotides to the tips of chromosomes to form stable ends (telomeres).

**telomere:** Telomeres are the physical ends of eukaryotic chromosomes. They protect the ends of chromosomes and confer stability so that genetic information is not lost after cell division (and replication of chromosomes). Telomeres consist of simple DNA repeats and the non-histone proteins that bind specifically to those sequences. In *Drosophila melanogaster*, the telomeres are derived from retrotransposons.

**telomere terminal transferase:** See Telomerase.

**template:** A macromolecular mold for synthesis of another macromolecule. For example, a single strand of DNA serves as the template for a complementary strand of DNA or mRNA, which are assembled based on complementary base pairing.

**10-nm chromatin fiber:** The least condensed form of chromatin; having a beads on a string appearance and consisting of nucleosomes connected by linker DNA.

**termination codon (stop codon):** One of the three codons in the standard genetic code that indicates where translation of a mRNA should stop: 5'-UAA-3', 5'-UAG-3', or 5'-UGA-3'.

**thelygenic:** When females produce only female progeny, as in the blowfly *Chrysomya rufifacies*.

**thelytoky:** Parthenogenesis in which no functional males are produced; unmated females produce female progeny only, or rarely, a few males. Thelytoky may be

caused by microbial symbionts such as *Wolbachia* or *Cardinium*.

**tiling microarrays:** These are microarrays that evaluate coding and noncoding sequences in an entire genome. Tiling microarrays are used to map binding sites for DNA binding proteins, map methylation patterns, map coding and noncoding RNA transcripts, and identify differences in sequences under different developmental or environmental conditions. Labeled target molecules are hybridized to unlabeled probes fixed to a solid surface. The arrays differ from microarrays in that they are supposed to cover the entire genome or areas of a genome. Tiling microarrays allow previously unidentified genes to be studied.

### Third Generation Sequencing

**Platforms:** Sequencing from a single DNA molecule involves sequencing by synthesis without a prior amplification step. These methods offer the possibility of high throughput, faster turn around time, longer read lengths to enhance *de novo* assembly, a requirement for small amounts of starting material, and lower cost. Several different approaches are being developed.

**30-nm fiber:** Condensation of DNA in eukaryotic chromosomes involves formation of 30-nm fibers from supercoils of six nucleosomes per turn. The 30-nm fiber has H1 histones binding to the linker DNA and to DNA bound to the histones of the nucleosome.

**tissue:** A specific set of cells that have a distinctive function.

**T<sub>m</sub>:** The interpolated temperature along a DNA melting curve at which 50% of the duplex DNA formed in a DNA–DNA hybridization is double stranded. The difference in T<sub>m</sub> between homoduplex and heteroduplex curves is called T<sub>m</sub>.

**totipotent:** The ability of a cell to develop into any cell type. Cells may become differentiated and unable to develop into all types of cells during development.

**tracer DNA:** In DNA–DNA hybridization, single-stranded single-copy DNA from one species is radioactively labeled (tracer DNA) and hybridized with unlabeled DNA (driver DNA) from the same species or from a different

species. DNA–DNA hybridization is used to determine the degree of sequence similarity between the DNAs.

**trailer segment:** A nontranslated sequence at the 3' end of mRNA following the termination signal, exclusive of the poly(A) tail.

**trans:** Trans elements are on the opposite strand of DNA when discussing transcription factors.

**transcript:** An RNA copy of a gene; in eukaryotes the initial transcript is pre-messenger RNA and it must be processed to become mRNA.

**transcription:** The process of producing an RNA copy of a gene.

**transcriptional activator proteins:** Elements that stimulate transcription by binding with particular sites in the DNA.

**transcriptome:** The transcriptome is the profile of the genes that are expressed or transcribed from genomic DNA within a cell or tissue, with the goal of understanding cell phenotype and function. The transcriptome is dynamic and changes rapidly in response to stress or during normal cell processes such as DNA replication and cell division.

Transcriptomes can also provide information on which genes are active during different developmental stages or sexes.

**transfection:** Infection of bacteria with a viral nucleic acid that lacks a protein coat.

**transinfection:** The insertion of a strain of endosymbiont into a new host. One example is the insertion of the “popcorn” strain of *Wolbachia* into the mosquito *Aedes aegypti*. This results in a reduction in lifespan of adults and reduces the ability of the mosquito to transmit the dengue-causing virus.

**transfer RNA (tRNA):** A family of small RNA molecules (usually >50 types per cell) that serve as adapters for bringing amino acids to the site of protein synthesis on the ribosome.

**transformant:** An individual organism produced by introducing exogenous DNA.

**transformation:** The process of changing the genetic makeup of an organism by introducing exogenous DNA.

Transformation may be transient or stable (transferred to succeeding generations.)

**transgene:** The DNA that is inserted into the nuclear genome of a cell or organism by recombinant DNA methods.

**transgene suppression (or silencing):** A variety of organisms, including insects, plants and mammals, can inactivate multiple copies of inserted genes that overexpress proteins or are abnormally transcribed. Transgene silencing may be induced by methylation of the DNA or by posttranscriptional and transcriptional processes.

**transgenic organism:** An organism whose nuclear genome contains genetic material originally derived from another organism (not its parents) or from a different species. The transgene(s) can be transmitted to subsequent generations (stable transformation because the transgene is in the germ-line chromosomes) or can be lost subsequently (unstable or transient transformation because the transgene is located only in the nuclei of somatic tissues and is not transmitted to progeny).

**transient transformation:** Transient transformation involves changing the genetic makeup by introducing exogenous DNA. If the genetic information is not incorporated into the germ line, transformation is unstable, or temporary.

**transitions:** Transitions are point mutations that involve changes between A and G (purines) or T and C (pyrimidines).

**translation:** The process by which the amino-acid sequence in a polypeptide is determined by the nucleotide sequence of a messenger RNA molecule on the ribosome.

**translational regulation:** Gene regulation by controlling translation. Translation of mRNA can be tied to the presence of a specific molecular signal; the longevity of an mRNA molecule can be regulated; or overall protein synthesis can be regulated.

**translocation:** A type of mutation in which a section of a chromosome breaks off and moves to a new position in that or a different chromosome.

**transovarial transmission:** Transmitted to the next generation through the egg.

**transposable element (transposon):** An element that can move from one site to another in the genome. Transposable elements (TEs) have been divided into two main classes, those that transpose with an RNA intermediate, and those that transpose as DNA. A rolling circle group was discovered recently (Helitrons), as well. Many families of each type are known.

**transposase:** An enzyme that catalyzes transposition of a TE from one site to another in a DNA molecule.

**transposition:** The movement of genetic material from one chromosomal location to another.

**transposon:** A transposable element carrying several genes, including at least one coding for a transposase enzyme. Many elements are flanked by inverted repeats. *Drosophila melanogaster* contains multiple copies of different kinds of TEs.

**transposon tagging:** A method of cloning genes from *Drosophila* after they have been "tagged" by having the *P* element insert into them to cause changes in the phenotype.

**transversions:** Transversions are point mutations that involve changes between a purine and a pyrimidine.

**Tree of Life:** A multibranched representation of the relationships between all living organisms.

**triplex DNA:** In triplex DNA, the usual A-T and C-G base pairs of duplex DNA are present and, in addition, a pyrimidine strand is bound in the major groove of the helix. DNA sequences that potentially can form triplex DNA structures appear to be common, are dispersed at multiple sites throughout the genome, and comprise up to 1% of the total genome.

**tubulin:** The protein from which microtubules are composed.

## U

**ubiquitin:** A protein that is present in cells of both prokaryotes and eukaryotes and is highly conserved. Ubiquitin contains 76 amino acids and plays a role in proteolysis in the proteosome. Ubiquitin-conjugating

enzymes add ubiquitins to proteins carrying degradation signals. The ubiquitin is recognized by proteosomes that then cut the proteins into fragments.

**unique genes:** Genes present in only one copy per haploid genome, which includes many of the structural (protein-encoding) genes of eukaryotes. Genes that have no apparent homologs in related organisms may be called "orphan genes" and orphans may provide unique traits to the organism.

**unrooted tree:** A phylogenetic tree in which the location of the most recent common ancestor of the taxa is unknown.

**UPGMA:** The use of distance measurements to group taxonomic units into phenetic clusters by the unweighted pair-group method using an arithmetic average.

**upstream:** Towards the 5' end of a DNA molecule.

**uracil:** A pyrimidine that is one of the nitrogenous bases found in RNA.

## V

**vector:** A DNA molecule capable of autonomous replication in a cell and that contains restriction enzyme cleavage sites for the insertion of exogenous DNA and, often, selectable markers.

**vertical gene transfer:** Transfer of a gene from parents to offspring. Cf. horizontal gene transfer, which involves transfer of a gene from organism to organism, perhaps via transposable elements, viruses, or by insertion of genes obtained by feeding.

**virus:** A noncellular particle that can reproduce only inside living cells, consisting only of a genetic material (either DNA or RNA) and a protein coat. Viruses are "alive" because they can reproduce, but they have no other traits of living organisms.

**vitellogenin:** The major yolk proteins in insects are derived from vitellogenins, which are produced by the fat body and secreted for uptake by maturing oocytes.

## W

**Western blots:** Proteins are separated electrophoretically and a specific protein is

identified with a labeled antibody raised against the protein in question.

**whole-genome shotgun sequencing:** When the whole genome is broken into partially overlapping pieces, cloned and sequenced and the genome sequence is assembled using bioinformatics tools. This is the most common method of sequencing insect genomes, now that it has been shown to be rapid and efficient.

**wild type:** The normal form of an organism--in contrast to that of mutant individuals.

**wobble hypothesis:** A hypothesis to explain how one tRNA may recognize two different codons on the mRNA. Anticodons are triplets with the first two positions pairing according to base-pairing rules. The third position "wobbles" and can recognize any of a variety of bases in different codons so that it can bind to either of two or more codons.

**Wolbachia:** An endosymbiont found in many arthropods (up to 76%), as well as nematodes. There are many strains of *Wolbachia*. *Wolbachia* are transovarially transmitted and may cause a variety of phenotypes, including modification of sex ratio, thelytoky, male killing, cytoplasmic incompatibility, and other effects. At least one strain ("popcorn"), derived from *Drosophila melanogaster*, can reduce the longevity of adult mosquitoes and reduce transmission of viruses and other disease-causing agents if transinfected. Transinfection (insertion of a novel strain of symbiont) of insects is being studied as a method of achieving the control of insect-vectored diseases such as dengue and malaria. *Wolbachia* may be horizontally transferred.

## X

**X-chromosome:** A sex chromosome that usually is present in two copies in females (XX) and in one copy (unpaired) in males (XO or XY).

**X-gal:** A lactose analogue (5-bromo-4-chloro-3-indolyl- $\beta$ -D-galactopyranoside). X-gal is cleaved by  $\beta$ -galactosidase into a product that is bright blue. If exogenous DNA has inserted into and disrupted the

$\beta$ -galactosidase gene, lambda plaques will appear white or colorless. Plaques without recombinant vectors will be blue.

## Y

**Y chromosome:** A sex chromosome that is characteristic of males in species in which the male typically has two dissimilar sex chromosomes (XY). The Y chromosome often has relatively few genes, but may have genes important for male fertility.

## Z

**Z chromosome:** One of the sex chromosomes found in heterogametic ZW female insects, such as Lepidoptera.

**Z-DNA:** A structural form of DNA in which the two strands are wound into a left-handed helix rather than a right-handed form.

**zinc-finger engineering:** The use of ZFNs to modify the genome.

**zinc-finger nucleases (ZFNs):** Zinc-finger nucleases are synthetic proteins that contain an engineered zinc-finger DNA-binding domain fused to the cleavage domain of the *Fok I* restriction endonuclease. ZFNs can be used to make double-stranded breaks in specific DNA sequences, resulting in site-specific homologous recombination and targeted modification of specific loci. These tools can perform targeted gene modifications or inactivations.

**zinc-finger protein:** Proteins with tandemly repeating segments that bind zinc atoms. Each segment contains two closely spaced cysteine molecules followed by two histidines. Each segment folds upon itself to form a finger-like projection. The zinc atom is linked to the cysteines and histidines at the base of each loop. The zinc fingers bind to DNA molecules and can regulate transcription.

**zygote:** A fertilized egg formed as the result of the union of male and female gametes.

# Index

Note: Page numbers followed by "f" and "t" refer to figures and tables, respectively.

## A

- A-DNA 14–15, 737
- abdominal-A* gene 283
- abdominal-B protein 431
- Abraxas grossulariata* 442
- Acari (mites) 138, 374,  
380–381, 682t–683t
- accessory chromosomes see  
chromosomes
- Ace* gene 495
- acentric fragments see  
chromosomes, acentric
- acetylcholine 495
- acetylcholinesterase
  - enzyme 495
  - gene 495, 682t–683t
- Acheta domestica* 469–470
- Accord* see transposable  
elements
- Acraea encedon* 442, 449
- acridine dyes 29
- Acromyrmex* 146, 280t, 284
- Acrythosiphon pisum* 107–108,  
110, 148, 280t, 283, 491,  
501–502, 681,  
682t–683t
- actin 18–19, 737
- actin* gene 354, 577–578
- action potential 494, 737
- Activator* see transposable  
elements
- active transport 54
- Adalia bipunctata* 433t,  
443t–444t
- Adalia decempunctata*  
443t–444t
- additive genes 737
- Adelges tsugae* 578

- adenine 3–4, 8, 10f, 15f, 737
- adenosine triphosphate see ATP
- adrenoleukodystrophy 504
- adzuki bean borer see *Ostrinia  
scapularis*
- Aedes aegypti*
  - domesticated populations  
641–642
  - genes cloned from 353–354
  - genetic modification 394,  
665t–667t, 682t–683t,  
691, 700, 703b–707b
  - genome 107–108, 280t, 281,  
294, 646
  - house-entering behavior  
471–474
  - meiotic drive 440–441
  - sex determination 439
  - transposable elements 394
  - Wolbachia* injection  
665t–667t, 679
- Aedes albopictus* 65, 130
- colonization and origins 599
- diapause in 289
- genome organization 57–58,  
73, 75
- transcriptome 289
- Aedes atropalpus* 475
- Aedes taeniorhynchus* 643
- Aedes triseriatus* 75
- Africanized honeybees see  
*Apis mellifera*
- agarose gel electrophoresis  
205–206, 737
- DNA recovery 210–211
- Ageniaspis citricola* 578
- Agrilus planipennis* 641
- alanine 18t

- alcohol abuse 504–505
- alcohol dehydrogenase* gene  
633
- alignment 556, 737
- alkaline phosphatase 195
- Allacma fusca* 437
- Allee effect 624
- allele frequency 737
- alleles 531–532, 737
- allochrony 624
- allopatric speciation 573–574
- allozymes 533–535, 596–599,  
609–612, 737
- electrophoresis 595–599
  - see also isozymes
- alphaviruses 678–679
- alternative splicing 62, 66, 737
  - see also genes, regulation
- Alul* see restriction  
endonucleases
- Alzheimer's disease 271, 504
- amber
- insects in 347, 348b
  - see also polymerase chain  
reaction
- Amblyomma* 355
- American foulbrood
  - see also Insect behavior
- ametabolous insects 161
- amino acids 18t, 528, 737
- aminoacyl tRNA synthetase  
48–50, 737
  - see also transfer RNA
- amnesiac gene 487–488, 488t,  
489–490
- ampicillin resistance 197, 198f
  - see also selectable markers
- amplicons 274–275, 737

