Bio351
Group Project 1:

The lack of fingerprints

changing the DNA codon from CGA to GGA. This changes the complementary mRNA codon from GCU to CCU, which specifies proline. A mutation that changes the amino specified by a codon is known as a **nonsynonymous** (or **replacement**) **substitution**. Switching an amino acid may alter the function of a protein. For example, having a proline versus an alanine as the 49th amino acid in taste receptor TAS2R38 influences a person's ability to taste bitter flavors.

A mutation is any change in sequence in the genome of an organism. Some mutations alter the phenotype; others do not.

Finally, consider a substitution that can occur in the third position of the codon specifying tryptophan. A C-to-T transition changes the DNA codon from ACC to ACT. This changes the mRNA codon from UGG to UGA. UGA is a stop codon. It signals that the protein is complete and no more amino acids should be added. A mutation that introduces a premature stop codon is called a **nonsense mutation**. Nonsense mutations often render the encoded protein nonfunctional (Yamaguchi-Kabata et al. 2008). Many humans carry loss-of-function nonsense mutations in both of their copies of the gene for the muscle protein alpha-actinin-3 (North et al. 1999). The nonsense allele is overrepresented in elite endurance athletes and underrepresented in elite sprint and strength athletes (Niemi and Majamaa 2005; Roth et al. 2008).

Changing the meaning of a codon is not the only way a point mutation can alter protein function or expression. Many genes in eukaryotes contain intervening sequences, or **introns**, embedded among the coding sequences, or **exons**. The introns are transcribed into the mRNA and must be spliced out before translation. Mutations in splice sites can prevent introns from being excised, resulting in production of abnormal proteins. Janna Nousbeck et al. (2011) discovered a splice-site mutation in humans that causes adermatoglyphia—the absence of fingerprints (Figure 5.23). Mutations in the promoter regions of genes, noncoding sequences that play a role in gene regulation, can alter gene expression.

Figure 5.23 Some people lack fingerprints The condition is called adermatoglyphia, and is informally known as immigration delay disease. From Nousbeck et al. (2011).

Like point mutations, insertions and deletions (collectively called **indels**) vary in their effects. And as with point mutations, the genetic code shows why. Insertion or deletion in a coding region of one, two, or any other number of nucleotides not a multiple of three results in a shift of the codon reading frame. This changes the meaning of every codon downstream from the mutation.

The mutational mechanisms we have considered in this section stock populations with a diversity of alleles. In the next few chapters, we will be concerned with the relative frequencies of different alleles in populations. **Computing Consequences 5.2 (next page)** shows how allele frequencies can be quantified.

5.3 Where New Genes Come From

As with mutations that create new alleles, many mechanisms generate new genes (Long et al. 2003; Kaessmann 2010). We cannot cover them all, but we can get a sense of where new genes come from by considering a few examples.

Gene Duplication

Two mechanisms of **gene duplication** are thought to be among the most common sources of new genes. The first is **unequal crossing over**, an error in the genetic recombination that happens during meiosis. In normal crossing over, homologous chromosomes (the maternal and paternal members of a pair) align side by side during prophase of meiosis I and exchange stretches of DNA containing the same loci. In unequal crossing over, the homologous chromosomes align incorrectly. This can happen if the same nucleotide sequence occurs in more than

Bio351
Group Project 2:

