

height, and physiology and morphology together can affect behavioral phenotypes such as courtship. Finally, all these lower level traits can affect life history traits like survival and reproduction, which determine the ultimate trait of individual fitness. The traits that are higher in this hierarchy are more complex and affected by more gene loci. The expression of most phenotypic traits, and especially the higher level ones, are also affected to varying degrees by the environment. This complexity means that the same genotype can produce different phenotypes, through the action of the environment. Conversely, the different genotypes can produce the same phenotypes, again due to the environment and also due to gene interactions. We will discuss complex phenotypic traits and fitness in more detail in Chapters 4 through 6.

2

Population genetics I: Genetic variation, random and nonrandom mating

What is Population Genetics?

In its broadest sense, **population genetics** is the study of naturally occurring genetic differences between organisms; these differences are called **genetic variation**. Genetic variation is important because it is the raw material for evolution. Genetic variation can occur at three hierarchical levels: within populations, between populations of the same species, and between different species. Therefore, to understand the purview of population genetics, we need to have a better understanding of populations and genetic variation.

What are populations and why are they important?

Populations are important because evolutionary processes occur primarily within populations; it is populations that evolve through changes in allele frequencies. In fact, most of the important concepts and processes in ecological genetics have meaning only at the level of the population, not the individual: These include genetic variation, allele and genotype frequencies, gene flow, drift, natural selection, heritability, and genetic correlation.

A species is rarely, if ever, a single interbreeding **panmictic** group, that is, one in which any member of the species can potentially mate with any other member of the opposite sex. Instead, species are divided into **populations**. A population can be defined as a local interbreeding (panmictic) group that has reduced gene flow with other groups of the same species. By reduced

gene flow we mean reduced movement of genes caused by migration and subsequent mating (see Chapter 3). The reduced gene flow between populations can be caused by **spatial structure**, which occurs when the individuals in a species are not evenly distributed in space (Figure 2.1).

Determining the boundaries of a population in nature can be quite simple when dealing with species that occur in discrete clumps in a patchy habitat, such as fish in small ponds or plants on small islands. This determination becomes more problematic with species that are not spatially structured but rather range continuously over a wide area, such as maple trees in a large forest or wide-ranging birds. Here the probability of interbreeding is a function of distance; individuals at opposite ends of the species distribution are certainly not interbreeding, but distinct population boundaries cannot be drawn. The key question from a genetic viewpoint is how much interbreeding and, thus, gene flow takes place at any given distance.

The terminology applied to populations can be confusing. Sometimes a group of populations connected by some level of gene flow is called a **metapopulation** (all the populations within the boundary in Figure 2.1), while other authors call the group a population and the subgroups within it **subpopulations, local populations, or demes**. In this book, whenever we are discussing spatially structured populations like those depicted in Figure 2.1, we will use *metapopulation* for the group and *subpopulation* for each unit to avoid ambiguity. Most mating occurs within subpopulations, but there is at least the potential for gene flow between the subpopulations in a single

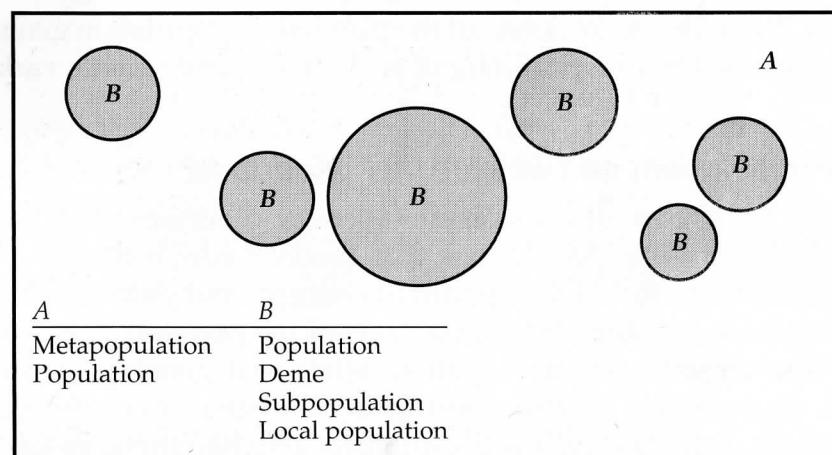


Figure 2.1 A simplified sketch of a spatially structured species. The shaded areas (denoted by *B*) are where the organisms live, whereas the area within which some gene flow occurs is denoted by *A*. *A* could potentially be as small as a cubic meter of soil for a microorganism, or perhaps as large as a continent for a migratory bird. The columns give names commonly used for *A* and *B*.