- amplification 737  
multiple displacement 359,  
760  
plateau 327  
amplified fragment length  
polymorphisms see  
polymerase chain  
reaction, AFLP  
AmpliTaq 358  
analysis of molecular variation  
(AMOVA) 632, 635  
anaphase 86, 738  
*Anaplasma* 95  
*Anarhichas lupus* 387  
*Anastrepha fraterculus* 629  
*Anastrepha ludens* 665t–667t  
*Anastrepha suspensa* 398, 453,  
665t–667t  
*Andrena* 607  
aneuploidy 421, 426–427, 738  
Angstrom 738  
annealing 14, 21–22, 23f, 93,  
185f, 738  
annotating genome data see  
genomics  
*Anopheles*, telomeres 83  
*Anopheles albimanus* 452b, 479  
*Anopheles arabiensis* 355–356  
*Anopheles culifacies* 439  
*Anopheles gambiae* 490–491,  
665t–667t, 673t  
gene silencing in 406–407  
genetic manipulation of 676  
genome 107–108, 279–280,  
281, 294, 610t  
insecticide resistance 355–356  
sex determination 439  
transformation of 396  
transposable elements in  
396  
*Anopheles maculipennis* 65  
*Anopheles quadrimaculatus* 75  
*Anopheles stephensi* 396,  
665t–667t  
*Anoplura* 81, 129, 138, 161  
*Antennapedia* gene 60,  
160–161, 163–164  
anterior-posterior polarity 154  
see also development  
*Antherea pernyi* 75, 111,  
126–127  
*Antherea polyphemus* 124–127  
*Anthonomus grandis* 116–117  
antiaphrodisiac 496  
antibiotic resistance 197, 198f,  
688  
antibody 738  
*Anticarsia gemmatalis* 243  
anticodons 49, 738  
antifreeze protein genes 387  
antigen 738  
antiparallel 9b, 738  
antisense strand 16  
antlions see *Myrmeleontidae*  
ants see *individual species*  
Aphelinidae 134  
aphids 109, 146–148  
see also species names  
*Aphis pisum* 109  
*Aphytis melinus* 691  
*Apis cerana* 432–433, 578  
*Apis mellifera* 737  
and Africanized honey bees  
478  
alfalfa pollen collection 479  
behavior 491–492  
European species 578  
foulbrood susceptibility  
472–473  
genetic improvement 662,  
664, 682t–683t  
genome 65–66, 75, 107–108,  
280t, 281–282, 288, 607,  
610t  
honey production  
heritability 473, 478  
hygienic behavior genetics  
472–473  
mitochondrial genome 540f  
personality 501–502  
scouting behavior 501–502  
sex determination 432, 433t  
social behavior 507–508  
stinging behavior 479  
and *Varroa* 280t, 473, 478,  
578, 608  
apomixis 137b, 423–424, 433t  
apomorphic 550, 738  
apoptosis 738  
Applied Biosystems SOLiD  
Sequencer 275  
*Arabidopsis* 640  
Araneae 138  
Archaea 17–18, 128, 142, 315,  
738  
Archaeabacteria 40–41, 564  
Archaeognatha 569  
Argentine ant see *Linepithema*  
*humile*  
arginine 18t, 76  
Argonaute 57  
*Armadillidium album* 446  
*Armadillidium nasatum* 446  
arrhenogenic females 738  
arrhenogenous gene 434–435  
arrhenotoky 104–105, 106,  
110, 134, 431, 432, 433t,  
738  
see also sex determination  
*Arsenophonus nasoniae*  
443t–444t  
arthropods  
classification of 523, 524t  
collection of 594–595  
endopolyploidy 107  
evolution 521  
fossil record of 551–553,  
568–571  
pathogen detection in  
354–355  
phylogeny 548, 571–573  
transgenic see genetically  
modified arthropods  
artifacts  
of PCR primers 326, 766  
phylogenetic trees 561–562  
primer-dimer 326, 766  
artificial insemination 673t  
artificial parthenogenesis see  
parthenogenesis  
artificial selection 738  
asexual see thelytoky  
*Asobara tabida* 478  
asparagine 18t  
aspartic acid 18t  
associative learning 485–486,  
491–492, 738  
assortment see independent  
assortment  
asymmetric PCR see  
polymerase chain  
reaction  
*Athalia rosae* 432–433, 434f

- see also sex determination  
 Atlantic wolfish see *Anarhichas lupus*  
 ATP 96, 738  
 ATPase 97f, 681, 682t–683t  
*Atta cephalotes* 280t, 284  
 attacins 114  
 attine leaf-cutter ants 143–146  
 AUG see codons  
 Australian sheep blowfly see *Lucilia cuprina*  
 autecology 593, 738  
 autoclaving 328  
*Autographa californica* NPV 243, 677–678  
 automixis 433t  
 autoparasitoids 138–139, 436–437  
 autoradiography 237, 738  
 autoregulatory control 739  
 autosomes see chromosomes, autosomes  
 avoidance of pesticides see behavior; insecticide resistance
- B**
- B chromosomes see chromosomes, accessory (B-type; supernumerary)  
*Bacillus* 315  
*Bacillus sphaericus* 348b  
*Bacillus stearothermophilus* 273–274  
*Bacillus subtilis* 216–217, 676–677  
*Bacillus thuringiensis* 131, 616–617, 684–685  
 back mutations see mutations, back  
 backcross 471–473, 479  
 BACs see bacterial artificial chromosomes  
 Bacteria 17–18, 739  
 see also individual species  
 bacterial artificial chromosomes (BACs) 216–217, 230, 739  
 bacterial conjugation 739  
 bacteriocytes 110, 128–129, 146–147  
 see also symbionts  
 bacteriomes 131, 148–149  
 see also symbionts  
 bacteriophage plaques see plaques  
 bacteriophages 184, 739, 763  
 cloning vectors 219–226  
 helper 230  
 lambda 194, 215–216, 219–226, 756  
 M13 216–217, 228–230, 759  
 WO 136  
 bacterioids 130, 138  
 see also symbionts  
*Bactrocera correcta* 453  
*Bactrocera cucurbitae* 398, 629  
*Bactrocera dorsalis* 398, 453, 682t–683t  
*Bactrocera oleae* 629  
*Bactrocera tryoni* 397, 607–608, 665t–667t  
 baculoviruses 216, 677–678, 739  
 expression vectors 198–200, 238–240, 242–243  
 see also vectors  
*Baetica ustalata* 65  
 Bal31 nuclease 234, 235t  
 balancer chromosomes see chromosomes, balancer  
*Bam* see restriction endonucleases  
 Barcode of Life Project 542–544  
 barcoding 523, 542–544, 546, 603, 607–608, 644, 739  
 base pairing 739  
 complementary 9b, 14, 15f  
 base-excision repair 29, 739  
 bases 3–6, 739  
 Basic Local Alignment Search Tool see BLAST  
 Bateson, William 159  
 BATWING 563  
 Bayesian inference 560–561  
*Bdelloura candida* 396  
 bean beetle see *Callosobruchus chinensis*  
 BEAST 563, 739  
 bees see individual species  
 behavior aggressive behavior in *D. melanogaster* 507  
 associative learning 485–486, 491–492, 738  
 avoidance of pesticides 479  
 cannibalism 479  
 chemotaxis 478–479, 741  
 collection of alfalfa pollen 479  
 conditioning of leg position 489–490  
 courtship 484–485, 497–499  
 death-feigning 501  
 definition 467  
 diapause 485  
 fate map analysis of 471  
 flight 475  
 foraging 471–472, 474–475  
 genetic analysis 465  
 geotaxis 475, 752  
 honey production by bees 473, 478  
 host-preference 475  
 host-selection 138b  
 house-entering in mosquitoes 471–474  
 hygienic in honey bees 472–473  
 learning 468–469  
 locomotor activity 481  
 mating 466, 497–499  
 migratory flight of *Oncopeltus* 476–477  
 olfactory learning 490–491  
 personality 501–502  
 pheromone communication 475–476, 492–493  
 photoperiodic clock 480–486  
 phototaxis 475, 764  
 polygenetically determined 478–479  
 preening 478–479  
 silk mat spinning by *E. kuhniella* 475  
 sitters 474–475  
 song cycle in *Drosophila* 484–485  
 stinging 479  
 walking 489, 501  
 see also period gene

- Bemisia tabaci* 640, 644, 682t–683t  
*bicoid* gene 155  
 biofilms 142–143  
 biogeography 540–541  
 bioinformatics 253, 277–279, 739  
 biological clock see *period gene*  
 biological species concept 573–575  
 biotechnology 740  
 biotin labels 241  
 biotype identification 644–645  
 bisexual 498  
*Biston betularia* 643  
 bithorax complex 161, 163–164  
   see also development  
*bithorax* gene 160f, 161  
 bivalent 740  
 BLAST 268, 292, 553–556, 740  
 blastoderm 151, 152f, 154, 740  
   see also development  
*Blattaria* 161, 682t–683t  
*Blatella germanica* 161–162, 673t, 683  
*Blattodea* 83, 523  
 blowfly see *individual species*  
 blunt ends 192–193, 195, 196, 196t, 740  
 BmNPV 243  
 boll weevil see *Anthonomus grandis*  
*Bombus terrestris* 642  
*Bombyx mori*  
   behavior genetics 475  
   BmNPV 243  
   chorion genes 124–127  
 DNA  
   methylation 65  
   unique-sequence 111  
 genetic improvement 662, 664–667, 665t–667t, 673t, 693  
 genome 57–58, 107–108, 123, 280t, 281, 288, 610t  
 nuclear polyhedrosis virus (NPV) 677–678  
 polyploidy 73  
 sex determination 453  
 silk glands 116  
 telomeric sequence 83  
 transformation of 395  
 transposable elements in 120, 395–396  
 vitellogenin genes 116–117  
 bootstrapping 559–560, 740  
*Borrelia burgdorferi* 322, 350–351  
 bottlenecks 740  
 boundaries 63–64  
   see also insulators  
 boundary elements 41  
*Brachys tesselatus* 443t–444t  
*Bracon hebetor* 432  
*Braconidae* 134, 139  
 branch length 557–558  
*Brevipalpus* 138b  
*Brevipalpus phoenicis* 433t, 449  
 broad-sense heritability 740  
 5-bromouracil 29  
*Bryobia sarotheamni* 138b  
*bubblegum* gene 504  
*Buchnera aphidicola* 110, 146–147, 529  
 buffers 316t–317t, 740  
 bumblebee see *Bombus terrestris*  
*Burkholderia* 131
- C**
- C-banding 535–536, 740  
 C-value paradox 73–74, 740  
 CAAT box 44–45, 44f  
*cabbage* gene 487, 488t, 489–490  
*cabbage looper* see *Trichoplusia ni*  
*Caenorhabditis elegans* 56–57, 72–73, 119, 268, 270–271, 377, 392, 568  
 calcium channels see channels, calcium  
 calcium chloride 201–202  
 calf intestinal phosphatase (CIP) 234, 235t  
*Calliphoridae* 395  
*Callosobruchus chinensis* 109  
 Cambrian period 493, 568, 570t  
 cAMP 487–488, 744  
*Campoletis sonensis* 139  
*Camponotus floridanus* 284  
*Candidatus Midichloria mitochondrii* 131–132  
*Candidatus Riesia pediculicola* 282–283  
 cannibalism 479  
 capping 46, 82, 740  
*cappuccino* gene 156–157, 430  
 carbohydrates 740  
 Carboniferous period 569, 570t  
*Cardinium* 132–139, 138–139, 359, 360b–361b, 740  
   modification of sex determination 420, 433t, 443t–444t, 449  
*Cardiochiles diaphaniae* 673t  
 Carnegie 20 virus 386f  
*caudal* gene 682t–683t  
 cDNA see complementary DNA  
*Cecidomyiidae* 92  
 cecropins 114, 675  
*Celera, Drosophila Genome Project* 269–270  
 cells 740  
   autonomous sex determination 427, 741  
   determination 745  
   division see meiosis; mitosis  
 cell cycle 25f, 26, 84–90, 741  
   check points 86  
   gap phases 26  
   M phase 84–85  
   S phase 26, 84–85, 87, 770  
 cell-penetrating peptides 406  
 cellular rhythms see *period gene*  
 cellularized blastoderm 151  
 centimorgan 741  
 Central Dogma 6–8, 38–39, 741  
 centromere-like elements 109  
 centromeres 71–72, 81–82, 741  
   see also chromosomes  
*Cephalothorax* gene 163–164  
*Cerambycidae* 130  
*Ceratina* 607

- Ceratitis capitata* 688  
 eradication of in California 452*b*, 625–639, 628*f*  
 gene amplification 125  
 genetic control 665*t*–667*t*  
 genome 125  
 geographic variation 630–637  
 population diversity 628–630  
 transformation of 395, 398  
 vitellogenins 116–117
- Ceratitis rosa* 629–630
- Ceratovacuna nekoashi* 334
- cercus* gene 439
- cesium chloride centrifugation 202
- Chaetophiloscia elongata* 446
- Chagas' disease 121–122, 284–285, 675
- Chain-terminating method of DNA sequencing see Dideoxy method and DNA sequencing
- channels 741  
 ion 493–496, 755  
 potassium 495  
 sodium 495
- chaotropic salts 188–189
- chaperones 56, 95–96, 741
- chaperonin* gene 95–96
- characters 549–550  
 see also molecular systematics
- Chaser* gene 474–475
- check points 86
- chelating agents 187–188, 741
- Chelex DNA preparation method 320–321, 320*t*, 607
- Chelicera 283–285
- chemiluminescent labels 210, 241
- chemotaxis 478–479, 741
- chiasmata 89  
 see also crossing over
- chickens 396–397, 674
- Chilopoda 524*t*
- ChIP-chip 290
- Chironomidae 92, 105–106
- Chironomus* polysomes 50–51, 51*f*  
 telomeres 83
- Chironomus pallidivittatus* 83
- Chironomus plumosus* 65
- Chironomus tentans* 75–76, 83
- Chironomus thummi* 111, 113
- chitin 741
- chitin synthase 406–407, 665*t*–667*t*
- chitinase 148–149, 741
- choline acetyltransferase (CHAT) 495
- chorion 124, 741  
 genes 18–19, 111, 124–127, 125*f*
- chromatids 33, 85*f*, 741  
 see also sister chromatids
- chromatin 741  
 10-nm fiber 774  
 see also different types
- chromomere 741
- chromosomes 9*b*, 53, 72, 76–77, 741  
 accessory (B type; supernumerary) 94, 110–111, 450, 739  
 acentric 81, 92, 737  
 amplification 108, 123–124, 354, 750  
 artificial 673*t*  
 autosomes 427, 739  
 balancer 384, 739  
 centromeres see centromeres  
 condensation 77–78, 77*f*, 742  
 crossing over 9*b*, 33, 89, 90, 420–421, 499–500, 744  
 damage 90–92  
 dicentric 92, 745  
 holocentric 81, 753  
 homologous 754  
 imprinting 64–66, 107, 742  
 interphase 26, 61–62, 76, 80, 755  
 library 742  
 linkage maps 757  
 metaphase see metaphase  
 mitochondrial 98  
 origins of replication 21–22, 26, 200, 687, 762, 768
- paternally derived 107, 110–111, 431
- polytene see polyteny
- puffing 91*f*, 93, 742, 767
- rearrangements 18–19, 80–81, 91, 119, 377
- replication see replication ring 769
- salivary gland 50–51, 51*f*, 91*f*, 92–93
- sex 94–95, 423, 771  
 W chromosome 432  
 X chromosome 95, 425–427, 777  
 Y chromosome 81, 94–95, 427, 432, 778  
 Z chromosome 432, 778
- structure 535
- telomeres see telomeres
- translocations 92
- walking 216, 237–238, 237*f*, 742  
 see also DNA
- Chrysomya rufifacies* 436
- Chymomyza costata* 485
- cicadas see *Magicicada*
- circadian 466, 504–505  
 clock 466, 480–486, 742  
 rhythms 742  
 of eclosion 480*t*, 482  
 of locomotion 480*t*, 482
- Circular DNA molecules see DNA cis 742
- cisternae 55
- CITES 594–595
- citrus mealybug see *Planococcus citri*
- clades 742
- cladistic systematics 526, 550, 742  
 see also systematics
- cladogenic speciation 549–550
- cladograms 551, 742
- clamp-loading proteins 27
- class I and II elements see transposable elements
- class I, II, and III genes 42
- clock  
 circadian 466, 480–486, 742  
 molecular 528, 541