A monkey with unusual diet

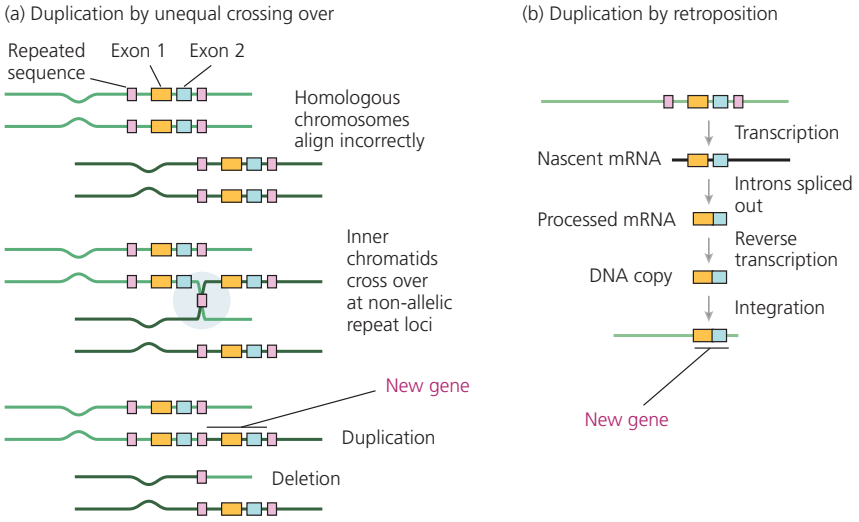

Figure 5.26 Two mechanisms of gene duplication (a) Unequal crossing over during meiosis. (b) Retroposition.

one place on the chromosome, as when copies of a transposable element have inserted at multiple loci (Figure 5.26a). The consequence of misalignment is that the DNA segments exchanged are out of register. One of the participating chromatids ends up with a duplication, while the other sustains a deletion.

The second mechanism, depicted in Figure 5.26b, is called **retroposition** or **retroduplication**. Retroposition begins when a processed messenger RNA, from which the introns have been spliced out, is reverse-transcribed by the enzyme reverse transcriptase to form a double-stranded segment of DNA. If this DNA segment integrates into one of the main chromosomes, the genome acquires a duplicated copy of the original gene. In many cases the new copy is a nonfunctional **pseudogene**, because it lacks regulatory sequences that cause it to be transcribed. If, however, the duplicate inserts near existing regulatory sequences, subsequently acquires them via a transposable element insertion, or evolves them from scratch, it may become a functional gene.

Retroposition and unequal crossing over leave distinctive footprints in the genome. Among other clues to their origin, retroduplicated genes lack introns and are usually found far from the original gene. Genes that were duplicated by unequal crossing over, in contrast, contain the same introns as their parental genes and are found in tandem with them on the same chromosome.

A New Gene Generated by Unequal Crossing Over

An example of a gene created by unequal crossing over comes from work by Jianzhi Zhang and colleagues (2002) on the douc langur (*Pygathrix nemaeaus*) of Southeast Asia. The douc has an unusual diet for a monkey: It eats leaves. The leaves are fermented by bacteria living in the monkey's foregut. Further along the digestive tract, the monkey digests the bacteria and absorbs the nutrients they contain. Like ruminants, which have a similar diet and digestive strategy, the douc maintains a relatively high concentration of RNASE1 in its foregut. RNASE1 is an enzyme, made by the pancreas, that breaks down RNA. This liberates the nitrogen in the RNA for recycling by the monkey's own metabolism.

Zhang and colleagues examined the genes for RNASE1 in douc langurs and other primates. Most primates have just one locus encoding RNASE1, but doucs have two. Zhang named the second enzyme RNASE1B. Figure 5.27a displays an

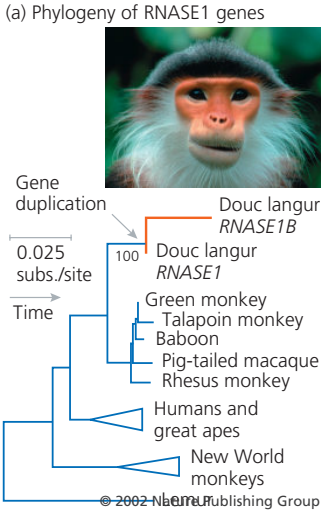

Figure 5.27 A new gene from unequal crossing over (a) Phylogeny of RNASE1 genes. (b) RNASE1 reaction norms.