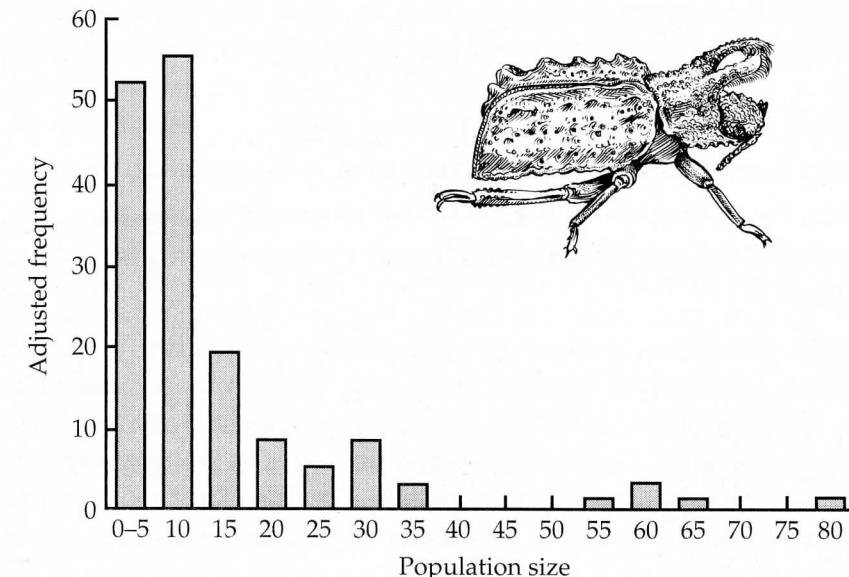


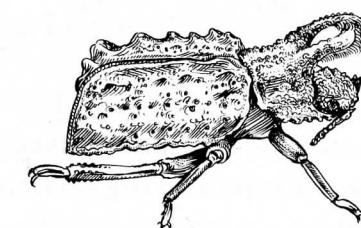
Figure 2.2 Frequency distribution showing the number of subpopulations at each size class in a metapopulation of fungus beetles, *Bolitotherus cornutus*. Most populations have very few beetles, but a few have many. (After Whitlock 1992.)

metapopulation. We will still use the word *population* in discussions not involving spatial structure.

A clear example of a structured population is given by fungus beetles in the forests of Virginia (Whitlock 1992). These beetles inhabit woody shelf fungi growing on dead trees, and only rarely move between trees. Therefore, the beetles on a group of fungi on a single tree represent a subpopulation, whereas all the beetles in a forested area are a metapopulation. There were 158 subpopulations in the metapopulation studied by Whitlock, with highly skewed subpopulation sizes (Figure 2.2); that is, two-thirds of the subpopulations had 10 or fewer beetles, but there were six subpopulations with more than 50 individuals. Subpopulation size was correlated with the age and size of fungi, with the larger subpopulations on the older and larger patches of fungi.

Genetic Variation

Genetic variation *within* a population occurs when there is more than one allele present in a population at a given locus; geneticists sometime refer to this as a population that is **segregating** or **polymorphic** at that locus. Not all loci are variable—some are **fixed**, which means that all members of the population are homozygous for the same allele. Genetic variation is ubiquitous in natural populations, but not for all traits or loci. When variation occurs



between populations of the same species it is called **genetic differentiation**. For example, two populations could be fixed for two different alleles at a given locus.

Population genetics has three interrelated goals:

1. To explain the **origin and maintenance** of genetic variation.
2. To explain the **patterns and organization** of genetic variation.
3. To understand the mechanisms that cause **changes in allele frequencies**.

The major processes studied by population geneticists are mutation, recombination, inbreeding, genetic drift, gene flow, and natural selection. Changes in allele frequencies within populations caused by natural selection can lead to adaptation. Genetic differentiation between populations can ultimately lead to the creation of new species. For these reasons, population genetics is central to the study of evolutionary change in nature.

Measuring Genetic Variation: Genetic Markers

The basic tools used to study genetic variation within and between populations are called **genetic markers**. Markers allow one to determine what alleles are present in populations, and are therefore extraordinarily useful for studying a wide variety of questions in ecology and evolution. Some of the main uses are (see Avise 1994; Mueller and Wolfenbarger 1999; Parker et al. 1998; Sunnucks 2000):

- Studying mating systems (discussed in this chapter). For example, how inbred is a population?
- Measuring gene flow and population structure (see Chapter 3). For example, how much migration occurs between subpopulations?
- Determining paternity to measure heritability (see Chapter 4) and male fitness (see Chapter 6). For example, are plants with larger flowers more successful at siring seeds?
- Producing genetic maps to find genes underlying complex traits (see Chapter 5). For example, how many loci affect body size?
- Conservation biology (see Chapter 7). For example, how genetically different are two populations of an endangered species?