- Clock* gene 466, 480*t*, 481, 484, 486, 504–505  
 cloning 184–185, 184–185, 185*f*, 215, 742  
 blunt-ended 192–193, 195, 196, 196*t*, 740  
 cDNA 215–216, 217, 230–234, 232*f*  
 competent cells 229  
 definition 216  
 DNA polymerase I of 232*f*  
 enzymes 184  
 expression of cloned genes 200, 238  
 F pili 228–229  
 heat shock genes 18–19, 112–113  
 into plasmid vectors 185*f*, 186–187  
 kits 185–186, 224  
 ligase 23–25, 184, 195–196, 235*t*, 746, 757  
 overhangs 193, 193*t*, 195  
 restriction endonucleases 184, 192–195, 768  
 shotgun 269–270, 771  
 strategies 219–226  
 transformation see transformation vectors 184, 185*f*, 186, 216–217, 742  
 cosmids 226–228, 227*f*  
 lambda 183–184, 215–216, 219–226  
 M13 216–217, 228–230, 759  
 pBR322 200, 217  
 phagemids 216–217, 230, 763  
 pUC series 199*f*  
 see also transformation  
 cloud computing 275–276  
 cloverleaf structures 49  
 ClustalW 556, 563  
 cluster analysis 563, 616–617, 742  
 Coccinellidae, male killing 450–451  
*Cochliomyia hominivorax* 644  
 eradication of 451–453, 452*b*, 668  
*Cochliomyia macellaria* 644  
 cockroaches, symbionts 142  
 coding strand 16, 743  
 codling moth see *Cydia pomonella*  
 codominant 743  
 codons 16–17, 743  
 bias 743  
 initiation 17, 755, 772  
 punctuation 17  
 termination (stop) 17, 49, 774  
 coefficient of gene differentiation (GST) 612, 743  
 coffee berry borer see *Hypothenemus hampei*  
*Coleomegilla maculata* 443*t*–444*t*  
*Coleoptera* 83, 105–106, 106–107, 120, 123, 129, 138, 161, 348*b*, 443*t*–444*t*, 547, 682*t*–683*t*  
 collecting insects for analysis 594–595  
*Collembola* 92, 138, 161, 569  
 colonization of insects 599  
 colony 743  
 colony hybridization 240, 743  
*Colorado beetle* see *Leptinotarsa decemlineata*  
 community structure 594  
 comparative genomics 253, 294–295, 526–527  
 compensasomes 426–427  
 competent cells 201*t*, 743  
*E. coli* 201*t*  
 complementary base pairing 9*b*, 14, 15*f*, 240, 743  
 complementary DNA (cDNA) 215–216, 230–234, 232*f*, 740, 743  
 libraries 217–218, 231  
 complementary sex determiner gene 433*t*  
 complex traits 272, 295, 743  
 concatemers 743  
 concerted evolution 126, 743  
 conditional lethal mutations see mutations, conditional lethal  
 conditioned stimulus 485–486, 743  
 conjugation see bacterial conjugation  
 consensus sequences 238, 743  
 constitutive enzymes 743  
 constitutive heterochromatin see heterochromatin, constitutive  
 contigs 265–266, 743  
 convergent evolution 140, 743  
*copia*-like elements see transposable elements  
 copper-zinc superoxide dismutase see superoxide dismutase gene  
*Coptotermes* 142  
 copy number 71–72, 98, 113, 121, 147, 217, 744  
 core DNA 76–77  
 corpora  
 allata 469  
 cardiaca 469  
 Correns, Carl 4–5  
 cos sites 219–220, 222, 226–227, 744  
 cosmids 216, 744  
 cloning vectors 219, 226–228, 227*f*  
 genomic library construction 228  
 cotton bollworm see *Helicoverpa armigera*  
 couch potato gene 467, 475  
 courtship behavior in *Drosophila* 497–499  
 courtship song 484–485  
*Cowdria ruminantium* 355  
 CpG island 277–278, 744  
 Creutzfeld-Jakob disease 7, 504  
 Crick, Francis 5–6  
 crickets see *Laupala* and species names  
 cristae 96  
 crossing experiments 471–476

- see also behavior, genetic analysis
- crossing over 9*b*, 33, 89, 90, 420–421, 499–500, 744
- Crustacea* 83, 569
- CRUZIGARD** 675
- cryptic species 577–578, 614–615, 622, 644–645, 744
- cryptochrome* gene 466, 480*t*, 481, 484
- CTAB** 322
- Culex*, transposable elements
- Culex bitaeniorhynchus* 65
- Culex pipiens* 75
- dispersal of resistance 119–120
  - genetic manipulation of 682*t*–683*t*
  - genome organization 81
  - genome sequencing 280*t*, 281
  - gynandromorphs 439
  - meiotic drive 440–441
  - reproductive incompatibility 689–691
  - resistance gene amplification 127
  - sex determination 439
- Culex quinquefasciatus* 57–58, 114–115, 127, 280*t*, 281, 441
- Culex tarsalis* 645
- Culicoides variipennis* 534–535
- Curculionidae* 97–98, 105–106
- cuticle genes 18–19
- cuticular hydrocarbons 499
- cycle* gene 466, 480*t*, 481, 484, 485, 504–505
- cyclic adenosine monophosphate see cAMP
- cyclical thelytoky see thelytoky
- cyclin-dependent protein kinases 86–87
- cyclins 86–87
- Cydia pomonella* 243, 695
- Cynipidae* 134
- Cyphomyrmex longiscapus* 145*f*
- Cyrtodiopsis* spp. 441
- cysteine 18*t*
- cystoblast 149–150
- cytochrome 744
- cytochrome b 97*f*, 348*b*, 540*f*
- cytochrome c 528–529
- cytochrome c oxidase subunit 1 gene 97*f*, 540*f*, 542, 607
- cytochrome P450 gene 646
- cytogenetics 105, 532
- cytokinesis 26, 744
- cytological map 92–93, 744
- cytoplasm 53–55, 744
- cytoplasmic
- incompatibility 132–133, 137*b*, 138*b*, 744
  - inheritance 744
  - sex ratio distorters 441–450, 744
- cytoplasmic polyadenylation element-binding proteins (CPED) 490
- cytosine 3–4, 8, 10*f*, 15*f*, 64–65, 744
- D**
- Dalton 744
- Danaus chrysippus* 442
- Danaus plexippus* 107–108, 280*t*, 284
- Darwin, Charles 524–525
- daughterless* gene 425*t*, 428–429
- ddNTPs 252, 255–258, 258–260, 258*f*
- de Vries, Hugo 4–5
- death-feigning behavior 501
- defensins 114
- degenerate genetic code 17, 744
- see also genetic code
- degenerate primers 330*t*–331*t*, 334–335, 353, 744
- deletions see mutations, deletions
- denaturation 745
- Dendroctonus frontalis* 146
- dendograms 551
- dengue fever 130, 705–706, 709
- Denhardt's solution 208
- density gradient centrifugation 122, 745
- densonucleosis viruses 678
- 2'-deoxyadenosine
- 5'-triphosphate (dATP) 12*f*
  - see also DNA
- 2'-deoxycytidine
- 5'-triphosphate (dCTP) 12*f*
- 2'-deoxyguanosine
- 5'-triphosphate (dGTP) 12*f*
- deoxyribonuclease (DNase) 234, 235*t*, 745
- deoxyribonucleic acid see DNA
- deoxyribonucleotides see dNTPs
- 2'-deoxyribose 9*f*
- 2'-deoxythymidine
- 5'-triphosphate (dTTP) 12*f*
- Dermoptera* 83, 105–106, 120, 161
- deuterotoky 106, 745
- deutocerebrum 469
- development 149–154
- D. melanogaster* 105, 149–154, 485
  - analyzed with mutants 154–162
  - embryonic 151–153
  - gene interactions 162
  - imaginal discs 153, 754
  - long-germ band 152*f*
  - maternal effect genes 150
  - oocyte 149–151
  - parasegments 156*f*
  - postembryonic 153–154
  - sensory organs 494
  - short-germ band 151
  - zygotic-segmentation genes 157–161
- evo-devo 105, 164–166
- of insects other than *Drosophila* 162–164
- long-germ band 151, 757
- short-germ band 151, 771
- Devonian period 569, 570*t*

- Diabrotica virgifera virgifera* 673t  
*Diadegma chrysostictos* 382–383, 432–433  
*Diadromus pulchellus* 82, 432–433  
 diamondback moth see *Plutella xylostella*  
 diapause 485, 646  
 dicentric chromosomes see chromosomes, dicentric  
*Dicer* 57, 402  
*Dicyphus tamaninii* 344  
 dideoxy method of DNA sequencing (Sanger chain-terminating) 252, 254, 255–261, 255f, 256f, 289, 537–538, 554f–555f, 741, 745 automated 264–265  
 dideoxynucleotides see ddNTPs  
 differential splicing 62  
*Diglyphus begini* 605–606  
 dimethyl sulfoxide 201–202, 325  
 diploidy 71–72, 87–89, 104–105, 422–423, 431, 745  
*Diplolepis rosae* 447–448  
*Diplura* 105–106, 161, 569  
*Diptera* 92, 105–106, 106–107, 120, 129, 138, 161, 547, 682t–683t  
 direct amplification of length polymorphisms see polymerase chain reaction, DALP  
 direct repair 29  
 direct repeats 745  
 discontinuous genes 19, 745  
 discrete character 558, 745  
 disease control evaluation 357, 645  
 dispersal 592–594, 603, 631  
*dissatisfaction* gene 498–499  
*distalless* gene 682t–683t  
 distance, genetic 558, 609–611, 745 estimates 609–611, 745  
 distance-matrix methods 558  
*Distorter* gene 441  
 divergence 745  
 divergence-time estimate 745  
 DMSO see dimethyl sulfoxide  
 DNA 4–6, 71–72, 745 amplification 314f, 597t–598t, 737 see also genes, amplification; polymerase chain reaction analysis 538–549 ancient 321–322, 321t, 347–350 barcoding 523, 546, 603, 607–608, 644, 739 base ratios 76 binding domains 158 chimeric 741 chips see microarrays circular molecules 97f cloning 184–185, 185f, 215–216, 686 complementary (cDNA) 215–216, 217–218, 230–234, 232f composition 76 core 76–77 cutting see restriction endonucleases damage see DNA, repair data banks 267–268 databases 253–254 degradation 349 denatured 745 dinosaur 347–350 double helix 5–6, 13–14, 746 double-stranded 193t, 233, 747 driver 747 engineering 220–221, 356–357, 356f exogenous 186, 748 extraction 186–187 fate of fed 695 fingerprinting 330t–331t, 534t, 538, 595–596, 746, 749 foldback 749 forms 14–15 A-form 14–15, 737 B-form 14–15, 739 C-form 14 D-form 14 E-form 14 H-form 14 Z-form 14, 778 gut 618–620 hairpin structure 11–13 heterochromatic 61–62 heteroduplex 600, 753 heterologous 753 highly-repetitive 111, 122–123, 753 history 4–6 homoduplex 600, 753 hybridization 209, 236, 532, 746, 755 insect 186–187 storage of 318–319 intergenic regions 29, 755 junk 19, 38, 73–74, 756 ligation 195–196 linker 76–77, 757 marker 758 melting 45, 758 methylation 64–65, 354, 746, 759 microarrays see microarrays microsatellite 122–123, 534t, 597t–598t, 602–604, 610t, 612, 613t, 636, 759 middle-repetitive 111–122 minisatellite 602–603 mitochondrial (mtDNA) see mitochondrial DNA moderately repetitive 759 noncoding 73–74 old 350–351 packaging 77–79 parasitic 73–74 polymorphism 746 precipitation 191t probes 209–210, 537, 746 rearrangements 18–19, 108 recombinant 662–663, 768 recovery from agarose gels 210–211 repair 28–29, 746 repetitive 74–75, 768, 774

- replication 4, 21–22, 25–27, 746  
*see also* replication  
 right-handed 14  
 satellite 75, 82, 122–123, 546, 770  
 selfish 73–74, 771  
 sequencing 251, 527–528, 537–538, 746  
 applications 534t  
 methods 358–359, 551–553, 552f  
 vectors 198–200  
 shearing 192  
 single copy nuclear 98, 546, 551–553  
 single-stranded (ssDNA) 229, 772  
 size markers 212, 339f  
 size-fractionating 226  
 spacer 404  
 storage in liquid nitrogen 188t, 201t  
 structure 5–6, 8–11  
 supercoiled 77–78, 773  
 synonymous changes 773  
 synthesis 21  
 template 22  
 tracer 775  
 transcription 16, 41–42, 52b, 746, 775  
 translation 16, 48–52, 776  
 triplex 776  
 unique-sequence 111  
 DNA ligase 23–25, 184, 195–196, 235t, 746, 757  
 DNA polymerase 21–22, 26–27, 233, 234, 235t, 309, 313–315, 324, 746  
 slippage 603  
 DNA topoisomerase 25, 746  
 DNA vaccines 746  
 DNA-binding protein 27, 746  
 DNA-dependent RNA polymerase 746  
 DNase I 234, 235t  
 dNTPs 209–210, 255–258, 310, 313, 316t–317t, 325, 744, 745  
 Dolly the sheep 216  
 domain 746  
 domains of life 746  
 dominant traits 4–5  
*see also* genetic terminology  
 dopamine hypothesis 504–505  
 dorsalizing genes 155  
 dorso-ventral polarity 155, 165  
 dosage compensation 425–427, 426f, 746  
 dot-blot 209  
 double helix 5–6, 13–14, 746  
 double switch gene 429–430  
 double-strand conformation polymorphism (DSCP) 592–593, 597t–598t, 599–600  
*doublesex* gene 47, 420, 425t, 429–430, 432, 436, 438, 438f, 449, 453  
*doubletime* gene 466, 481, 484, 504–505  
 downstream 438, 747  
 drift *see* genetic, drift  
 driver DNA 747  
*Drossha* 57  
*Drosophila*  
 centromeres 81–82  
 chorion gene amplification 124–127, 125f  
 development 149–154  
 developmental mutants 154–162  
 DNA barcoding 543  
 genes cloned from 353  
 Genome Project 72–73, 265, 268–272  
 helix-turn-helix DNA-binding proteins 60  
 histone genes 113  
 immune response genes 113–115  
 maternal effect genes 150, 155–157  
 meiosis 90  
 minichromosomes 81  
 mitochondrial DNA 96  
 polytene salivary gland chromosomes 91f  
 replication sites 21  
 reproductive incompatibility 689–691  
 segmentation genes 155, 157–161  
 song cycle behavior 484–485  
 telomeres 83  
 vitellogenin genes 116–117  
*Drosophila ananassae* 109, 280t, 294, 500  
*Drosophila bifasciata* 379, 443t–444t  
*Drosophila birchii* 576  
*Drosophila buzzatii* 31  
*Drosophila cut* gene 45  
*Drosophila erecta* 280t, 294  
*Drosophila funebris* 120–121  
*Drosophila grimshawi* 280t, 294  
*Drosophila guanche* 379  
*Drosophila hawaiiensis* 113, 390  
*Drosophila hydei* 113, 120, 675  
*Drosophila mauritiana* 121, 381, 396–397, 446–447, 576–577  
*Drosophila melanica* 440  
*Drosophila melanogaster*  
 aggressive behavior 507  
 behavior genetics 465  
 chorion gene amplification 111, 124–127  
 circadian locomotor activity 481  
 comparative genomics 57, 58f  
 courtship behavior 484–485, 497–499  
 cytoplasmic incompatibility 136, 137b  
 development 105, 149–154, 485  
 analyzed with mutants 154–162  
 embryonic 151–153  
 gene interactions 162  
 imaginal discs 153, 754  
 long-germ band 152f  
 maternal effect genes 150  
 oocyte 149–151  
 parasegments 156f