(a, b) Reprinted by permission from Macmillan Publishers Ltd: Zhang, J., Y. P. Zhang, and H. F. Rosenberg. 2002. "Adaptive evolution of a duplicated pancreatic ribonuclease gene in a leaf-eating monkey." *Nature Genetics* 30: 411–415.

evolutionary tree, based on nucleotide sequence data, for the singleton RNASE1 genes of 15 primates, plus the two RNASE1 genes of doucs. The langur genes are each other's closest relatives. The simplest explanation is that the RNASE1B gene in doucs arose as a recent duplication of the monkey's RNASE1 gene. The two langur genes each have an intron, and their introns are nearly identical. This suggests that the duplication arose by unequal crossing over.

Zhang and colleagues found nine amino acid substitutions distinguishing the proteins encoded by the douc langur's two RNASE1 genes. To see whether the enzymes had diverged in function, the researchers tested their RNase activity at a variety of pHs. The reaction norms appear in Figure 5.27b. RNASE1B is more active at the relatively low pHs characteristic of the douc langur small intestine. RNASE1, on the other hand, retains an optimal pH similar to that of other primate RNASE1s. In addition, the researchers found that douc RNASE1B has lost the ability to break down double-stranded RNA, a capacity seen in other RNASE1s and thought to play a role in defense against viruses.

These results are consistent with the hypothesis that following the duplication creating RNASE1B, the new gene evolved to encode an enzyme specialized for the digestive demands of the douc langur's unusual diet, while the parental gene retained its ancestral, generalist function. Similar duplications and specializations of RNASE1 genes appear to have occurred independently in ruminants (Zhang 2003) and African leaf-eating monkeys (Zhang 2006; Yu et al. 2010).

Genes that are duplicated within a genome and later diverge in function, like RNASE1 and RNASE1B in douc langurs, are described as **paralogous**. Paralogous genes can be contrasted with **orthologous** genes. These are genes that are derived from a common ancestral sequence and separated by a speciation event, like RNASE1 in douc langurs and RNASE1 in humans.

A New Gene Generated by Retroposition

An example of a new gene created by retroposition comes from the work of Heidi Parker and colleagues (2009) on dogs. The researchers sought to find the gene or genes responsible for chondrodysplasia, the short-legged condition characteristic of corgis, dachshunds, bassets, and a variety of other breeds **(Figure 5.28a)**. Scanning the dog genome for differences between breeds with and without chondrodysplasia, Parker and colleagues found, on chromosome 18, a duplicate copy of the gene for fibroblast growth factor 4 (*fgf4*). Possession of the duplicate is strongly associated with chondrodysplasia. Figure 5.28b sorts breeds with and without the condition into three categories. In most breeds all individuals tested either carried the duplicate on both copies of chromosome 18 (*DD*), or on neither copy (*NN*). A few breeds were polymorphic, meaning that different individuals had different genotypes. (The breeds classified as Other in which all dogs carried two copies of the duplicate—cairn terrier, Norwich terrier, and shih tzu—have short legs but do not meet a stringent definition of chondrodysplasia.)

The duplicate copy of *fgf4* lacks introns and is located some distance from the parent copy. These and other clues indicate that the duplicate arose by retroposition. Parker and colleagues found that in dogs that carry it, the duplicate gene is expressed in the joint-forming cartilage in the long bones of puppies, as is the parent copy. It appears that the duplicate acquired a promoter by serendipitously inserting into the middle of a transposable element. Fine control over when and where the encoded protein is actually made likely involves regulatory sequences in the untranslated portion of the mRNA.

(a) Who's a good girl?

(b) *Fgf4* retrogene genotypes in dogs of various breeds

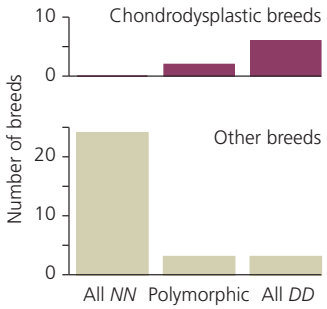

Figure 5.28 A new copy of the gene *fgf4* from retroposition (a) Like other Welsh corgis, Juno has chondrodysplasia—a condition characterized by short legs. (b) Across dog breeds, chondrodysplasia is strongly associated with the presence of a retroduplicated copy of the gene for fibroblast growth factor 4. Prepared using data from Parker and colleagues (2009).