In the descriptions of different types of markers discussed later in this chapter, the markers are discussed in the historical order in which they were developed. The first markers available revealed little of the underlying genetic variation; with each successive innovation more variation was revealed. This trend has culminated with gene sequencing, which reveals all the genetic variation, albeit only for a small portion of the genome for most

organisms to date. The amount of marker variation necessary depends on the specific question being asked and the populations and organisms under study. Examples of the use of most of these types of markers can be found in later chapters.

Visible polymorphisms

The first type of marker used by geneticists was **visible (discrete) polymorphisms**, in which given phenotypic traits have only a few (usually two or three) distinct types or **morphs**. For these traits the majority of phenotypic variation is due to one or sometimes two gene loci; therefore, they are not strongly affected by the environment. The traits studied by Mendel in pea plants were visible polymorphisms—for example, white vs. purple flowers and round vs. wrinkled seeds. Visible polymorphisms were the only markers available to geneticists until the 1960s, so many of the classical studies in population and ecological genetics used these markers. Some visible polymorphisms are merely used as markers and are not traits of intrinsic interest, for example, round vs. wrinkled seeds in peas, red vs. white eyes in *Drosophila*. Others are directly of interest because they may be adaptations, such as flower color, heterostyly in plants (see Chapter 5), and melanic morphs in insects (see below).

When they are available, visible polymorphisms are still very useful for a variety of studies. However, only a very small fraction of phenotypic traits are controlled by one or two loci and little affected by the environment. Therefore, visible polymorphisms are not representative of the entire genome and do not reveal enough genetic variation for most research questions.

Molecular markers

Much of modern molecular genetics is based on some type of **electrophoresis**, in which macromolecules (proteins, RNA, DNA) are separated on a gel using an electric charge. Most empirical work in modern population genetics is based on molecular markers. All of the many molecular markers are based on the same basic principles. Samples containing proteins or DNA from a number of individuals are placed separately at one end of a gel (Figure 2.3). Then an electric field is applied to the gel, and different molecules in the sample move at different rates depending on differences in size, shape, molecular weight, and electric charge. Molecules that are identical or very similar end up clustered in the gel into **bands**. Finally, the gel is treated to **visualize** the bands. There are several materials that the gels can be made of, including starch, agarose, cellulose acetate, and polyacrylamide, but in all cases the purpose of the gel is to provide a matrix for the differential movement of the different molecules. There are also many types of visualization systems, including chemical stains, radioactivity, and

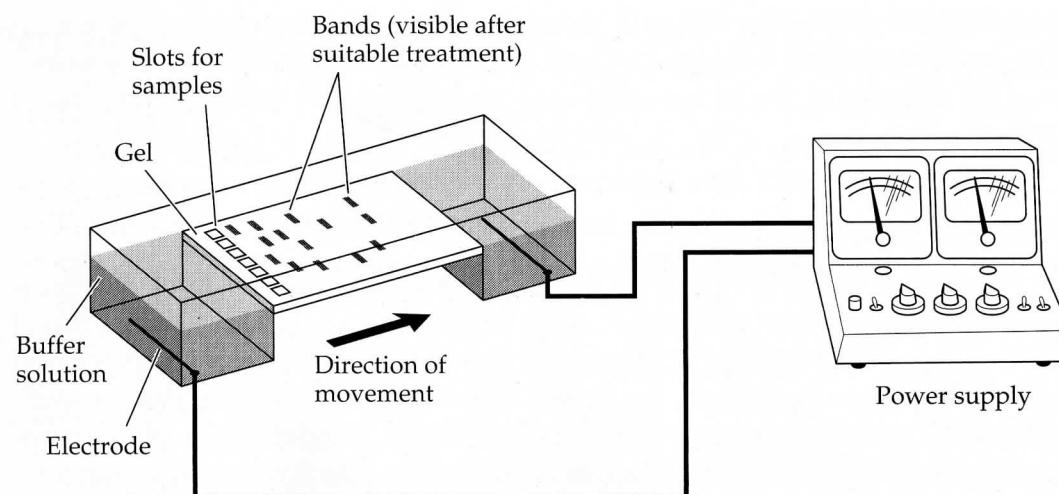


Figure 2.3 One type of laboratory apparatus for electrophoresis. Each sample slot contains material from one individual. The procedure is widely used to separate protein or DNA molecules. Proteins separate based mainly on differences in charge, while DNA fragments migrate in proportion to the logarithm of their size in base pairs (smaller fragments migrate faster).

fluorescence. For more information on many of these markers, see Hillis, Moritz, and Mable (1996).