- Drosophila melanogaster*  
 (Continued)  
 postembryonic 153–154  
 sensory organs 494  
 short-germ band 151  
 zygotic-segmentation genes 157–161  
 diapause 485  
 DNA 76  
   extraction method 187–190, 188t, 323t  
   methylation 65  
   repetitive 75  
*Est-5* and *Est-6* gene  
   regulation 496–497  
 foraging behavior 471–472, 474–475  
 freeze resistance in 387  
 gene silencing in 394, 399  
 Genome Project 72–73, 265, 268–272, 269, 270–271, 280t, 568, 610t  
 genomic imprinting 65  
 heat shock genes 112–113  
 heterochromatin 80–81  
 histone genes 113  
 homeotic genes 159–161  
 and human diseases 503–506  
 incipient speciation 574t  
 learning in 468–469, 486–490  
 M' strains 391  
 meiotic drive 440, 758  
 methylation of DNA 65  
 microarray 243–247  
 microinjection of 384–385  
 minichromosomes 109  
 mitochondrial DNA 96  
 mitochondrial genome 96  
 odor behavior of 490–491  
*P* cytotype 390–391  
*P* element-mediated transformation 116, 121, 379–381, 672, 694  
*para* gene 353, 475  
*period* gene 466, 481, 482–484, 483f, 485, 486  
 position effect variation 63–64, 387, 765  
 Q strains 391  
 replication 26  
 reproductive isolation 576  
 resistance genes 387f  
 ribosomal genes 115  
 salivary gland chromosomes 91f, 92–93  
 sex determination 47, 420, 424–431, 440, 497  
*in situ* hybridization 93, 436  
 sleep 505–506  
 somatic-sex determination 427–430, 428f  
 speciation genes in 499–501  
 systems genetics 295–297, 506–507  
 telomeres 83, 287  
 transformation with TEs 381–386  
 transposable elements in 119–120  
*Wolbachia* transfection 134  
   Y chromosome 81  
*Drosophila miranda* 81  
*Drosophila mojavensis* 280t, 294  
*Drosophila nasutoides* 75  
*Drosophila nebulosa* 391  
*Drosophila obscura* 440  
*Drosophila palidosa* 500  
*Drosophila paulistorum* 443t–444t, 445  
*Drosophila persimilis* 280t  
*Drosophila pseudoobscura* 280t, 294, 482–484, 496–497, 572  
*Drosophila quinaria* 440  
*Drosophila sechellia* 280t, 479, 576–577  
*Drosophila serrata* 576  
*Drosophila simulans* 97–98, 119–120, 134, 280t, 381, 390, 479, 482–484, 499, 576–577, 696  
*Drosophila subobscura* 478  
*Drosophila testacea* 440  
*Drosophila tripunctata* 440, 478–479  
*Drosophila virilis* 65, 82, 280t, 294, 482–484  
*Drosophila willistoni* 280t, 379, 381, 443t–444t  
*Drosophila yakuba* 96, 97f, 280t, 482–484, 542  
 Drosophilidae 92, 395  
 drug addiction 504–505  
*DSC1* gene 353  
 Duchenne muscular dystrophy 271  
*Dugesia tigrina* 396  
*dunce* gene 467, 470f, 487, 488–489, 488t, 488t, 489
- E**
- eag* gene 475  
 ecdysis 60  
 ecdysone 60, 161–162, 747  
 Ecdysozoa 571, 747  
 eclosion 387, 480t, 482, 482–484  
*EcoRI* see restriction endonucleases  
 EDTA see ethylenediaminetetraacetic acid  
 effective population size 596–599, 609, 613  
 egg chamber 149–150  
*Ehrlichia* 95, 341  
 electroelution 210–211  
 electrophoresis 183–184, 204–206, 533–535, 747  
   allozyme 595–599  
 gel 204–206, 291–292, 533, 750, 762, 764  
 protein 533–535, 597t–598t, 599  
 pulsed field 136, 767  
 two-dimensional 291  
 electroporation 202, 673t, 747  
 ELISA 351, 354–355  
 Embiididae 115–116  
 Embioptera 105–106, 161  
 EMBL Data Library see European Molecular Biology Library Nucleotide Sequence Data Library  
 embryogenesis 151–153, 430  
 embryonic development see development

- EMERALD 246  
 emerald ash borer see *Agrilus planipennis*  
 emergent properties 296–297  
 enantiomers 475–476, 747  
 encapsulation of pathogens 114  
*Encarsia*, sex determination 138–139, 436–437  
*Encarsia formosa* 136, 443t–444t, 448  
*Encarsia pergandiella* 138–139, 436–437, 443t–444t  
 Encyrtidae 134  
 end-labeled probes 241  
 endangered species 594–595  
 endonucleases 192–195, 747  
     homing 375, 400, 401b, 405–406  
     restriction see restriction endonucleases  
 endoplasmic reticulum (ER) 54–55, 747  
     membrane channel 55  
     rough 55, 770  
     smooth 55, 772  
 endopolyploidy 107, 747  
 endosymbionts see symbionts  
*engrailed* gene 60, 157, 547  
 enhancers 41, 45, 61–62, 747  
     trap 489, 747  
*Enterobacter agglomerans* 142–143, 143f  
     see also Symbionts  
*Enterobacter cloacae* 695–696  
*Entomocorticium* 146  
 environmental sex determination 436–437, 747  
     see also sex determination  
 enzymes 16, 234, 235t, 747  
     constitutive 743  
     DNA degradation 5  
     DNA replication 4, 27  
     inducible 754  
     see also specific enzymes  
*Eotetranychus suginamensis* 138b  
 Ephemeroptera 115–116, 161  
*Ephestia kuehniella* 318–319, 401, 475  
 epigenetic inheritance 748  
 epigenetic regulation 59b  
 epigenetics 64, 65b  
*Epiphyas posvittana* 682t–683t  
 epistasis 748  
 eradication programs see pest management strategies  
*Erwinia herbicola* 695–696  
*Escherichia coli* 112, 184, 196, 196–201, 218f, 270, 397, 683, 748  
     DNA polymerase I of 209–210  
*Escovopsis* 144  
 Esterase  
     *Est-5* and *Est-6* in *Drosophila* 496–497  
     genes 127  
 ethanol precipitation see DNA, precipitation  
 ethidium bromide (EtBr) 190, 203–204, 203f, 205f, 206, 211, 212, 537, 748  
 ethology 469  
 ethomics 506, 748  
 ethyl methanesulfonate 391–392  
 ethylenediaminetetraacetic acid (EDTA) 187, 747  
 ETS see ribosomal RNA, ETS  
 Eubacteria 40–41, 128, 313, 564, 748  
 euchromatin 57, 71–72, 80–81, 748  
*Eucoilidae* 134  
 Eukarya 17–18  
     see also Eukaryote  
 eukaryote 17–18, 564, 748  
     chimeric origin of 565–568  
     genome duplication in evolution 66–67  
     see also individual species  
*Eulophidae* 134  
*Eupelmus vitellii* 82  
*Eurema hecabe* 433t, 740  
 European corn borer see *Ostrinia nubilalis*  
 European Molecular Biology Library (EMBL)  
 Nucleotide Sequence Data Library 267, 553  
*Euscelidium variegatus* 130  
*even-skipped* gene 157, 157, 158, 158f, 159  
 Evo-Devo 105, 164–166  
 evolution 66–67, 521, 748  
     analyses 357–358  
     concerted 126  
     convergent 140, 743  
     molecular 526–527, 531–538, 760  
     nearly neutral theory 530–531, 760  
 evolutionary developmental biology see Evo-Devo  
 evolutionary species concept 574, 574t  
 excisions see mutations  
 exons 9b, 19, 40–41, 62, 748  
     shuffling 748  
 exonucleases 234, 235t, 748  
 exportins 54  
 expressed sequence tags (EST) 285–288, 608, 640, 748  
 expression  
     cloned genes 200, 238  
     screening 238–240  
     vectors see vectors  
 expression microarray analysis 243–247  
*extramachrochaetae* gene 425t, 429  
 extranuclear inheritance 95–98  
*exuperantia* gene 155  
*Eyprepocnemis plorans* 65
- F**
- F family see transposable elements  
 F pili 228, 748  
 F statistics 749  
 F1 generation 471, 749  
 F2 generation 749  
 facultative heterochromatin see heterochromatin  
 thelytoky see thelytoky  
 fall armyworm see *Spodoptera frugiperda*

- FASTA 292  
 fat body 92, 114  
 fate maps 471, 749  
   see also Behavior  
 FB transposons 749  
 fertility genes 120, 427, 447  
 fertilization 151, 749  
 fibroins 116  
   see also silk  
 Ficoll 208  
 Fire, Andrew 56–57, 401–402  
 fire ant see *Solenopsis invicta*  
 Fitch-Margoliash algorithm 558–559  
 fitness 749  
*flamenco* gene 678  
 flanking sequence 749  
 flight behavior 475  
 flightlessness 475  
 Florida carpenter ant see *Camponotus floridanus*  
 flour beetle see *Tribolium castaneum*  
 FLP recombinase 680, 749  
 FLP-FRT mediated recombination 680  
 fluorescence labeling 210, 601, 602f  
 fluorometric analysis 190  
 fluorophores 245  
 FlyBase 269–270  
 foldback (FB) transposon see transposable elements  
 follicle cells 149–150  
*foraging* gene 474, 508  
 forward genetics 749  
 fossil insects 320t, 321–322, 347–350, 551–553, 568–571  
   see also amber, insects in  
 foulbrood disease in *A. mellifera* 472–473  
 founder effects 592–593, 749  
 fragment length polymorphisms  
   see also restriction fragment length polymorphisms  
 amplified see polymerase chain reaction, AFLP  
 frameshift mutations see mutations, frameshift  
 Franklin, Rosalind 5–6  
 freeze resistance 387  
   see also genetic improvement  
*fruitless* gene 425t, 429, 498  
 FST 749  
 functional genomics 292–293  
*fushi tarazu* gene 158–159  
   see also development  
 fusion proteins 226, 750
- G**
- G element see transposable elements  
 G-banding 750  
 GABA (gamma-aminobutyric acid) 496  
 beta-galactosidase 224, 489  
 gall wasps 106  
 gametes 60, 87–89, 750  
 gamma-aminobutyric acid (GABA) 496  
 gap genes 157–158, 750  
 gap repair 378, 389f  
 gastrulation 445  
 gating 750  
*gemini* gene 507–508  
 GenBank 267, 269–270, 542, 553, 612  
 Gene(s) 16, 750  
   additive 737  
   amplification 108, 123–124, 354, 750  
   boundaries 750  
   chips see microarrays  
   chorion 18–19, 124–127, 125f  
   classes 45  
   cloning 352–354, 750  
   conversion 126, 750  
   discontinuous 745  
   double switch 429–430  
   duplication 123–127, 750  
   essential 748  
   expression 387, 750  
   extrachromosomal 748  
   function 528  
   gap 157–158, 750  
   gun 673t, 685–686, 750  
   homologous 754
- housekeeping 44–45, 59, 528, 754  
 immune-response 113–115  
 imprinting 64–66  
 jumping 378, 756  
 knockout 665t–667t  
 libraries 217–230, 234–240, 750  
 maternal effect see maternal effect genes  
 mitochondrial 95–98  
 multiple-copy 8, 595  
 nonprotein-coding 16  
 numerator 427  
 ontology 279, 750  
 organization 18–21  
 orphan 285–287, 762  
 products 238  
 protein-coding 9b, 37–38, 42f, 546  
   RNA transcripts 45  
 regulation 58–63, 750  
 regulatory 30t, 156f, 768  
 replacement 388, 751  
 reporter 489, 768  
 ribosomal 115  
 segmentation 157–161, 770  
 sequencing see sequencing  
 silencing 80, 399, 665t–667t, 692, 751  
 silk 115–116  
 single-copy 532, 535  
 site-specific modifications 679–680  
 speciation 575–577, 772  
 structural 773  
 targeting 751  
 transfer 95–96, 109–110, 388–390, 751  
 trees 549  
 unique 777  
   see also individual gene names/types  
 Geneious 563  
 general model for sex determination see sex determination  
 general transcription factors 44  
 genetic bottlenecks 751

- code 5–6, 16–18, 751  
control of pests 451–453,  
  689  
distance 558, 609–611, 745  
diversity 751  
drift 633, 751  
duplication 294–295  
engineering 181, 220–221,  
  751  
hitchhiking 541  
improvement 662, 664,  
  682t–683t  
linkage 751, 757  
load 669  
maps 751  
markers 751  
modification for pest  
  management 661  
relatedness see relatedness  
sex determination see sex  
  determination  
sexing method 453  
terminology 32  
variation 634
- genetically modified  
  arthropods (GMA) 662,  
  662–663, 663–664, 664–  
  668, 665t–667t, 752  
containment of 670t–671t,  
  672  
deployment 689–692  
identification 687–689  
methods to develop 672–  
  685, 673t  
nucleic acid delivery 685–686  
release 707b–709b  
  outcome prediction  
  697–699  
regulation of 697–700,  
  700, 701b–702b,  
  702b–703b, 710t–711t  
risk issues 690t, 710t–711t  
risk assessment 690t,  
  692–697  
traits to modify 614, 667
- genetically modified  
  organisms (GMO) 752
- genetics 65b, 105–107  
genic balance model of sex  
  determination 751  
genome 108–110, 751
- analysis see genomics  
arthropod phylogenies  
  572–573  
databases 610t  
dilution 421  
multiple 127–149  
sequencing see sequencing  
Genome Annotation  
  Assessment Project  
  (GASP) 270  
genomic(s) 253, 738, 743,  
  750–751  
comparative 253, 294–295,  
  526–527  
*Drosophila* 268–272  
duplication 568  
footprinting 751  
functional 292–293  
imprinting 354, 751–752  
libraries see libraries,  
  genomic  
other arthropods 279–287  
quality of analysis 286b  
stress 377–378  
structural 293–294
- genotypes 752  
geographic variation 630–637  
geological time scale 570t  
geotaxis 475, 752  
germ cells 149–150, 752  
germ line 151, 424, 752  
  determination 430–431  
transformation see  
  transformation
- GFP 393, 468, 665t–667t, 669,  
  676, 702–703, 752  
gigabase pairs (Gbp) 752  
*Glossina* 280t  
  symbionts 148–149, 676  
*Glossina morsitans morsitans*  
  130, 442, 682t–683t  
*Glossina pallidipes* 452b  
*Glossina palpalis* 452b  
glucose dehydrogenase gene  
  497–498  
glucose-6-phosphate  
  dehydrogenase gene  
  632  
glue protein gene (sgs-4) of  
  *Drosophila* 18–19  
glutamic acid 18t
- glutamic acid decarboxylase  
  (GAD) 496  
glutamine 18t  
glutathione-S-transferase  
  682t–683t  
glycine 18t  
glycosylation 55, 752  
Golgi complex 38, 55, 752  
gooseberry gene 158f  
grape phylloxera 332–333  
grasshoppers see species  
  names  
green fluorescent protein see  
  GFP  
*groucho* gene 425t, 429  
Grylloblattodea 105–106, 525t  
*Gryllotalpa fossor* 65, 425–426  
*Gryllus bimaculatus* 665t–667t,  
  682t–683t  
GST see coefficient of gene  
  differentiation  
guanidinium hydrochloride  
  188–189, 321–322,  
  551–553  
guanidium thiocyanate 188–  
  189, 321–322, 551–553  
guanine 8, 10f, 15f, 737  
gynandromorph 439  
  see also sex determination  
gynogenesis 106  
gypsy moth see *Lymantria*  
  dispar  
gypsy-like elements see  
  transposable elements
- H**
- H1, H2A, H2B, H3, and H4 see  
  histone proteins  
*Habrobracon*, sex  
  determination 432  
*Habrobracon juglandis* 475  
Hae see restriction  
  endonucleases  
*Haemaphysalis longicornis*  
  673t  
*Haematobia irritans* 340f, 397,  
  682t–683t  
hairpin structure see DNA;  
  RNA  
Haldane's Rule 442  
halteres 547

- Hamiltonella defensa* 148, 503  
 haplo-diploidy 431, 436–437, 472t  
 haploidy 71–72, 87–89, 104–105, 752  
 haplotype 613–614, 752  
 Hardy-Weinberg equilibrium 609, 612, 752  
*Harmonia axyridis* 443t–444t  
*Harpegnathos saltator* 284  
 heat shock genes 18–19, 112–113  
 proteins 112  
 response 56  
*Helena* see transposable elements  
 helicases 21–22, 23f, 752  
*Heliconius* 280t  
*Helicoverpa armigera* 243, 478, 682t–683t  
*Helicoverpa zea* 398, 673t  
*Heliothis subflexa* 476  
*Heliothis virescens* 353–354, 476, 479  
*Helitrons* see transposable elements  
 helix 752  
 helix-turn-helix DNA-binding proteins 60  
 helper plasmids 382–383, 752  
 hemimetabolous insects 161  
 Hemiptera 81, 83, 105–106, 106–107, 120, 123, 129, 138, 161, 523, 682t–683t  
 hemoglobin 111, 528  
 Hennig 86 563  
 heritability ( $h^2$ ) 477, 752  
 broad-sense (VG/VP) 477, 740  
 narrow sense (VA/VP) 477  
*hermaphrodite* gene 425t, 429  
*Hermes* see transposable elements  
*hermit* see transposable elements  
*Herves* see transposable elements  
*HeT-A* see transposable elements  
 heterochromatin 79–81, 90, 753  
 centromeric 71–72  
 constitutive 535–536, 743  
 facultative 749  
 position effect variation 63–64, 387, 765  
     see also DNA  
 heteroduplex analysis (HDA) 600–601  
 heterogametic sex 94, 431, 753  
 heterogeneous nuclear ribonucleoproteins (hnRNPs) 54, 753  
 heterologous probes 537  
 heterologous recombination 753  
 heteroplasm 96, 541–542, 753  
 Heteroptera 444  
*Heterorhabditis bacteriophora* 396  
 heterosis 753  
     negative 760  
 heterozygosity 609, 753  
*hexamerin* gene 682t–683t  
 Hexapoda 525t, 753  
 highly-repetitive DNA see DNA, highly-repetitive  
*Hin* see restriction endonucleases  
 histidine 18t  
 histone(s) 61, 76, 150–151, 753  
     genes 113  
     octamer 76–77, 77f, 113  
     proteins 76  
 hitchhiking 541  
*hobo* see transposable elements  
 Hogness box see TATA box  
 Holocentric chromosomes  
     see chromosomes, holocentric  
 holometabolous insects 153, 161, 489, 571  
 HOM-C 163–164  
 homeo domain 60  
 homeobox 159, 266–267, 753  
 homeodomain proteins 159  
 homeosis 159  
 homeotic 154, 753  
     bithorax complex 154  
     genes 159–161, 753  
     mutants 154, 753  
 homing 377–378  
 homing endonucleases 375, 400, 401b, 405–406, 665t–667t  
*Homo sapiens* 270  
 homoduplex DNA 600, 753  
 homogametic sex 95, 754  
 homology 165, 266–267, 496–497, 531, 754  
 homoplasy 561–562, 754  
 Homoptera see Hemiptera  
 homosexual 498  
 homozygosity 754  
 honey bee see *Apis mellifera*  
 Hood, Leroy 296  
 horizontal gene transfer 95–96, 109–110, 134, 567, 662–663, 693–696, 754, 756  
 hormones 60, 427  
 host-plant preference  
     see behavior, host-preference  
 host-preference behavior  
     see behavior, host-preference  
 hotspots 90  
 house fly see *Musca domestica*  
 house-entering behavior of mosquitoes see behavior  
 housekeeping genes 44–45, 59, 528, 754  
*Hox* genes 159–160, 283  
*hsp* genes see heat shock genes  
 human disease genes 271  
 human diseases 503–506  
     see also individual diseases  
 Human Genome Project 230, 268–270  
*hunchback* gene 158  
 Huntington's disease 271  
*Hyalophora cecropia* 111, 121  
 hybrid breakdown 135  
 dysgenesis 374, 376–377, 754  
 sterility 135, 442  
 zones 624  
     see also *P* element  
 hybridization 621