Bio351
Group Project 3:

Short leggedness in dogs

evolutionary tree, based on nucleotide sequence data, for the singleton RNASE1 genes of 15 primates, plus the two RNASE1 genes of doucs. The langur genes are each other's closest relatives. The simplest explanation is that the RNASE1B gene in doucs arose as a recent duplication of the monkey's RNASE1 gene. The two langur genes each have an intron, and their introns are nearly identical. This suggests that the duplication arose by unequal crossing over.

Zhang and colleagues found nine amino acid substitutions distinguishing the proteins encoded by the douc langur's two RNASE1 genes. To see whether the enzymes had diverged in function, the researchers tested their RNase activity at a variety of pHs. The reaction norms appear in Figure 5.27b. RNASE1B is more active at the relatively low pHs characteristic of the douc langur small intestine. RNASE1, on the other hand, retains an optimal pH similar to that of other primate RNASE1s. In addition, the researchers found that douc RNASE1B has lost the ability to break down double-stranded RNA, a capacity seen in other RNASE1s and thought to play a role in defense against viruses.

These results are consistent with the hypothesis that following the duplication creating RNASE1B, the new gene evolved to encode an enzyme specialized for the digestive demands of the douc langur's unusual diet, while the parental gene retained its ancestral, generalist function. Similar duplications and specializations of RNASE1 genes appear to have occurred independently in ruminants (Zhang 2003) and African leaf-eating monkeys (Zhang 2006; Yu et al. 2010).

Genes that are duplicated within a genome and later diverge in function, like RNASE1 and RNASE1B in douc langurs, are described as **paralogous**. Paralogous genes can be contrasted with **orthologous** genes. These are genes that are derived from a common ancestral sequence and separated by a speciation event, like RNASE1 in douc langurs and RNASE1 in humans.

A New Gene Generated by Retroposition

An example of a new gene created by retroposition comes from the work of Heidi Parker and colleagues (2009) on dogs. The researchers sought to find the gene or genes responsible for chondrodysplasia, the short-legged condition characteristic of corgis, dachshunds, bassets, and a variety of other breeds **(Figure 5.28a)**. Scanning the dog genome for differences between breeds with and without chondrodysplasia, Parker and colleagues found, on chromosome 18, a duplicate copy of the gene for fibroblast growth factor 4 (*fgf4*). Possession of the duplicate is strongly associated with chondrodysplasia. Figure 5.28b sorts breeds with and without the condition into three categories. In most breeds all individuals tested either carried the duplicate on both copies of chromosome 18 (*DD*), or on neither copy (*NN*). A few breeds were polymorphic, meaning that different individuals had different genotypes. (The breeds classified as Other in which all dogs carried two copies of the duplicate—cairn terrier, Norwich terrier, and shih tzu—have short legs but do not meet a stringent definition of chondrodysplasia.)

The duplicate copy of *fgf4* lacks introns and is located some distance from the parent copy. These and other clues indicate that the duplicate arose by retroposition. Parker and colleagues found that in dogs that carry it, the duplicate gene is expressed in the joint-forming cartilage in the long bones of puppies, as is the parent copy. It appears that the duplicate acquired a promoter by serendipitously inserting into the middle of a transposable element. Fine control over when and where the encoded protein is actually made likely involves regulatory sequences in the untranslated portion of the mRNA.

(a) Who's a good girl?