Protein electrophoresis

The first molecular markers used in population genetics were proteins, usually soluble enzymes, also referred to as **allozymes** or **isozymes**. Lewontin and Hubby first used allozymes for population genetics in 1966, and the technique quickly revolutionized the field. For years, population geneticists had debated on theoretical grounds about how widespread genetic variation is in natural populations. Application of the allozyme technique to a variety of species revealed that variation is nearly ubiquitous in nature.

In protein electrophoresis, a part or all of the organism is typically ground up in a buffer (a solution that helps maintain a constant pH), and a sample of the buffer, now containing all of the soluble enzymes, is placed at the **origin** of the gel. After electrophoresis, a stain is used that takes advantage of the specificity of enzyme-substrate reactions to visualize only one enzyme. A variety of different stains can be used to visualize different enzymes on different gels. This solves a basic problem in all molecular marker techniques—how do you isolate one gene locus for study out of the myriad proteins and millions of bases of DNA in each cell? The allozyme staining solution typically contains the substrate and a dye that precipitates where the enzyme-catalyzed reaction takes place. When the gel is incubated

in the staining solution, a colored band is produced wherever that one enzyme occurs in the gel (see Murphy et al. (1996) for technical details).

Some amino acids are positively, and some negatively, charged. When the electric field is applied to the gel, enzymes with different charges migrate at different rates. Because different allelic DNA sequences can code for slightly different amino acid sequences in the protein, this difference in mobility reveals underlying genetic variation. Note that we are talking about differences within the same enzyme, so different bands represent different forms of the same enzyme. These different forms are caused by differences in amino acid sequence, but these variants are all functional versions of the same enzyme. Figure 2.4 shows a typical banding pattern for a simple enzyme locus with two alleles present in this sample. This shows a key advantage of enzyme electrophoresis over many other markers: In most cases it produces **codominant** markers, which means that no allele is dominant over others, and therefore heterozygotes can be distinguished from both homozygotes. This is because heterozygotes appear as two distinct bands on the gel, while homozygotes produce only one band (see Figure 2.4).

Compared to visible polymorphisms, protein electrophoresis reveals far more of the underlying genetic variation, and this variation is also a more random sample of the genome, an important feature for the uses and questions listed above. Another advantage of allozymes is that they are almost always codominant. Their biggest disadvantage is that they reveal only a small subset of the actual variation in DNA sequences between individuals, for a variety of reasons:

- Since proteins are gene products, variation in proteins does not reveal variation in non-coding regions of the genome.
- For a large number of proteins, particularly structural proteins, specific stains are not available, so only the loci that code for a subset of the proteins in an organism can be examined.

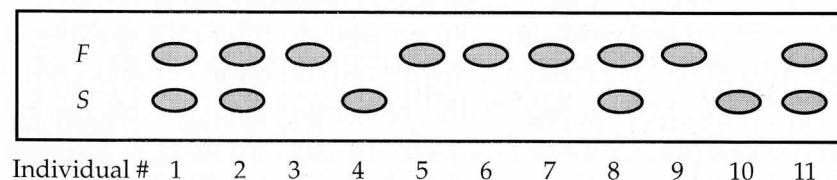


Figure 2.4 Banding pattern for a simple allozyme locus with two codominant alleles. There are 11 vertical lanes, each representing one individual. The two alleles are labeled *F* for faster migrating (the origin is at the bottom) and *S* for slower. Homozygotes appear as single bands, since the enzyme coded for by both alleles in an individual are identical and thus migrate the same distance. In heterozygotes, two bands appear, one for the enzyme coded for by each of the two different alleles present in heterozygous individuals.

- Even for proteins that can be stained, there can be many variations in amino acid sequence that do not cause differences in mobility on a gel, particularly if the overall charge of the proteins are the same.
- Many changes in the DNA sequence in a coding region do not cause changes in amino acid sequence of the protein. This is especially true of changes at the “silent,” or third position, of the codon.

These are all consequences of the fact that protein electrophoresis examines the enzyme phenotype, rather than documenting the genotype directly. Therefore, the main limitation of protein electrophoresis is that it reveals only a small fraction of the genetic variation in a population. However, it is still a useful technique in organisms with a high degree of allozyme variation and for questions not requiring large amounts of genetic variation.

DNA markers

Molecular methods for examining variation in DNA sequences directly first became available in the 1970s, and new techniques continue to be developed at a dizzying pace. In essence, all these techniques are ways of converting genetic variation into pieces of DNA of different sizes, which can then be separated by gel electrophoresis. Proteins separate mainly based on differences in charge, whereas DNA separates based on differences in the size of the fragment. There are a wide array of techniques that are changing rapidly, so we will focus on the following techniques that have gained the widest use in ecological genetics.