- DNA 209, 236, 532, 755  
*in situ* 209, 535–536, 755  
 of *Magicicada* populations 621–622  
 probes 236, 754  
 screening 238–240  
*Hydra littoralis* 396  
*Hydra vulgaris* 396  
 hydrogen bonding 3–4, 13–14, 754  
 Hydrophilidae 115–116  
 2-hydroxyecdysone 117  
 hygienic behavior 472–473  
 Hymenoptera 83, 105–106, 106–107, 115–116, 120, 123, 129, 138, 161, 596–599, 610t, 682t–683t  
*Hyperkinetic* gene 475  
 hypertranscription 123, 426–427, 754  
 Hypervariable DNA sequences see DNA  
*Hypothenemus hampei* 110  
*Hypothenemus obscurus* 110
- I**
- Ichneumonidae 134, 139  
 IGS see ribosomal RNA, IGS  
 Illumina (Solexa) Genome Analyzer 274  
 imaginal discs 153, 754  
 IMMANC program 635–637  
 immune response 56, 113–115, 283  
   genes 113–115  
   modification 136, 137t, 140  
 immunochemical screening 238–240, 239f  
 importation of live organisms 595  
 imprinting 433t  
   genomic 64–66, 751–752  
   maternal 433t  
   parental 763  
*in silico* biology 754  
*in situ* hybridization 209, 535, 755  
*in vitro* packaging 221–222, 222f, 224, 225f, 226, 755  
 inactivated X chromosome 425–426  
 inadvertent selection 689  
 inbreeding 285, 432  
 inbreeding coefficient 754  
 indels 31, 546–547, 754  
 independent assortment 4–5, 33, 89, 420–421, 754, 766  
 industrial melanism 643  
 inheritance  
   epigenetic 748  
   extranuclear 95–98  
 initiation codon see codons, initiation  
 insecticide resistance 127, 645–646  
   amplified esterases 340–341  
 in biological control agents 645–646  
   detection 355–356  
   genes cloned 127  
   knockdown (kdr) 684–685, 685t  
   as selectable marker 197, 198f  
   symbionts conferring 131  
 insect(s)  
   behavior see behavior  
   collection of 594–595  
   development see development  
   diversity 594  
   DNA see DNA  
   fossil record 320t, 321–322, 347–350, 551–553  
   genetic systems 105–107  
   personality 501–502  
   plant interactions 479, 639–640  
   population origins 641–643  
   transgenic 665t–667t  
 insertion sequences 755  
 insertions see mutations, insertions  
 instantaneous speciation 622  
 insulators 63–64, 755  
 interactomes 295, 755  
 intercalating agents 755  
 interference see RNA, silencing  
 intergenic DNA see DNA, intergenic  
 intergenic regions see DNA, intergenic  
 interphase 26, 61–62, 76, 80, 755  
 intersex 427  
*intersex* gene 425t, 429  
 intracellular microorganisms see symbionts  
 intraspecific  
   reproductive incompatibility see cytoplasmic incompatibility  
   variability 434–436  
 introgression 755  
 introns 9b, 19, 20, 40–41, 63f, 111, 113, 546, 634, 755  
   splicing out 48  
 introns-early hypothesis 19–20  
 introns-late hypothesis 19–20  
 inversion see mutations  
 inverted repeats 266, 755  
 ion channels 493–496, 755  
 ion sequencing 276  
 Ion Torrent Personal Genome Machine 276–277  
 ionic selectivity 494, 755  
 insertion see mutations, insertions  
 isochores 277–278  
 isoleucine 18t  
 Isoptera 83, 129, 161, 682t–683t  
 isozymes 531–533, 755  
   see also allozymes  
 Israeli acute paralysis virus 700  
 ITS see ribosomal RNA, ITS  
*Ixodes* 107–108, 131–132, 138, 280t, 284–285, 350–351, 673t
- J**
- JcDNV gene vector 401  
 Jerdon's jumping ant see *Harpegnathos saltator*  
 jewel wasps see *Nasonia*  
 jockey see transposable elements  
*Juan* see transposable elements  
 jumpstarter 387f  
   see also P-elements, vectors

- junk DNA see DNA, junk  
*Junonia coenia* 401  
*Jurassic Park* 347  
juvenile hormone 117, 161–162
- K**  
karyotype 756  
*kete* gene 499  
kilobases (Kb) 16  
definition 756  
kilodaltons (dKa), definition 756  
kin selection 756  
kinship 594, 600, 603, 605, 609  
*Klebsiella* 144–145  
*Klebsiella oxytoca* 142–143  
see also symbionts  
Klenow fragment of DNA polymerase I 233, 235t, 241, 255, 313, 315, 324–325, 325, 756  
see also DNA polymerase  
*knirps* gene 158, 158f  
knockdown resistance (*kdr*) gene 684–685, 685t  
see also insecticide resistance  
Kpn see restriction endonucleases  
Krüppel gene 158, 158f
- L**  
L-form bacteria 445  
Labiata 524t  
*lacZ* gene 217, 226, 230, 261, 688–689  
as selectable marker 224  
lagging strand of DNA 4, 22–25, 756  
see also replication  
lambda see bacteriophages, lambda  
*Laodelphax striatellus* 679  
Large milkweed bug see *Oncopeltus*  
lariat 46f, 47  
larval serum genes 111–112  
Last Universal Common Ancestor (LUCA) 565
- Lateral gene transfer see Horizontal gene transfer  
*latheo* gene 487, 488t, 489–490  
*Laupala* 74  
law  
of Independent Assortment 89, 420–421, 756  
of Segregation 89, 420–421, 756  
leader sequences 45, 756  
leading strand of DNA 4, 22–25, 756  
see also DNA; replication  
leafhoppers 130, 354, 569, 684–685  
see also *Euscelidium*  
learning 468–469, 486–490  
olfactory 490–491  
*Leishmania* 355, 396–397, 674  
leishmaniasis 676–677  
*leonardo* gene 488t  
Lepidoptera 81, 83, 105–106, 106–107, 111, 115–116, 120, 123, 161, 682t–683t  
*Leptinotarsa decemlineata* 353–354, 646  
*Leptopilina boulardi* 134, 679  
*Leptopilina clavipes* 433t  
*Leptopilina heterotoma* 443t–444t  
lethal mutations see mutations, lethal  
leucine  
amino acid 18t  
zippers 60, 756  
Lewis, Edward B. 154, 159, 161  
libraries 756  
cDNA 217–218, 231  
genomic 217–230, 234–240, 752  
microsatellite 613t  
screening 234  
shotgun 219, 771  
ligases see DNA ligase  
ligation 23–25, 757  
*Ligia oceanica* 446  
likelihood methods 559, 757  
*Linepithema humile* 280t, 284
- linkage 751, 757  
linkage maps 757  
linker DNA 76–77, 757  
Linnaeus, Carl 523  
*linotte* gene 487, 488t, 489–490  
lipids 757  
*Lipolexis oregmae* 618  
locomotor activity see behavior  
locus 757  
*Locusta migratoria* 65, 116–117, 544  
long intergenic noncoding RNA (lincRNA) 757  
long interspersed elements (LINEs) 119–120  
long noncoding RNA (lncRNA) 39b, 757  
long PCR see PCR, long or high-fidelity  
long period interspersion pattern 757  
see also genome organization  
long terminal repeats 118–119, 545–546  
long-branch attraction 562, 757  
long-germ band development 151, 757  
longevity 703–704  
Lou Gehrig's disease 504  
love songs see behavior, song cycle in *Drosophila*  
luciferase 273–274  
*Lucilia cuprina* 75, 82, 381, 397, 453, 665t–667t  
*Lucilia sericata* 665t–667t  
*Lymantria dispar* 353–354  
nuclear polyhedrosis virus (NPV) 242–243, 502–503, 677–678  
lysine 76  
*Lysiphlebus fabarum* 737  
*Lysiphlebus testaceipes* 618  
lysis 320, 757  
lysogenic stage 219–220, 757  
lysozymes 114, 204t, 757  
lytic stage 219, 220, 220f, 757

**M**

M phase *see* cell cycle, M phase  
**M13** bacteriophage 216–217, 228–230, 229f, 759  
 universal primer 759  
*see also* bacteriophages  
**MacClade** 563  
**MAFFT** 563  
*Magicicada* 620–624, 623f  
*Magicicada neotredecim* 621–622, 624  
*Magicicada septemdecim* 620–621  
*Magicicada tredecim* 620–621  
*mago nashi* gene 430  
 major groove of DNA 14, 757  
 malaria 646–647, 647, 665t–667t  
 male  
   feminization 433t  
   gamete loss 431  
   sterility 381  
 male drive (MD) 440  
 male killing 132–133, 136, 137b, 148, 433t, 443t–444t, 445, 446–447, 449, 450–451  
*male-specific lethal* genes 425t, 426–427, 429  
**Mallophaga** 81, 129, 161  
 mammalian cancer gene 271  
 mammals 57–58, 65, 74, 122, 393, 399  
*Manduca sexta* 111, 116–117, 681, 683  
 gene regulation 60  
*Maniola jurtina* 442  
**Mantodea** 161  
**Mantophasmatodea** 525t  
 map unit 757  
 mapping with RFLP markers  
   *see* restriction fragment length polymorphisms  
*mariner* *see* transposable elements  
 markers 758  
 Markov process 560  
*Mastotermes electrodominicus* 348b

maternal effect genes 150, 155–157, 428–429, 442, 758  
 maternal inheritance 98, 540–541, 758  
 maternal microinjection 673t  
 maternal sex ratio (MSR) 437, 441, 450  
*see also* sex determination  
 mating behavior *see* behavior  
*Maverick* transposon 122  
 Maxam-Gilbert method  
   of DNA sequencing (chemical method) 252, 254, 261–262, 262f, 758  
*maxillopedia* gene 163–164  
 maximum likelihood methods 559, 562  
 maximum parsimony methods 559, 562, 758  
 Mayr, Ernst 565  
 mealybugs *see species names*  
*Mecoptera* 83, 105–106, 123, 161  
*Medea* genes 442, 665t–667t  
*Medfly* *see* *Ceratitis capitata*  
 median melting temperature 758  
*Mediterranean fruit fly* *see* *Ceratitis capitata*  
*MEGA* 758  
 megabase pairs (Mbp) 758  
*Megaloptera* 105–106, 161  
 meganucleases 375, 400, 401b, 405–406  
*Megaselia scalaris* 109, 436  
*Megoura criassicauda* 148  
*Megoura viciae* 65  
 meiosis 3–4, 33, 150–151, 758  
   in eukaryotes 87–90, 88f, 447–448  
     meiosis I 89  
     meiosis II 89  
 meiotic drive 440–442, 758  
*Melipona quadrifasciata* 432–433  
 Mello, Craig 56–57, 401–402  
 membrane 758  
 memory 468–469, 487, 490  
 Mendel, Gregor 4–5

Mendelian inheritance 4–6, 758  
 Mendelian population 758  
 Mendel's Law of Independent Assortment 89, 420–421, 756  
 Mendel's Law of Segregation 89, 420–421, 756  
 messenger RNA (mRNA) 758  
   antisense 738  
   cloning 233–234  
   in eukaryotes  
   function 39b  
   isolation 233–234  
   precursor *see* pre-mRNA  
   processing of transcript to mRNA 6–8, 7f  
   sequencing 289  
   stability 55–56  
   surveillance 52–53  
 metabolic pathways 758  
 metabolomics 645–646, 758  
 metagenomics 290–291, 758  
 metamorphosis 153, 161–162, 489  
 metaphase 86, 759  
*Metaseiulus occidentalis* 121, 138b, 280t, 354, 673t, 698f, 701b–702b  
*Methanobacterium thermoautotrophicum* 315  
 methionine 18t, 334–335  
 methotrexate 127  
 methylation of DNA *see* DNA, methylation  
 MIAME 246  
 microarrays 216, 243–247, 592–593, 601–602, 602f, 610t, 639–640, 741, 745  
   tiling 285–287, 290, 775  
 microgram, definition 759  
 microinjection 673t  
 micron, definition 759  
 microorganisms *see* Bacteria; symbionts  
*micropia* *see* transposable elements  
 micropyle 151

- microRNAs 39*b*, 56–58, 547–548, 759  
 miR-34 272  
 small hairpin (shmiRNA) 680–681
- microsatellite DNA see DNA, microsatellite
- middle-repetitive DNA see DNA, middle-repetitive
- minichromosomes 109
- minigenes 19–20
- minipreps 204*t*
- minisatellite DNA see DNA, minisatellite
- minor groove of DNA 14, 759
- Minos* see transposable elements
- mirtrons 759
- mismatch repair 29
- missense mutations see mutations, missense
- mitochondria 95, 127–128, 539–544, 759
- mitochondrial DNA (mtDNA) 96, 536, 539–544, 595, 610*t*, 760  
*A. mellifera* 540*f*  
*D. melanogaster* 96  
*D. yakuba* 97*f*  
 molecular evolution 98  
 mutations in 539  
 paternal transfer of 96  
 PCR primers 597*t*–598*t*  
 structure 540–541
- mitochondrial genes 95–98
- mitosis 3–4, 25–27, 84–87, 85*f*, 741, 759  
 see also individual stages
- mitotic spindle 759
- mobile genetic elements 759
- mode of inheritance 343*f*, 391, 432, 633
- model organisms 504–505, 696, 759
- modENCODE Consortium 279
- moderately repetitive DNA 759
- molecular biology 760  
 clock 528, 541, 760
- cytology 535–536  
 data analysis methods 531–538  
 ecology 591  
 evolution 526–527, 573–578, 760  
 genetic manipulation see transgenic insects  
 genetics 184–187, 465, 591, 668–672, 760  
 phylogeny 760  
 systematics 521, 760  
 traits 527–528
- molecular analysis of parasitoid linkages (MAPL) 617–618
- molecular domestication 287–288
- Mollusca 57
- molting 60
- monarch butterfly see *Danaus plexippus*
- monoclonal antibodies 760
- monophyly  
 of arthropods 548  
 of Chelicerata 548  
 of *D. pseudoobscura* 572  
 of Mandibulata 548
- Morinda citrifolia* 479
- morphogens 155, 760
- morphological traits 527–528
- mosaic analysis 427, 449
- mosquitoes  
 transgenic 707*b*–709*b*  
 see also *Aedes*; *Anopheles*; *Culex*
- MrBayes software 563
- mRNA see messenger RNA
- mtDNA see mitochondrial DNA
- Muller's ratchet 422, 760
- multigene families 18–19, 760
- multiple allele model of sex determination 432
- multiple cloning sites 197–198, 199*f*
- multiple displacement amplification (MDA) 359, 760
- multiple genomes 127–149
- multiple-locus, multiple-allele model 433, 760
- Mus musculus* 268
- Musca domestica* 673*t*
- DNA methylation 65
- genes cloned from 353
- mating behavior 479
- repetitive DNA 75
- sex determination 434–435, 435*f*, 439
- sex ratio distorters 442
- short-period interspersed pattern genome 75
- transposable elements in 381, 395
- Muscidae* 395
- Muscidifurax uniraptor* 135, 447–448
- MUSCLE 563
- muscle of Lawrence 429, 498
- mushroom bodies 469–470, 470*f*, 489, 491
- mutagens 29, 760
- mutant phenotype 29, 760
- mutations 760  
 back 30*t*, 31–32, 739  
 conditional lethal 30*t*, 31, 743  
 deletions 30*t*, 31, 92, 744  
 effects of 30*t*  
 favorable 422  
 frameshift 30*t*, 31, 749
- genome 29–32  
 homeotic 160  
 insertions 30*t*, 31, 92, 755  
 inversions 31, 755  
 lethal 30*t*, 31, 756  
 load 422  
 missense (nonsynonymous) 30*t*, 31  
 in mitochondria 539  
 neutral 530–531  
 nonsense 30*t*, 31, 761  
 point 30–31, 30*t*, 764  
 preadaptive 127  
 regulatory 30*t*, 31–32  
 repair 28–29  
 reversions 30*t*, 31–32  
 silent (synonymous) 30*t*, 31, 771