(b) *Fgf4* retrogene genotypes in dogs of various breeds

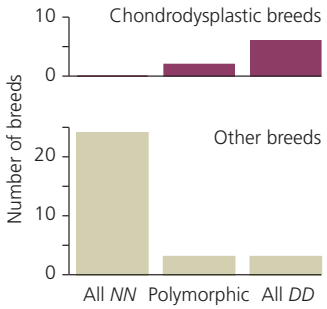

Figure 5.28 A new copy of the gene *fgf4* from retroposition (a) Like other Welsh corgis, Juno has chondrodysplasia—a condition characterized by short legs. (b) Across dog breeds, chondrodysplasia is strongly associated with the presence of a retroduplicated copy of the gene for fibroblast growth factor 4. Prepared using data from Parker and colleagues (2009).

All the chondroplastic breeds that Parker and colleagues tested carried the same duplicate gene at the same locus. It thus appears that the duplicate arose only once, before the development of today's breeds. It has subsequently been driven to high frequency in modern short-legged breeds by artificial selection.

Duplicated Genes and Evolution

Gene duplication is an important category of mutation for evolution. An increase in copy number may itself be adaptive. Duplication followed by divergence in function generates gene families.

Gene duplication accounts for a substantial fraction of the genomic variation among individuals, and thus the raw material for evolution. Recent estimates suggest that more of the genome is affected by copy number variation than by differences derived from point mutations (Mills et al. 2011). Variation among individuals in gene copy number can, by itself, serve as a substrate for adaptive evolution (Perry et al. 2007). And serial duplication followed by gene divergence has generated families of functionally related genes that, as in the case of olfactory receptor genes, can include hundreds of members (Young et al. 2008).

New Genes from Scratch

Although most new genes are born as duplicates of existing genes, some genes appear to be born from noncoding DNA. David Knowles and Aoife McLysaght (2009) found evidence for three examples in humans. We will briefly discuss one.

C22orf45 is a gene of unknown function unique to humans. It is transcribed in a variety of tissues, and its mRNA is known to be translated into protein.

Even though the gene is unique to humans, similar nucleotide sequences occur at the homologous locus in chimpanzees, gorillas, orangutans, gibbons, and macaques. The sequences in all of these nonhuman primates contain elements, including premature stop codons, that would substantially alter the encoded protein were the sequence transcribed and translated. One of the premature stop codons is shared by all five nonhuman species **(Figure 5.29)**.

Figure 5.29 A new human gene from noncoding DNA In other apes and in macaques, the sequence homologous to the human gene *C22orf45* contains a premature stop codon. From Knowles and McLysaght (2009).

The most parsimonious explanation for this pattern is that the sequence was noncoding in the last common ancestor of humans and the other five species, and became a coding gene in the lineage leading to modern humans.

Having discussed mutations that create new alleles and new genes, we now turn to the most drastic mutations: those that alter large portions of chromosomes or even the entire genome.

5.4 Chromosome Mutations

The mutations discussed thus far occur on the scale of a single base pair in DNA to segments containing tens of thousands of base pairs. These alterations pale in comparison to mutations that alter the gross morphology of chromosomes. Some of these mutations affect only gene order and organization; others produce duplications or deletions that affect the total amount of genetic material. They can also

Bio351
Group Project 4:

Inversion in *Drosophila subobscura*

involve the entire DNA molecule or segments of varying sizes. Here we focus on two types of chromosome alterations that are particularly important in evolution.

Inversions

Chromosome **inversions** often result from a multistep process that starts when radiation causes two double-strand breaks in a chromosome. After breakage, a chromosome segment can detach, flip, and reanneal in its original location. As Figure 5.30 shows, gene order along the chromosome is now inverted.

In addition to involving much larger stretches of DNA than point mutations and gene duplications, inversions have different consequences. Inversions affect a phenomenon known as genetic **linkage**. Linkage is the tendency for alleles of different genes to assort together at meiosis. Genes on the same chromosome tend to be more tightly linked (that is, more likely to be inherited together) than genes on nonhomologous chromosomes. Similarly, the closer together genes are on a chromosome, the tighter the linkage. Crossing over at meiosis, on the other hand, breaks up allele combinations and reduces linkage (see Chapter 8).