RFLP. The first DNA technique to gain wide use in population genetics was **DNA restriction fragment length polymorphisms (RFLPs)**. In this technique, a sample of DNA is mixed with restriction enzymes, which cut DNA at restriction sites with specific short DNA sequences (Figure 2.5A). The resulting DNA is then separated by electrophoresis and visualized with radioactively labeled or fluorescent DNA probes using a Southern blot (Figure 2.6). If the restriction site occurs rarely in the DNA sample, then the cut fragments will be large and the result will be bands that do not migrate very far during electrophoresis. If the restriction enzyme occurs more frequently in the sample, then the DNA will be cut into smaller fragments that will migrate farther on the gel (Figure 2.5B). Thus, differences in banding patterns between individuals reveals genetic variation in the number and position of restriction sites that have the specific DNA sequence recognized by the restriction enzyme. RFLPs have the advantages that they are codominant and have been developed for many organisms, but they are considerably more labor-intensive and require more DNA than the more recent techniques based on PCR.

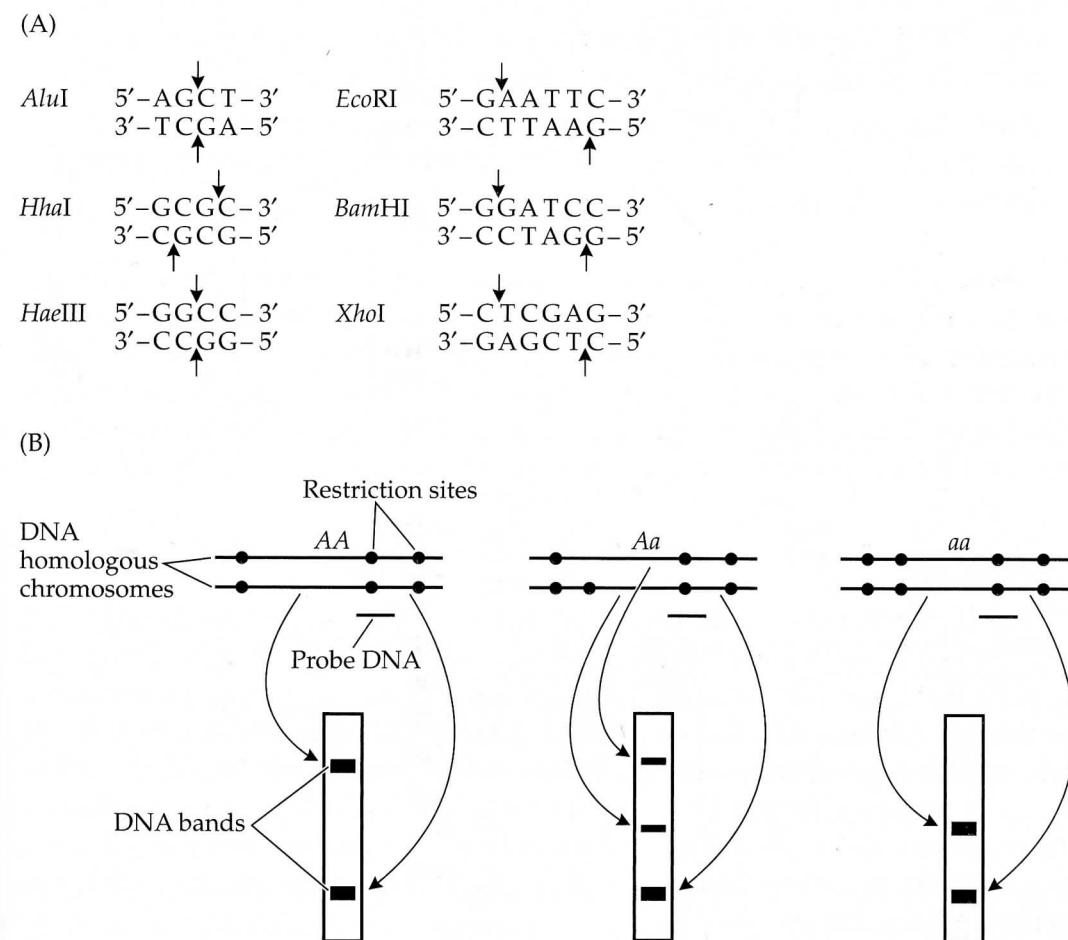


Figure 2.5 (A) The restriction sites for six of the more than 500 commercially available restriction enzymes. In each strand cleavage occurs at the position of the arrowhead. For example, the enzyme *Alu*I cuts DNA at sites containing the four-base sequence AGCT, and each strand is cleaved between the G and the C. By contrast, *Eco*RI cuts at the six-base sequence GAATTC, and each strand is cleaved between the G and the A. (B) RFLP analysis. Three genotypes are shown. At the top is a schematic of portions of a pair of homologous chromosomes with the location of restriction sites shown. The bottom shows the resulting banding patterns for each genotype on a Southern blot, with larger DNA fragments toward the top. One