- suppressions 30t  
 temperature-sensitive 475  
 transitions 30–31, 30t, 776  
 transversions 30t, 776  
 X chromosome 482
- mutator** 387f  
*see also P elements, vectors*
- Mycocetus smithii* 423–424
- mycoplasmas 130
- Mylothris spica* 442
- Myriapoda* 524t
- Myrmeleontidae* 132
- Myzus persicae* 65, 109, 127, 682t–683t
- Myzus persicae nicotianae* 642
- N**
- NAD 195, 235t  
 NADH dehydrogenase 97f, 540f
- nanogram, definition 760
- nanometer, definition 760
- nanopore sequencing 277
- nanos* gene 156–157
- nanotechnology 406–407, 673t
- Nasonia* 107–108, 136, 162–163, 468, 471–472  
 and B chromosomes 94, 110–111
- Nasonia giraulti* 280t, 450, 475
- Nasonia longicornis* 280t, 450
- Nasonia vitripennis* 65–66, 109–110, 133–134, 244, 280t, 282, 432, 433t, 443t–444t, 450, 475, 682t–683t
- Nasutitermes corniger* 600
- NCBI, National Center for Biotechnology Information 553–556
- nearly neutral theory of evolution 530–531, 760
- negative heterosis 760
- Nei's standard genetic distance 611, 636–637
- Nemomychidae 348b
- Neochrysocharis formosa* 433t, 443t–444t, 446
- Neodiprion nigroscutum* 432–433
- Neonympha mitchellii mitchellii* 594–595
- Neoptera 161
- Nephila clavipes* 283
- nervous system 469–471, 493–496
- neurobiochemistry 493–496
- neurodegenerative diseases in *Drosophila* and humans 503–506
- neuropeptides 470–471, 760
- neuropil 469
- Neoptera 115–116
- neurotransmitters 495
- neutral theory of molecular evolution 760
- NextGen sequencing 253, 272–277, 273t, 348b, 537–538, 761
- Ngbo* gene 499
- nick, definition 761
- nick translation 209–210, 209f, 240, 761
- nicotinamide adenine dinucleotide *see* NAD
- nitrogenous bases 761
- 10-nm chromatin fiber 774
- 30-nm fiber 78–79, 775  
*see also* chromosomes
- Non-preference* gene 475
- noncoding DNA *see* DNA, noncoding
- noncoding RNA (ncRNA) 38, 40, 761
- noncoding strand 761
- nonhomologous recombination 680
- nonsense mutations *see* mutations, nonsense
- Northern blot analysis 183–184, 206, 240, 761
- Nosema apis* 340–341, 355
- Nosema ceranae* 340–341, 355
- NPV *see* nuclear polyhedrosis virus
- nuclear DNA *see* DNA
- envelope 48, 53, 79, 82, 761
- genome 761
- histone acetylating enzymes 61
- membrane 53
- organization 53
- transplantation 673t
- nuclear mitochondrial pseudogenes *see* numts
- nuclear polyhedrosis virus (NPV) 242–243, 502–503, 677–678
- nuclear pore complex 53–54, 761
- nuclear transplantation 673t
- nucleases
- meganucleases 375, 400, 401b, 405–406
- TALENs 375, 400, 401b, 404–405
- zinc-finger* *see* zinc-finger nucleases
- nucleic acids 761
- delivery into arthropod tissues 685–686
- detection 206
- hybridization 240–242, 761
- precipitation 190–191
- nucleoli 79, 761
- nucleoporins 53–54
- nucleosides 8, 11f, 761
- nucleoskeleton 73–74
- nucleosomes 71–72, 76–78, 761
- nucleotide-excision repair 29
- nucleotides (nt) 8, 762
- nucleus 53–54, 79, 762
- null allele 762
- numerator genes 427
- numts 523, 544, 607, 762
- nurse cells 149–150
- Nusslein-Volhard, Christiane 154
- O**
- obligate symbiosis 129
- odd-paired* gene 158, 158f
- odd-skipped* gene 158f
- Odonata 105–106, 120, 161
- odorant binding protein 493, 762
- Odysseus* gene 576–577
- 3'-OH terminus 10

- Okazaki fragments 4, 22–25, 762  
 oligonucleotides 762  
*Oncopeltus fasciatus* 443t–444t, 476–477, 682t–683t  
 onionskin structure 123–125  
     see also Gene amplification  
*Onychophora* 548  
 oocytes 762  
     formation 149–151  
*Ooencyrtus fecundus* 447  
*Ooencyrtus submetallicus* 447  
 oogenesis 430  
 open reading frames (ORFs) 254, 266, 553, 762  
*Opomyzidae* 381  
 Ordovician period 570t  
 organelles 98, 127–128, 131, 762  
 oriental fruit fly see *Bactrocera dorsalis*  
 origin of replication see replication, origins  
 Origin of Replication Complex (ORC) 26  
 orphan genes see genes, orphan  
 orthologs 162–163, 561–562, 762  
*Orthoptera* 83, 105–106, 106–107, 120, 123, 129, 682t–683t  
*oskar* gene 156–157, 430  
*Ostrinia nubialis* 475–476  
*Ostrinia scapularis* 443t–444t, 449  
 ovarian tumor gene 430–431  
 overhangs 193, 193t, 195  
 ovo gene 430–431
- P**
- P* cytotype 390–391  
*P* elements 373, 479–480, 762  
     cross mobilization of 398  
     as driver of genes into populations 121, 392  
     horizontal transfer of 374–375, 380–381, 392  
     hybrid dysgenesis 374, 376–377  
     mediated transformation 375, 381, 382f, 393–394, 672  
     of *Drosophila* with a *Bombyx* silk gene 116  
     of *Drosophila* with Est-5 496–497  
     of insects other than *Drosophila* 390, 395–398  
     origin 374–375, 379–381  
     resistance to 390–392  
     structure 376f, 377–378  
     transposition method 378–379  
     vectors 374–375, 381–386, 382f, 386f, 387f, 752  
     see also transposable elements  
*p25.7wc* see *P*-element, vectors  
*P450* gene 665t–667t, 682t–683t  
 pair-rule genes 157, 159, 762  
     see also development  
*paired* gene 157, 158, 158f  
 Paleozoic era 569, 570t  
 palindromic sequences 318, 762  
*PALM* 563  
*Palorus ratzeburgii* 82  
*PAML* 4 563  
*PAML* 762  
*Pao* elements see transposable elements  
*para* gene 353, 475  
 parahaploidy 104–105, 107, 431  
 paralogs 561–562, 762  
 parapatric species 500  
 parasegments 762  
 parasitic DNA see DNA; transposable elements  
 parasitoid wasps 139–141, 617–618  
 parasitoids  
     genetic improvement 453–454  
     see also individual species  
 parathion hydrolase gene (*opd*) 393  
     see also insecticide resistance  
 paratransgenesis 662, 663, 665t–667t, 675–677, 677t, 695–696, 762  
 parental imprinting 763  
*Paris* see transposable elements  
 Parkinson's disease 271, 504  
 parsimony methods 556, 559, 562, 763  
 parthenogenesis 105–106, 134, 137b, 423–424, 433t, 664–667  
*Parvoviridae* 678  
 paternal  
     derived chromosomes  
     see chromosomes, paternally derived  
     leakage 541–542  
     see also sex determination  
 paternal sex ratio (PSR) 94, 450, 763  
 paternity 592–593  
 pathogens 763  
     detection 354–355  
 Pauling, Linus 5–6  
 PAUP 268, 563  
*Pauridia peregrina* 447  
*Pauropoda* 524t  
*pBR322* see plasmids, *pBR322*  
*pcosEMBL* see vectors  
 PCR see polymerase chain reaction  
 pea aphid see *Acyrthosiphon pisum*  
*Pectinophora gossypiella* 395, 673t, 702b–703b  
*Pediculus humanus humanus* 107–108, 280t, 282–283  
 PEG see polyethylene glycol  
*Penelope* see transposable elements  
 7-pentacosene 499  
 peppered moth see *Biston betularia*  
 peptide bonds 49, 763  
 period gene 466, 480t, 481, 482–484, 483f, 485, 486, 504–505  
     see also behavior  
 periodical cicadas see *Magicicada*

- Periplaneta americana* 142, 353–354, 469–470  
 peritrophic membrane 113–114  
 Permian period 569, 570t  
 personality 501–502  
 pest insects 667–668  
 pest management strategies 441, 451–453, 614–620, 625–639, 643–647, 661  
 pesticide resistance see insecticide resistance  
*Pfu* DNA polymerase 336, 352  
 pH 763  
 phage M13 see bacteriophages, M13  
 phagemids 216–217, 230, 763  
*phase-angle* gene 484  
 Phasmida 161  
*Pheidole pallidula* 508  
 phenetic systematics 526, 550, 574t, 763  
 phenograms 551, 763  
 phenol DNA extraction method 187, 188t, 211, 322  
 phenotype 29, 502–503, 565, 763  
 phenylalanine 18t  
 pheromone binding protein 763 biosynthesis-activating neuropeptide (PBAN) 492 communication 475–476, 492–493  
*Phialophora* 144  
*Philodoria potatoria* 442  
*Phlebotomus* 676–677  
*Phormia regina* 479, 487  
 phosphodiester bonds 10, 763  
 phosphoric acid group 8  
 phosphorylation 763  
 photoperiodic clock 480–486  
 photoreceptor 486  
 phototaxis 475, 764  
 phyletic speciation 764  
 PHYLIP 563  
*Phyllocoptis citrella* 577–578  
 PhyloCode 526
- Phylogenetic Analysis Using Parsimony see PAUP  
 phylogenetic 764 methods 549–564 software 562–564 species 574–575 systematics 522–523, 526–527, 538–539 trees 549, 551, 557, 764 phylogenies 522–523, 526–527, 764 construction of 556–561 Phylogeny Inference Package see PHYLIP phylogeography 634 phylogenograms 551 physical map of the genome 764 Phytoseiidae 138, 189 picogram, definition 764 piggyBac see transposable elements piled egg gene 475 pink bollworm see *Pectinophora gossypiella* *Pissodes* 97–98 *Pissodes remorensis* 97–98 *Pissodes strobi* 97–98 *Pissodes termonalis* 97–98 piwi-interacting RNA (piRNA) 39b, 377, 764 PJB8 see vectors *Planococcus citri* 64, 131 *Planococcus lilacius* 65 plants defenses to herbivory 639–640 RNAi in 684–685 plaques 223, 223f, 764 hybridization 223, 764 screening 236–237, 236f, 238–240, 764 plasmids 184, 764 *E. coli* cloning 196–201 *E. coli* transformation 201–202 helper 382–383, 752 pBR322 200, 217 preparations see Minipreps promiscuous 197 pUC 199f purification 197–198, 200, 202–204 as vectors 197–198 wild-type 197 see also cosmids; phagemids Plecoptera 105–106, 161 pleiotropism 764 plesiomorphic 550 *Plodia interpunctella* 401 *Plutella xylostella* 353–354, 682t–683t *Pogonomyrmex barbatus* 280t, 284 point mutations see mutations, point polarity anterior-posterior 154 dorso-ventral 155, 165 see also development pole cells 149–150, 383–384, 430, 764 Polintons 267 polonies 272–273 poly(A) signal 277–278 tail 47, 55–56, 352, 764 poly(A) polymerase 47 polyacrylamide gel electrophoresis 205–206 polydnnaviruses 139, 764 polyethylene glycol 201–202, 325 polygenic inheritance 478–479 polygyny 508 polyhedrin promoter 243 polylinkers 197–198, 217, 230, 765 see also multiple cloning sites polymerase 765 polymerase chain reaction (PCR) 98, 217–218, 307, 765 AFLP 322, 329–333, 331, 595–596, 597t–598t, 599, 615–616, 629 allele-specific 310f–311f, 314f, 317, 330t–331t, 596, 597t–598t

- polymerase chain reaction (PCR) (Continued)  
     amplification of DNA 314f, 354  
     amplification plateau 327  
     anchored 330t–331t, 333, 738  
     ancient DNA 321–322, 347–350  
     annealing 316t–317t, 325  
     applications 347–359  
     arbitrary primers (AP-PCR) 330t–331t, 333–334, 738  
     asymmetric 320t, 334, 738  
     automation 324  
     basic 312–329  
     bridge 274  
     buffers 316t–317t, 740  
     contamination 327–329, 360b–361b  
     cycle number 316t–317t, 326–327, 327t  
     cycle sequencing 332t  
     DALP 334  
     degenerate primers 330t–331t, 334–335, 353  
     denaturing DNA for 312–313  
     detecting pathogens in arthropod vectors 354–355  
     digital 346–347  
     direct sequencing 358–359  
     disease control evaluation 357  
     DNA extraction 189, 322  
     DNA preparation 312, 319–323  
     emulsion 273–274  
     engineering DNA 356–357, 356f  
     false negatives 320–321, 326, 344–345  
     false positives 329, 344–345  
     fluorescence-detecting thermocyclers 345  
     hot-start 335, 754  
     inhibition 322–323  
     insects in amber 347, 348b  
     inverse 330t–331t, 335, 337f, 755  
     jumping 349  
     long or high-fidelity 336–338, 596, 597t–598t  
     modifications 329–347  
     multiplex 330t–331t, 339–341, 340f, 760  
     negative controls 329  
     nested 341  
     old DNA 350–351  
     optimization 340–341  
     PCR-RFLP 330t–331t, 341, 597t–598t, 604–605, 632, 763  
     positive controls 328–329  
     power of 313  
     primer artifacts 326  
     primers 312, 315–318  
     protocols 313, 314t, 332t  
     quantitative (qPCR) 342, 767  
     RAPD 189, 309, 330t–331t, 342–345, 592–593, 595–596, 597t–598t, 610t, 767  
     applications 343, 534t, 538  
     data analysis 613  
     genetic variation  
         estimation 343–344  
         method 342–343, 343f  
     phylogeny analysis 605–606, 606f  
     species identification 343–344, 577–578  
     real-time 345, 768  
     reverse transcription (RT-PCR) 346, 546, 770  
     RNA amplification 330t–331t, 351–352  
     RNA quantitation 330t–331t  
     sequencing 263–264, 263f, 358–359, 597t–598t  
     specificity 324–326  
     SSP 332t  
     storing insects for 318–319  
     SWAPP 345  
     TaqMan 345–346, 774  
     tDNA-PCR 774  
     temperature cyclers 313, 320, 344  
     template DNA 316t–317t, 319  
     universal primers 357  
         see also primers  
     polymers 765  
     polymorphism 609, 746, 765  
         single nucleotide 467, 608, 610t, 772  
     polynucleotide kinase 234, 235t  
     polynucleotides 10–11, 765  
     polypeptides 242–243, 765  
     polyphyly 572  
     polyploidy 73, 105–106, 108, 111, 150, 765  
     polyribosomes (polysomes) 50–51, 51f, 765  
     polyteny 71–72, 91f, 92–93, 108, 765  
     population ecology and genetics 591  
     population isolation 620–624  
*Porcellionides pruinosus* 446  
     position effect variation 63–64, 387, 765  
     postbithorax gene 160f  
     postembryonic development see development  
     postreplication repair 29  
     posttranscriptional control 765  
     posttranscriptional processing 766  
     postzygotic reproductive isolating factors 573  
     postzygotic sex determination 437  
     potassium channels see channels, potassium  
     pre-adaptive mutations see mutations, pre-adaptive  
     *pre-intermolt* gene 488–489  
     pre-mRNA 9b, 20–21, 39b, 40–41, 45, 46f, 54, 766  
     precipitating of nucleic acids see nucleic acids, precipitation  
     predatory mite see *Metaseiulus*; *Phytoseiidae*  
     preening behavior 478–479  
     prey species 618–620