When inversions are heterozygous, meaning that one chromosome copy contains an inversion and the other does not, the inverted sequences cannot align properly when homologs synapse during prophase of meiosis I. Successful crossing-over events are rare. The result is that alleles inside the inversion are locked so tightly together that they are inherited as a single "supergene."

Inversions are common in *Drosophila*. Are they important in evolution? To answer this question, consider a series of inversions found in populations of *Drosophila subobscura*. This fruit fly is native to western Europe, North Africa, and the Middle East, and has six chromosomes. Five of these chromosomes are **polymorphic** for at least one inversion (Prevosti et al. 1988), meaning that chromosomes with and without the inversions exist. Biologists have known since the 1960s that the frequencies of these inversions vary regularly with latitude and climate. This type of regular change in the frequency of an allele or an inversion over a geographic area is called a **cline**. Several authors have argued that different inversions must contain specific combinations of alleles that function well together in cold, wet weather or hot, dry conditions. But is the cline really the result of natural selection on the supergenes? Or could it be a historical accident, caused by differences in the founding populations long ago?

A natural experiment has settled the issue. In 1978 *D. subobscura* showed up in the New World for the first time, initially in Puerto Montt, Chile, and then four years later in Port Townsend, Washington, USA. Several lines of evidence argue that the North American population is derived from the South American one. For example, of the 80 inversions present in Old World populations, precisely the same subset of 19 is found in both Chile and Washington State. Within a few years of their arrival on each continent, the *D. subobscura* populations had expanded extensively along each coast and developed the same clines in inversion frequencies found in the Old World (Figure 5.31). The clines are even correlated with the same general changes in climate type: from wet marine environments to mediterranean climates to desert and dry steppe habitats (Prevosti et al. 1988; Ayala et al. 1989). This is strong evidence that the clines result from natural selection and are not due to historical accident.

Which genes are locked in the inversions, and how do they affect adaptation to changes in climate? In the lab, *D. subobscura* lines bred for small body size tend to become homozygous for the inversions found in the dryer, hotter part of the

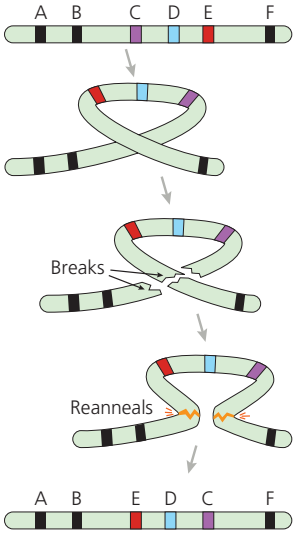

Figure 5.30 Chromosome inversion Inversions result when a chromosome segment breaks in two places and reanneals with the internal segment reversed. Note the order, before and after, of the genes labeled C, D, and E.

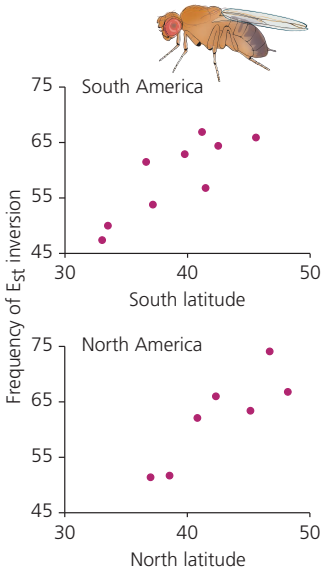

Figure 5.31 Inversion frequencies form clines in *Drosophila subobscura* From data in Prevosti et al. (1988); see also Balanyà et al. (2003).

range (Prevosti 1967). Research by George Gilchrist and colleagues (2004) has confirmed that pronounced and parallel clines in body size exist in fly populations from North America, South America, and Europe. These results hint that alleles in the inversions affect body size, and that natural selection favors large flies in cold, wet climates and small flies in hot, dry areas. The fly study illustrates a key point about inversions: They are an important class of mutations because they affect selection on groups of alleles.