restriction site is missing in the *A* allele, resulting in a larger fragment than in the *a* allele. The smallest fragment is produced in both alleles. Arrows show which DNA fragment produces each band; thicker bands are those fragments present in both chromosomes (they sometimes do appear thicker on an actual gel). Only fragments that the probe DNA binds to are visualized as bands.

PCR. Almost all recent molecular markers are based on the **polymerase chain reaction (PCR)**, which has revolutionized the entire field of genetics. PCR is a simple, automated method to produce many copies of a specific

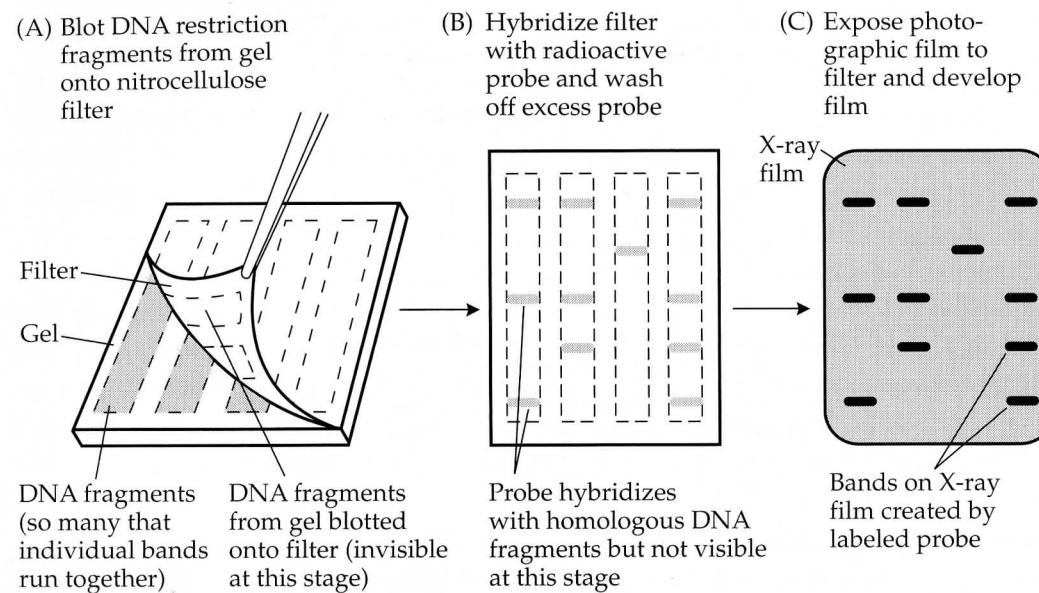


Figure 2.6 Southern blot procedure. (A) DNA fragments separated by electrophoresis are transferred and chemically attached to a filter. (B) The filter is mixed with radioactive or fluorescent probe DNA, which hybridizes with homologous DNA molecules on the filter. (C) After washing, the filter is exposed to photographic film, which develops dark bands caused by radioactive or light emissions from the probe.

fragment of DNA. To do this, short oligonucleotides (usually 20 to 30 nucleotides in length) that are complementary to the two ends of the fragment to be amplified are synthesized to be used as a primer for DNA synthesis. These are then mixed with a sample of the DNA to be studied along with DNA polymerase (to catalyze DNA synthesis) and large amounts of the four nucleotides (A, T, C, and G). This mixture is then put through repeated cycles of (1) denaturing the DNA sample into single strands to serve as a template for DNA synthesis, (2) binding (annealing) of the primers to these strands, and (3) synthesis of new copies of the DNA downstream of the primer (Figure 2.7). Since each new strand synthesized can serve as a template in the next round of synthesis, the number of copies of the fragment of interest doubles with each cycle, numerically overwhelming any other DNA in the sample.