- prezygotic reproductive isolating factors 573
- primary transcript 62, 766
- primase 23, 26–27, 766
- primer-dimer artifacts 326, 766
- primers 21–22, 766  
arbitrary for PCR (AP-PCR) 330t–331t, 333–334, 738  
artifacts of PCR 326, 766  
degenerate for PCR 330t–331t, 334–335, 353  
for DNA replication 23, 24f  
extension 241  
G/C clamp 318  
for PCR 312, 315–318,  
330t–331t, 597t–598t  
random 241, 342–345  
SCAR 344  
universal for PCR 357
- prions 7, 766
- probes 209–210, 234, 537, 766  
heterologous 537  
labeling 240–242
- proboscipedia* gene 163f
- Procecidochares* 475
- Proctolaelaps regalis* 380–381
- Proctotrupidae 134
- programmed cell death 56
- prokaryotes 20, 67, 766
- proline 18t
- prometaphase 86
- promoters 41, 44–45, 62, 63f,  
766
- proofreading 315, 325, 766
- prophage 220
- prophase 26, 85–86, 447–448,  
766
- protease 56, 766
- Protein Information Resource (PIR) 292
- proteins 51–52, 766  
domain 746  
electrophoresis 533–535,  
597t–598t, 599  
families 294  
kinases 86–87, 474–475  
sequences 51–52  
stress 773  
structure 51–52  
synthesis 38, 48–52
- transcriptional activator 60,  
775  
transport 54–55
- Proteobacteria 95, 539,  
565–567
- proteome 766
- proteomics 253, 291–292, 766
- proteosome 56, 767
- prothoraxless* gene 163–164
- protocerebrum 469
- Protura 105–106, 161
- Prusiner, Stanley 7
- pseudoarrhenotoky 431
- Pseudococcidae 107, 131
- Pseudococcus calceolariae* 65
- Pseudococcus obscurus* 65
- pseudogamy 106
- pseudogenes 19, 111, 767
- Pseudonocardia* 144, 145f
- Psocoptera 115–116, 161
- PSR see sex ratio, paternal
- Pst see restriction endonucleases
- Pteromalidae 134
- pUC see plasmids
- puffing of chromosomes see chromosomes, puffing
- pulsed field electrophoresis see electrophoresis
- purines 8, 10f, 767
- Pycnogonida* 524t
- Pyrgomorpha conica* 65
- pyrimidines 8, 10f, 767
- Pyrococcus* 315
- Pyrococcus furiosus* 235t
- pyrosequencing 273–274, 767
- Pyrrhocoris apterus* 501–502
- Q**
- Q-banding 535–536, 767
- quality control 246
- quantitative genetics 767
- quantitative trait loci (QTL)  
471, 476, 506–507, 767
- R**
- R1 and R2 transposable element see transposable elements
- rabbit hemorrhagic disease virus 346
- radiolabeling 240–241, 767
- radish gene 487, 488t, 489–490
- random fragment length polymorphisms see restriction fragment length polymorphisms
- random primers see primers
- RAPD-PCR see polymerase chain reaction
- reactomes see interactomes
- reading frame 768
- recessive traits 4–5, 768  
see also genetic, terminology
- reciprocal cross 377, 768
- recombination 33, 420–421,  
768
- heterologous 753  
see also plasmids; recombination
- red flour beetle see *Tribolium castaneum*
- red harvester ant see *Pogonomyrmex barbatus*
- reductional division 89–90
- redundancy 768
- refractory 130
- Regiella insecticola* 148
- regulatory mutations see mutations sequences 768 signals 686–687
- reinforcement 485–486, 768
- relatedness 531–532, 534t,  
536
- release of insects carrying a dominant lethal (RIDL)  
665t–667t, 707b–709b
- Remebe 700
- Repbase 267
- repeated sequences see DNA, repetitive
- repetitive DNA see DNA, repetitive
- replacement vectors see vectors, replacement
- replica plating 200–201, 768

- replication  
 directional 22  
 DNA 4, 21–22, 25–27, 746  
 fidelity 28–29  
 fork 21, 768  
 ligation of DNA in 23–25  
 origins 21–22, 26, 200, 687, 762, 768  
 rolling circle 220  
 semi-conservative 9*b*, 21, 22*f*, 771
- reporter genes see genes, reporter
- reproductive incompatibility 360*b*–361*b*, 689–691
- reproductive isolation 135, 573, 699
- resistance see insecticide resistance
- response to selection (R) 477
- restriction endonucleases 184, 192–195, 193*t*, 224, 235*t*, 322, 353, 768  
 blunt ends 192–193, 195, 196, 196*t*  
 digest 194–195  
 maps 212  
 overhangs 193, 193*t*, 195
- restriction sites 353, 769  
 mapping 211–212, 769
- restriction fragment length polymorphisms (RFLP) 534*t*, 536–537, 595–596, 597*t*–598*t*, 604, 613–614, 768
- Reticulitermes flavipes* 141, 682*t*–683*t*
- retroelements 377–378, 696, 769
- retroposition 769
- retroposon integrations 546–547
- retrosequences 769
- retrotransposable elements  
*micropia* 120  
 see also transposable elements
- retrotransposons 267, 545–546, 769
- retrovirus-like 118*t*
- retroviruses 678, 769
- reverse genetics 769
- reverse transcriptase 231, 232–233, 234, 235*t*, 241, 769
- reverse transcription 111, 231, 346, 769
- reversions 622, 769  
 see also mutations
- RFLP-PCR see polymerase chain reaction, PCR-RFLP
- RFLPs see restriction fragment length polymorphisms
- Rhagoletis completa* 676
- Rhagoletis pomonella* 142–143, 143*f*, 573–574, 577, 599
- Rhodnius prolixus* 121–122, 280*t*, 284–285, 675, 682*t*–683*t*
- Rhyniella praecursor* 570*t*
- ribonuclease (RNase) 233, 769
- ribonucleic acid see RNA
- ribose 9*f*
- ribosomal RNA (rRNA) 16, 79, 544–546, 545*f*, 610*t*, 769  
 5S 42, 79  
 5.8S 79  
 16S 529  
 18S 79, 544, 547  
 28S 79, 544–545  
 ETS 544–545  
 function 39*b*  
 genes 115  
 IGS 544–545  
 ITS 544–545, 629–630
- ribosomes 38, 48  
 copy number 544  
 function 38  
 large subunit (28S) 79, 115, 120  
 small subunit (18S) 79, 115
- ribozyme 49, 769
- rice weevil see *Sitophilus oryzae*
- Rickettsia* 95, 128, 131–132, 355  
 modification of sex  
 determination 420, 433*t*, 443*t*–444*t*, 443*t*–444*t*, 446
- see also symbionts
- Rickettsiella* 148
- ring chromosomes 769
- Riptortus pedestris* 131, 485
- RISC protein complex 57, 402
- risk analysis for transgenic arthropods 692–697
- RNA 3–4, 8, 9*f*, 769  
 amplification 330*t*–331*t*, 351–352  
 antisense 57  
 double-stranded (dsRNA) 57, 402, 673*t*, 684  
 editing 62, 769  
 functions 39*b*  
 leader sequence 45, 63*f*  
 long intergenic noncoding (lncRNA) 757  
 long noncoding (lncRNA) 39*b*, 757
- messenger (mRNA) see messenger RNA
- noncoding (ncRNA) 38, 40, 761
- piwi-interacting (piRNA) 39*b*, 117, 764
- primer 23
- probes 209–210, 242
- processing 770
- ribosomal (rRNA) see ribosomal RNA
- sequencing 534*t*
- silencing (interference; RNAi) 40, 56–58, 375, 400*b*, 401–403, 485, 662, 665*t*–667*t*, 685*t*, 769  
 crop plants 684–685  
 pest control 680–685, 682*t*–683*t*
- small hairpin (shRNA) 680–681
- small interfering (siRNA) 39*b*, 684, 772
- small nuclear (snRNA) 16, 54, 772
- small nucleolar (snoRNA) 16, 772
- splicing 48, 770
- structure 11–13
- surveillance 52–53, 770

- synthesis 41–42  
transcription 770  
transfer (tRNA) see transfer RNA  
RNA polymerase 24*f*, 41–42, 43*f*, 770  
DNA dependent 746  
type II 42, 43*f*  
type III 42  
RNA-Seq analysis 289, 770  
RNase H 233  
Roche GS FLX sequencer 273–274  
rolling circle replication 220  
rooted tree 549, 550*f*, 557–558  
*rosy* gene 386*f*  
rough endoplasmic reticulum 55, 770  
*roughest* gene 109  
rovers 474  
rRNA see ribosomal RNA  
rTth DNA polymerase 336  
*runt* gene 158, 158*f*, 425*t*, 427  
*rutabaga* gene 487, 488*t*, 489
- S**  
S phase of cell cycle 26, 84–85, 87, 770  
S1 nuclease 234, 235*t*, 770  
*Saccharomyces cerevisiae* 72–73, 216–217, 268, 270–271, 568, 680  
salivary  
  secretions 132  
  toxins 132  
  glue 18–19  
  genes 488–489  
salivary gland chromosomes  
  see chromosomes, salivary gland  
sand flies see *Phlebotomus*  
Sanger method of DNA sequencing see dideoxy method  
*sans fille* gene 430–431  
*Sarcophaga bullata* 65, 75  
satellite DNA see DNA, satellite  
sawfly see *Athalia*  
*Scaptomyza pallida* 379  
*Schistocerca gregaria* 73  
*Schizaphis graminum* 75  
*Sciara coprophila* 65, 115, 544  
Sciaridae 92  
screwworm see *Cochliomyia hominivorax*  
SDS see sodium dodecyl sulfate  
secondary structure 770  
secondary transposition see transposition, secondary  
segment-polarity genes 157, 158*f*, 159, 770  
segmentation 153–154, 156*f*  
segmentation genes 157–161, 770  
segregation 4–5, 89, 420–421, 766  
Segregation Distorter (SD) 440  
selectable markers 197, 670*t*–671*t*, 688, 770  
selection  
  artificial 738  
  differential (S) 477, 771  
  experiments 476–478  
  response to 477  
selfish DNA see DNA, selfish  
semi-conservative replication  
  see replication, semi-conservative  
sense strand 16  
sensory transduction 494, 771  
SEQBOOT program 636–637  
Sequenase 255, 257*t*, 358  
sequencing 237–238, 251, 746, 771  
  data analysis 265–267, 614  
  data banks 267–268  
  direct 358–359  
  first-generation 252, 254  
    dideoxy method 252, 253, 254, 255–261, 255*f*, 256*f*, 264–265, 289, 537–538, 554*f*–555*f*, 739, 745  
    Maxam-Gilbert method 252, 254, 261–262, 262*f*, 758  
  gels 259*f*  
  kits 255  
  in molecular ecology 606–608  
  in molecular systematics 534*t*, 537–538, 551–553, 552*f*  
next-generation (NextGen) 253, 272–277, 273*t*, 348*b*, 537–538, 761  
PCR 263–264, 263*f*, 358–359  
standards 275–276  
third generation 276–277, 775  
whole-genome shotgun 262–263, 265–266, 777  
sericin 116, 771  
serine 18*t*  
*Serratia symbiotica* 148, 289  
sex  
  origin of 423–424  
  pheromone communication 475–476, 492–493  
  see also pheromones  
sex chromosomes see chromosomes, sex  
sex combs reduced gene 681*f*  
sex determination 417, 751  
  model 437–440  
sex hormones 427  
sex ratio  
  distorters 441–450, 744  
  female-biased 441  
  maternal 437, 441  
  paternal 94, 450, 763  
*Sex-lethal* gene 47, 420, 425*t*, 426–427, 428–429, 434–435, 436, 437  
sexual isolation 136, 476  
sexual reproduction 33, 420  
  costs and benefits 421–424  
*Shab* gene 493, 495  
*Shaker* gene 475, 493, 495  
*Shal* gene 493, 495  
*Shaw* gene 493, 495  
sheep blowfly see *Lucilia cuprina*  
short inverted terminal repeats see repeated sequences  
short period interspersion pattern of genome organization 771  
short tandem repeats 771

- short-germ band development 771  
 shotgun cloning 269–270, 771  
 libraries 219, 771  
 sequencing 262–263, 265–266  
 sibling species 381, 499  
 signal transduction 86–87, 493–495, 640, 686–687  
 silent mutations see mutations, silent  
 silk 749, 771  
 genes 115–116  
 mat spinning 475  
 moths see *Antheraea pernyi*; *Bombyx mori*  
 silkworm see *Bombyx mori*  
 Silurian period 568, 570t  
 silverleaf whitefly see *Bemisia tabaci*  
 similarity 771  
 and homology definitions 266–267  
 and systematics 279, 531  
 simple sequence repeats (SSRs) 602–603  
 Sindbis virus 678–679  
 single copy nuclear DNA see DNA, single copy nuclear  
 single locus, multiple-allele sex determination model 432–433, 433t, 771  
 Single Molecule Real Time (SMRT) sequencer 277  
 single nucleotide polymorphism (SNP) 467, 608, 610t, 772  
 single-stranded binding proteins (SSBs) 21–22, 23f, 771  
 single-stranded DNA see DNA, single-stranded  
 single-stranded probes 241–242  
 singletons 634  
 Siphonaptera 105–106, 115–116, 161  
 SIRM see Sterile Insect Release Method  
 sister chromatids 87, 89–90  
 sister taxa 556  
*sisterless* genes 425t, 427  
 site-directed mutagenesis 388, 772  
*Sitophilus oryzae* 128–129  
 sitters 474–475  
 sleep 505–506  
 sliding-clamp proteins 27  
 slot blot 629, 772  
*Sma* see restriction endonucleases  
 small hairpin microRNA (shmiRNA) 680–681  
 small hairpin RNA (shRNA) 680–681  
 small interfering RNA (siRNA) 39b, 684, 772  
 small nuclear RNA (snRNA) 16, 39b, 54, 772  
 small nucleolar RNA (snoRNA) 16, 39b, 772  
*smell-impaired* gene 506–507  
*Sminthuris viridis* 437  
 smooth endoplasmic reticulum 55, 772  
 “So What test” 646–647  
 social behavior see behavior  
 SOD see superoxide dismutase  
*Sodalis glossinidius* 148–149, 676  
 sodium channel genes 353, 495  
 sodium dodecyl sulfate (SDS) 203–204, 291  
 sodium hypochlorite 327  
 sodium hydroxide 203–204  
*Solenopsis invicta* 107–108, 280t, 284, 432–433, 508  
 somatic cells 149–150, 772  
 somatic-sex determination 427–430, 428f  
 Son Killer 450  
 song cycles in *Drosophila* species 465  
 see also behavior  
 Southern blot analysis 183–184, 206–208, 207f, 240, 335–336, 687–688, 772  
 southern pine beetles see *Dendroctonus frontalis*  
*Spaeroma rugicauda* 446  
 speciation 135, 448, 499–501, 573–578, 772  
 allopatric 573–574  
 cladogenic 549–550  
*D. melanogaster* 499–501  
 genes 575–577, 772  
 incipient 574t  
 instantaneous 622  
 phyletic 764  
 sympatric 500, 573–574, 614–615, 773  
 species 772  
 concepts 573–575, 574t  
 cryptic 577–578, 614–615, 622, 644–645, 744  
 temporal isolation in 621–622  
 trees 549  
 see also speciation  
 specific activity of probe label 240–241, 772  
 sperm as vectors 673t  
 spin-columns 211  
 spindle see meiosis; mitosis  
*spinster* gene 498–499  
*spire* gene 156–157, 430  
*Spiroplasma* 148, 420, 433t, 443t–444t, 445  
 spliceosomes 47, 772  
 splicing 20–21, 48, 62  
*Spodoptera frugiperda* 395, 614–617  
*spongecake* gene 504  
 squash-blot procedure 629  
 SSBs see single strand binding proteins  
 stable transformation see transformation, stable  
 staggered cuts 192–193  
 stalk-eyed flies 441  
 start codon see codons, initiation  
*staufen* gene 156–157, 430  
 sterile insect release method (SIRM) 451–453, 452b, 625, 665t–667t, 667–668, 689, 707b–709b, 772  
 sterile males 452b, 772  
 sterility, hybrid 135, 442