Chromosomal inversions suppress recombination and thus help maintain combinations of alleles across nearby loci.

Genome Duplication

The final type of mutation we will consider occurs at the largest scale possible: entire sets of chromosomes. For example, if homologous chromosomes fail to segregate during meiosis I or if sister chromatids do not separate properly during meiosis II, the resulting cells may have double the number of chromosomes of the parent cell. In plants, because the germ line is not segregated, similar mutations can occur during the mitotic cell divisions that lead up to gamete formation. Mutations like these can lead to the formation of a diploid gamete in species where gametes are normally haploid.

Figure 5.32 shows one possible outcome of a chromosome-doubling mutation. In the diagram, the individual that produces diploid gametes contains both male and female reproductive structures and can self-fertilize. When it does so, a tetraploid ($4n$) offspring results. If this offspring self-fertilizes when it matures, or if it mates with its parent or a tetraploid sibling that also produces diploid gametes, then a population of tetraploids can become established.

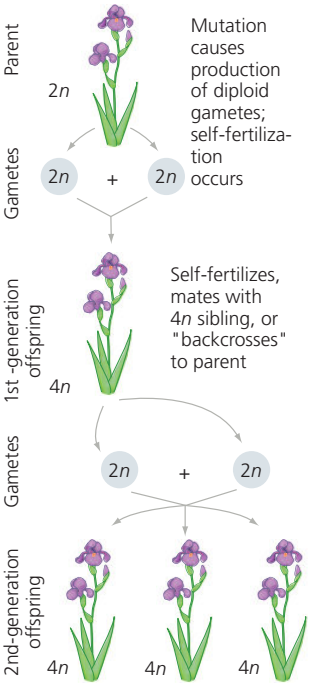

Figure 5.32 How tetraploid plants are produced

Organisms with more than two chromosome sets are said to be **polyploid**. Polyploid organisms can be tetraploid ($4n$), hexaploid ($6n$), octoploid ($8n$), or more. Polyploidy is common in plants and rare in animals—probably because self-fertilization is more common in plants than animals. Nearly half of all flowering plant species and the vast majority of the ferns are descended from ancestors where polyploidization occurred. In animals, polyploidy occurs in taxa like earthworms and some flatworms where individuals contain both male and female gonads and can self-fertilize. It is also present in animal groups that are capable of producing offspring without fertilization, through a process called parthenogenesis. In some species of beetles, sow bugs, moths, shrimp, goldfish, and salamanders, a type of parthenogenesis occurs that can lead to chromosomal doubling.

There are at least two reasons that polyploidy is an important type of mutation in evolution. First, it can lead to new species being formed. Second, it alters cell size, cell geometry, and gene dosage, and thus may endow individuals with new phenotypes that allow them to colonize and adapt to new environments.

Polyploidy and Speciation

To see why genome duplication can lead to speciation, imagine the outcome of matings between individuals in a tetraploid population and the most closely related diploid population. If individuals from the two populations mate, they produce triploid offspring. When these individuals mature and meiosis occurs, the homologous chromosomes cannot synapse correctly, because they are present in an odd number. As a result, the vast majority of the gametes produced by triploids end up with the wrong number of chromosomes and fail to survive. Triploid individuals have extremely low fertility.

Duplications of the entire genome are an important mechanism of speciation—particularly in plants.

In contrast, when tetraploid individuals continue to self-fertilize or mate among themselves, then fully fertile tetraploid offspring will result. In this way,

Bio351
Group Project 5:

The case of *Achillea borealis*

natural selection should favor polyploids that are reproductively isolated from their parent population. Diploid and tetraploid populations that are genetically isolated are on their way to becoming separate species.