In the PCR-based marker techniques described in the next section, it is only the target segment of DNA that is amplified. This means that the visualization step does not need to be specific as in allozymes and RFLP; any compound that stains DNA can be used, such as ethidium bromide, silver stain, or a variety of fluorescent dyes. This is a primary reason why the PCR-based markers are faster and simpler than most of the other techniques. Another advantage of PCR-based markers is that less sample DNA from each organism is needed

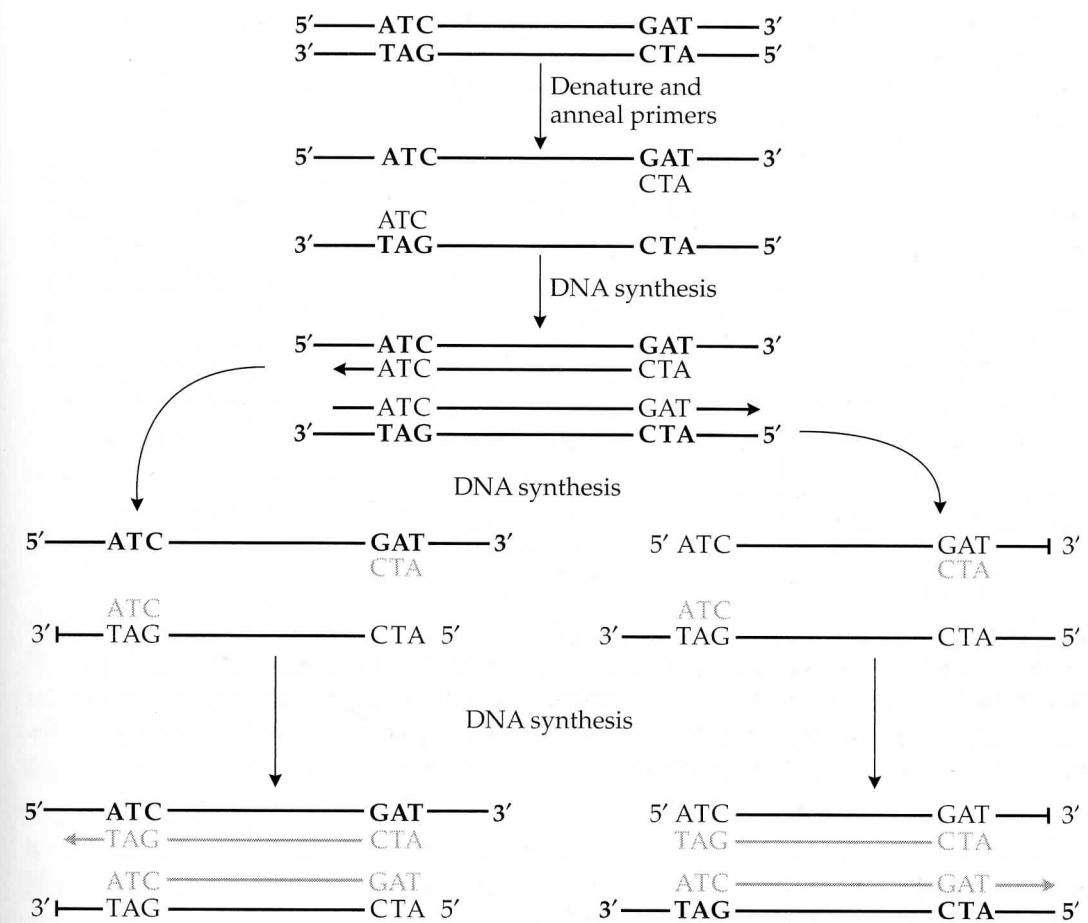


Figure 2.7 The polymerase chain reaction (PCR). PCR amplifies the fragment of DNA between the two primers exponentially, because the fragments produced in each cycle are used as templates in later cycles, leading to a doubling of the number of fragments in each cycle. The original sample DNA is shown in boldface type, the DNA synthesized in the first round in regular type, and the DNA from the second round in gray. After many cycles the vast majority of the fragments include only the primer sequences and the intervening region. For simplicity, only two cycles are shown, and primers are depicted as three bases long; in actual applications 20 or more cycles are common and primers are 10 or more bases long. Binding sites for the same primer sequence are shown close together on the two strands. Pairs of different-sequence primers for the two ends of the fragments are also used.

because it is amplified by the PCR reaction. There is a basic tradeoff between the markers discussed later in this chapter. Some, such as the VNTR and CAPS markers, have the great advantage of being codominant, but are much more difficult to develop for each new species because the PCR primer sequences must be determined. Others, such as RAPD and AFLP (to be discussed next), have the advantage of being based on random primer sequences, therefore,

specific primer sequences do not need to be identified, but they have the drawback of being dominant and so heterozygotes cannot be distinguished.