- sticky end 772  
 stinging behavior 479  
*Stoffel* fragment 315  
 stop codons 17, 49, 773  
*Strepsiptera* phylogeny 547  
*Streptococcus pneumoniae* 5  
 stress proteins 773  
 stringency 208, 773  
 structural genomics 293–294  
 stuffer fragment of lambda phage 221f, 222, 224, 226  
*Stylochus zebra* 396  
 subclones 261, 264, 773  
 subspecies identification 534–535  
*sulfakinin* gene 682t–683t  
*Sulfolobus acidocaldarius* 315  
 supercoiled DNA see DNA, supercoiled  
 supernumerary chromosomes see chromosomes, accessory (B-type; supernumerary)  
 superoxide dismutase (SOD) gene 528–529  
*swallow* gene 155  
 SWISS-PROT 292  
 symbionts 114, 121, 502–503, 747, 773  
 diversity of 129  
 evolution in aphids 146–148  
 feminization of hosts 138  
 functions 130  
 genetically engineered see paratransgenesis  
 gut 131, 141–143  
 horizontal transfer of 109–110, 134, 663  
 and insect behavior 502–503  
 intracellular 129  
 microbial 130, 143–146  
 mitochondria as 127–128  
 multiple types of 128–132  
 PCR detection of 360b–361b, 596  
 polydnnaviruses 139–141, 764  
 toxin production in 132  
 vertical transfer of 134, 147  
 virus-like 130  
 see also *Wolbachia*
- sympatric speciation 500, 573–574, 614–615, 773  
*Sympyla* 524t  
 synapses 773  
 synaptonemal complexes 90, 117–118  
 syncytium 151, 383–384, 773  
 syncatology 593, 773  
 syngamy 151, 437, 773  
 synteny 294, 773  
 synthetic biology 773  
 synthetic probes 242  
 systematics 773  
 characters 549–550  
 cladistic 526, 550  
 definition 522  
 maximum likelihood methods 559, 562  
 molecular methods for 521  
 parsimony methods 556, 559, 562, 763  
 phenetic 526, 550, 763  
 systems genetics 295–297, 506–507
- T**
- T4 DNA ligase see DNA ligase  
 T4 DNA polymerase see DNA polymerase  
 T4 polynucleotide kinase see polynucleotide kinase  
 T7 DNA polymerase see DNA polymerase  
*Tachinidae* 134  
*TAHRE* see transposable elements  
*takeout* gene 499  
*Talaeporia tubulosa* 442  
 TALENs 375, 400, 401b, 404–405, 662, 665t–667t, 679–680, 773  
*Tam3* see transposable elements  
*Tamarixia radiata* 339f  
 tandem repeats 74, 266, 768, 774  
 Taq DNA polymerase 255, 310f–311f, 313–315, 316t–317t, 324–325, 336, 634, 774  
*Tardigrada* 548
- targeted gene insertion 390  
 gene replacement 378–379  
 gene transfer 387, 389f, 774  
 mutagenesis 774  
*TART* see transposable elements  
 TATA box 44–45, 277–278, 753  
 TATAAT sequence 44–45  
 taxonomy 522, 524, 774  
 Tay-Sachs disease 504  
*Tc1* see transposable elements  
 tDNA-PCR see polymerase chain reaction (PCR), tDNA  
 telomerase 28, 83, 234, 774  
 telomere terminal transferase see telomerase  
 telomeres 9b, 28, 287, 774  
 DNA 28  
 function 82–84  
 Het-A sequences in 83  
 see also chromosomes  
 telophase 86  
 template 316t–317t, 774  
*Tenebrio molitor* 82  
*Tenebrionidae* 82  
*Tephritidae* 395, 673–674  
 terminal transferase see telomerase  
 termination codons 17, 49, 773  
 termites  
 symbionts 141–142  
 see also individual species  
*tetanic* gene 487  
 tetracycline 197  
 antibiotic use 132  
 tetracycline-repressible transcriptional activator (tTAV) 707–708  
*Tetrahymena* 83  
 tetramethylammonium chloride (TMAC) 318  
*Tetraneura ulmi* 442  
*Tetranychus urticae* 107–108, 280t, 283, 353–354, 360b–361b, 437, 549, 673t, 683  
 tetrotransposable element see transposable elements  
 TFIID 44

- TFPGA program 635  
 thelygenic females 436, 774  
 thelytoky 104–105, 106, 106,  
   431, 433t, 447, 507–508,  
   774  
 cyclical 106  
 evolution 423–424  
 facultative 106–107  
 induced 132–133, 136,  
   138–139  
 and sex determination  
   423–424  
 thermal cyclers 324  
*Thermococcus litoralis* 315  
 thermostable DNA polymerase  
   315  
   see also Taq DNA  
   polymerase  
*Thermus aquaticus* 315  
*Thermus litoralis* 315, 336  
*Thermus thermophilus* 315,  
   336  
 theta 635  
 threonine 18t  
 thymine 3–4, 8, 10f, 15f  
 Thysanoptera 115–116  
 Thysanura 161  
 ticks 322  
 tiling microarrays 285–287,  
   290, 775  
 tilling 293  
*timeless* gene 466, 480t, 481,  
   484  
 tissue, definition 775  
 tobacco budworm see  
   *Heliothis virescens*  
 tobacco hornworm see  
   *Manduca sexta*  
*Togaviridae* 678–679  
*Toll* gene 155  
*Torymidae* 134  
 totipotent 775  
*Toxotrypana curvicauda* 398  
 Tracheata 569  
 trailer segments 45, 775  
 trans elements 775  
 transcript 775  
 transcription 16, 41–42, 52b,  
   746, 775  
   regulation 60  
   repression 768  
 transcription activator-like  
   effector nucleases see  
   TALENs  
 transcriptional activator  
   proteins 60, 775  
 transcriptome 289  
 transcriptomics 253, 288–290,  
   601, 775  
 transfection 221–222, 775  
 transfer RNA (tRNA) 16, 38, 40,  
   48–52, 775  
   function 39b  
   genes 116  
   middle-repetitive genes 74,  
   111–122  
   mitochondria 539–540  
   structure 50f  
 transformant 775  
 transformation 775  
   of arthropods 163f  
   of *E. coli* 201–202  
   germ line 381–386  
   P-mediated 394  
   with recombinant plasmid  
     DNA 185f, 199f  
   stable 384, 672, 772  
   transient 383–384, 776  
*transformer* genes 47, 425t,  
   428, 429, 432, 433t,  
   434–435, 438, 453  
 transgenes 63–64, 80–81, 384–  
   385, 686–687, 776  
   horizontal transfer 693–696  
   suppression 399, 692, 776  
 transgenic arthropods see  
   genetically modified  
   arthropods (GMA)  
 transgenic crops 616–617  
 transgenic flying vaccinator  
   665t–667t  
 transgenic insects 661,  
   665t–667t, 670t–671t  
 transgenic organisms 776  
*Transib* see transposable  
   elements  
 transient transformation  
   see transformation,  
   transient  
 transitions see mutations,  
   transitions  
 translation 16, 48–52, 776  
 translational regulation 62–63,  
   776  
   see also gene regulation  
 translocations 92, 776  
 transovarial transmission 776  
 transposable elements 56–57,  
   72, 99, 104, 108, 117–  
   122, 266–267, 373, 422,  
   776  
   17.6 element 119–120  
   *Accord* 119–120  
   *Activator* 120–121  
   Class I 118t  
   Class II 118t, 122  
   conversion to activity 398–  
   399, 743  
   *copia*-like 118t  
   cross mobilization of 398  
   as drivers to insert genes  
   into populations 376,  
   392, 696  
   F 118t  
   FB (foldback) 118t, 749  
   and genome evolution  
   287–288  
   *gypsy* 678  
   *gypsy*-like 118t  
   hAT 398  
   *Helena* 377–378  
   *Helitrons* 118t, 119, 122  
   *Hermes* 375, 393, 395–396,  
   398, 672–674, 753  
   *hermit* 395  
   *Herves* 395–396  
   *HeT-A* 123, 271  
   *hobo* 120–121, 271, 375, 379,  
   393, 398, 672–674, 753  
   horizontal transfer of  
   693–696  
   *Hsmar1* 397  
   inverted repeats 266, 755  
   *jockey* 120–121  
   *Juan* 119–120  
   LINEs 119–120  
   long terminal repeats  
   118–119, 545–546  
   *mariner* 121, 375, 393,  
   396–398, 672, 674, 694,  
   696, 757  
   method of transposition  
   378–379

- micropia* 120  
*Minos* 375, 393, 396, 398, 672, 675, 759  
 movement 32f  
 and mutations 32  
*P* elements see *P* elements  
*Pao* 118t  
*Paris* 377–378  
*Penelope* 377–378  
*piggyBac* 375, 393, 395, 398, 664–667, 672, 675, 702–703, 764  
*pogo* 397  
*Q* 391  
 R1 and R2 elements 118t  
 resistance to 390–392  
 retrotransposons 267, 545–546, 769  
 retrovirus-like 118t  
 rolling-circle transposons 118t, 122  
*Sleeping Beauty* 397, 674  
 and speciation 104  
*TAHRE* 271  
*Tam3* 120–121  
*TART* 123, 271  
*Tc1* 392–393, 397  
*TE1* 109  
*Tigger* 397  
*Transib* 119  
 transposase 382–383, 388, 776  
*Ulysses* 377–378  
 transposase 382–383, 388, 776  
 transposition 378–379, 776  
 secondary 384, 770  
 transposons 267, 776  
 jumping 378, 387  
 tagging 374–375, 386–387  
   see also individual transposons  
 transversions see mutations, transversions  
 tree(s)  
   gene 549  
   of life 564–568, 566f, 776  
   phylogenetic 549, 551, 557, 764  
   rooted 549, 550f, 557–558  
   terraces 562  
   types 549–551
- unrooted 549, 550f, 557–558, 777  
*Trialeurodes vaporariorum* 344  
 Triassic period 569, 570t  
*Triatoma infestans* 675  
*Tribolium* 683  
   developmental genetics 162–163  
*Tribolium castaneum* 75, 107–108, 163–164, 163f, 280t, 282, 396, 442, 607, 673t, 681, 681f  
*Tribolium confusum* 501–502  
*Tribolium madens* 122–123, 443t–444t  
*Trichogramma* spp. 134, 443t–444t, 447, 453–454  
*Trichomalopsis dubius* 450  
*Trichoplusia ni* 353–354, 395, 398, 401, 675  
*Trichoptera* 83, 123, 161  
 7-tricosene 499  
*Trilobita* 524t  
 trinucleotides 602–603  
 triosephosphate isomerase gene 18–19  
*Trioxys pallidus* 343f, 605–606, 606f  
 triplex DNA see DNA, triplex  
 triploidy 421  
 tritocerebrum 469  
*Trixosclididae* 381  
 tRNA see RNA  
 tryptophan 18t, 334–335  
 tsetse fly see *Glossina*  
 tubulins 18–19, 150–151, 776  
*tudor* gene 156–157, 430  
*turnip* gene 487, 488t, 489–490  
*turnip sawfly* see *Athalia rosae*  
 two-spotted spider mite see *Tetranychus urticae*  
 tyrosine 18t
- U**
- UAA see codons  
 UAG see codons  
 ubiquitin 56, 776  
 UGA see codons  
*Ultrabithorax* gene 161  
 ultraviolet radiation 206
- Ulysses* see transposable elements  
 unique-sequence DNA see DNA, unique-sequence  
 Universal Tree of Life 564–568, 566f  
 unrooted tree 549, 550f, 550f, 557–558, 777  
 unweighted pair-group method using an arithmetic average see UPGMA  
 UPGMA 541, 558, 777  
 upstream 777  
 uracil 9b, 10f, 777
- V**
- VA/VP see heritability  
 vaccination 242–243  
 valine 18t  
*valois* gene 156–157, 430  
*Varroa destructor* 280t, 473, 478, 578, 608, 673t  
*Varroa jacobsoni* 578  
*Varroa underwoodi* 578  
 vasa gene 156–157, 430  
 VectorBase 267  
 vectors 184, 216–217, 777  
   BACs 216–217, 230, 739  
   cosmid 219, 226–228, 227f  
   disease 645  
   EMBL series 226  
   expression 198–200, 231, 748  
   gt 226  
   insertion 220–221, 755  
   M13 216–217, 228–230, 229f, 759  
   pantropic retroviral 678  
   parvovirid 678  
   pcosEMBL 228  
   pJB8 228  
   polydnnaviral 139–141  
   replacement 220–221, 221f, 768  
   sequencing 198–200  
   Sindbis virus 678–679  
   viral 677–679  
   see also baculovirus vectors; plasmids
- velvet bean caterpillar see *Anticarsia gemmatalis*

- velvet worms see *Onychophora*  
 Vent DNA polymerase 336  
 Venter, Craig 269  
 vertical transfer 448, 777  
*VG/VP* see heritability  
 virus-like symbionts 130  
 viruses 121–122, 316, 677–679, 695, 777  
 vitelline membrane 150, 152f  
*Vitellogenin* genes 116–117, 682t–683t  
 vitellogenins 150, 777  
*Volado* gene 488t, 490  
 volatile organic compounds (VOCs) 640  
 von Tschermark, Erich 4–5
- W**
- W chromosome 432  
 walking behavior 489, 501  
 walnut husk fly see *Rhagoletis completa*  
 Watson, James 5–6  
 West Nile virus 114–115, 355  
 Western blot analysis 183–184, 206, 777  
 white gene 109, 383–384, 388, 475  
 white-alcohol dehydrogenase transgene 399  
 whole-genome shotgun sequencing 235t, 262–263, 777  
 Wieschaus, Eric 154  
*Wigglesworthia glossinidiae* 148–149, 676  
 wild-type allele 29, 32, 777  
 Wilkins, Maurice 5–6  
 wing length and propensity to fly 476–477  
*wingless* gene 431
- wings clipped plasmid 386f  
 wobble hypothesis 38, 49–50, 777  
*Wolbachia* 104, 128–129, 132–139, 282, 359, 360b–361b, 777  
 bacteriophage in 136  
 bidirectional incompatibility 136  
 blocking transmission of disease-causing agents 137b, 148–149  
 cause of cytoplasmic incompatibility 130, 132–133, 133f, 137b  
 detection 338  
 as driver of genes into populations 136, 689  
 evolution 134–135  
 female killing 446–447  
 feminization 136, 433t  
 and genetic control of pests 663, 665t–667t, 689–691, 703b–707b  
 genome project 136  
 horizontal gene transfer 109–110, 134, 693–696  
 immune response modification 137b  
 male killing 132–133, 136, 137b, 446–447, 449  
 modification of sex determination 420, 433t, 443t–444t, 443t–444t, 446–449  
 and speciation 134–135, 448  
 and thelytoky 132–133, 136, 447  
 transfer from other arthropods 679  
 unidirectional incompatibility 135
- Wright's correlation (FST) 612  
*Wuchereria bancrofti* 114–115
- X**
- X:A chromosome ratio 420, 427  
*see also* sex determination  
 X chromosome 95, 777  
 mutations 482  
 X-Gal 224, 777  
 X-irradiation 29  
 xanthine dehydrogenase gene 383–384  
 xenologs 561–562  
*Xenopus* 160
- Y**
- Y chromosome 81, 94–95, 778  
*yar* gene 505–506  
 yeast artificial chromosome (YACs) 254  
 yeasts 128, 142  
*see also* *Saccharomyces*  
*Yersinia pestis* 342, 355  
 yolk 152f  
 protein genes 18–19  
 proteins 117
- Z**
- Z chromosome 432, 778  
 zinc-finger nucleases 375, 400, 401b, 403–404, 662, 665t–667t, 679–680, 778  
 zinc-finger proteins 60, 61f, 778  
 Zoraptera 105–106  
*zucchini* gene 487  
*Zygina pullata* 433t, 740  
 zygote 778  
 zygotic-segmentation genes 157–161