Genome Duplication and Adaptation

Justin Ramsey (2011) tested the hypothesis that polyploidy facilitates the colonization of, and adaptation to, new environments by performing a common garden experiment with wild yarrow **(Figure 5.33a)**. Along the coast of northern California, where Ramsey worked, yarrow populations with different ploidy occupy distinct habitats. Tetraploid populations live in coastal grasslands, conifer forests, and mountain meadows. Hexaploid plants live in sand dunes and oak woodlands. Because tetraploid plants occasionally produce hexaploid offspring, Ramsey suspected that the hexaploid populations were derived from tetraploid ancestors, and that the increase in ploidy aided their shift to drier habitats.

Ramsey grew tetraploid plants, first-generation hexaploid plants from tetraploid parents (neo-hexaploids), and hexaploid plants—all from wild populations—next to each other in sand dunes. He monitored them for three years.

The data on survival, displayed in 5.33b, show that hexaploids, which ordinarily live in dunes, are better adapted to dunes than tetraploids are. This is no surprise. The key result is that the neo-hexaploids did substantially better than the tetraploids, though not as well as the ordinary hexaploids. The data plotted in Figure 5.33c show that neo-hexaploids were also intermediate in flowering time.

Ramsey's results are consistent with the idea that changes in ploidy, by themselves, can alter phenotypes in a way that makes individuals better adapted to new environments. And they are also consistent with the idea that once a population with a novel ploidy level has colonized a new habitat, it can evolve by natural selection to become even better adapted.

This and the other research we have covered has illuminated the myriad ways that mutation supplies raw material for evolution. There are two more things we will need to know about mutation in the next several chapters: how often mutations happen, and how they affect the fitness of the individuals that carry them. These are these the issues we turn to in the final section of this chapter.

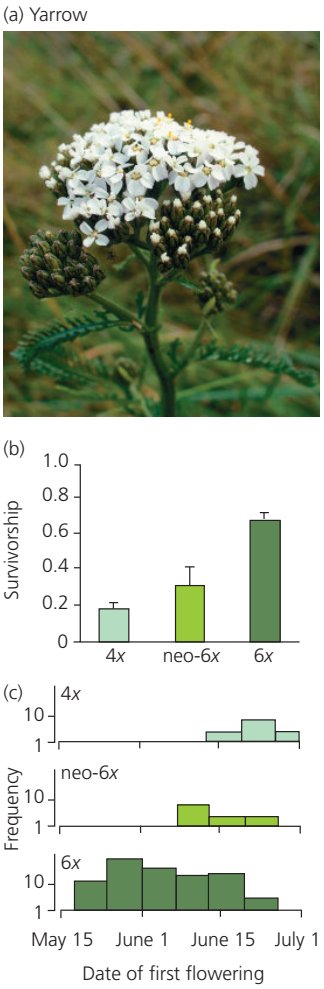

Figure 5.33 Genome duplication as a mechanism of adaptation (a) Wild yarrow (*Achillea borealis*). (b) Survival of plants with different ploidy in a common garden in dunes. (c) Flowering time distributions in the same plants. From Ramsey (2011).

5.5 Rates and Fitness Effects of Mutations

The rates and fitness effects of mutations have been hard to study, because mutations are rare and their consequences are often—though by no means always—subtle. Recently, however, advances in DNA sequencing and genetic engineering have allowed researchers to begin investigating these issues with new precision.

Mutation Rates

Traditionally, geneticists have estimated mutation rates by studying genes that, when disrupted, yield easily observable phenotypic changes (see Nachman 2004). Now that genes, and even whole genomes, can be sequenced quickly and cheaply, researchers can measure mutation rates more directly and on a larger scale.

For example, in a type of study called a **mutation accumulation** experiment, Stephan Ossowski and colleagues (2010) sequenced the genomes of five lineages of thale cress (*Arabidopsis thaliana*) derived from an already-sequenced common ancestor. Each of the five lineages had been grown under optimal conditions for 30 generations, and propagated each generation from a single randomly chosen