RAPD. Perhaps the simplest and quickest method of generating markers using PCR is called **randomly amplified polymorphic DNA**, or **RAPD** (Williams et al. 1990). Here, randomly generated primers, 10 bases in length, are used for PCR with the entire genome of the study organism serving as a template. If a sequence complementary to the primer(s) occurs on both strands of DNA, and they are close enough together, then this fragment will be amplified and appear as a band on the gel when the products of the PCR reaction are electrophoresed (see Figure 2.7). If the binding sites on the two strands are too far apart, then synthesis of the fragment will not be completed when the synthesis phase of PCR ends, and the fragment will therefore not be amplified.

A panoply of different markers have come from the use of RAPD technology. If a fragment can be amplified using a single primer, then it is called a **SPAR (single primer amplified region)**. When a particular amplified fragment is isolated and its nucleotide sequence determined, it becomes a **SCAR (sequence-characterized amplified region)**. A particular amplified fragment may also include a restriction-site polymorphism, in which case the polymorphism is called a **CAPS (cleaved amplified polymorphic site)**. A CAPS is the analog of an RFLP, except that the genotype is identified by amplification and a nonspecific DNA stain, rather than in a Southern blot with a radioactive probe. This is why a CAPS is sometimes called a **PCR-RFLP**. CAPS alleles, like RFLP alleles, are codominant.

AFLP. This technique is a hybrid of RFLP and RAPD that combines many of the advantages of each. In **AFLP** (sometimes called **amplified fragment length polymorphism**, although it does not reveal length polymorphisms), the genome of the study organism is cut with restriction enzymes as in RFLP, and then some of these fragments are selectively amplified with PCR using random sequences as in RAPD (see Mueller and Wolfenbarger 1999 for details). Since AFLPs are based on PCR, they are faster and easier than RFLPs. They are more repeatable than RAPDs because the primers used for amplification are typically about twice as long, so errors in binding to the template DNA are less frequent.

Both RAPD and AFLP reveal genetic variation in a population by screening for the presence or absence of a specific random (or partially random) DNA sequence, rather than different lengths of fragments as in the other DNA markers. In RAPD, if there is a location in the genome with two sites on opposite strands that the random primers bind to, then this fragment is amplified and a band is formed on the gel. If another individual in the population has a different DNA sequence at one of these binding sites, then the

primer will not bind there and no band will be formed. Similarly in AFLP, if a fragment produced by digestion with restriction enzymes has sequences at both ends matched by the primers, including the random portions, then that fragment will be amplified. In individuals with a different sequence at one or both ends, the fragment will not be amplified and no band will be produced. One consequence of this is that most RAPD and AFLP markers are inherited in a dominant fashion, meaning that heterozygotes and one of the homozygotes produce the same pattern. Heterozygotes produce a single band in exactly the same position as the band produced by one of the homozygotes (although there is twice as much DNA in the band from homozygotes). The other homozygote produces no band, which is often called a **null allele**. Therefore, only the band-absent (null) genotype can be scored definitively. Each band on a RAPD or AFLP gel is from a separate location on the genome, instead of having different bands representing different alleles at the same locus as in the other methods.

VNTRs (variable number of tandem repeats). VNTRs are noncoding regions of the genome that consist of several to many copies of the same sequence. The repeated sequence can be anywhere from two to 64 nucleotides long. Longer sequences (from 10 to 64 nucleotides in the repeating unit) are called minisatellites, whereas shorter repeats (2 to 9 nucleotides) are called microsatellites, simple sequence repeats (SSR), simple sequence repeat polymorphisms (SSRP), or short tandem repeats (STR). There is a large amount of variation between individuals for the number of repeat units at each VNTR locus, especially for the shorter repeat units. For example, a common microsatellite repeat unit is AC, so one allele might have this unit repeated 10 times, ACACACACACACACACAC, which is often written $(AC)_{10}$, and another allele might be $(AC)_{12}$. Note that these are double-stranded nuclear DNA, but for simplicity only the sequence of one of the strands is used to denote the marker.

For minisatellites, the DNA is cut with a restriction enzyme, and then a radioactively labeled DNA probe that has a complementary sequence to the repeated unit is used for visualization (similar to RFLP). For microsatellites, the regions on either side of the repetitive sequence (the **flanking regions**) are sequenced, and then complementary primers are constructed so that the microsatellite can be amplified by PCR and visualized with stain or fluorescence, similar to RAPDs. The different number of repeat units in different alleles causes each fragment to migrate a different distance on the gel, so genetic variability is revealed. Multilocus probes are often used for minisatellites, and microsatellite loci can have as many as 30 to 50 different alleles in a population, so these techniques can reveal a tremendous amount of variation. Indeed, these markers are often referred to as **DNA fingerprints**, because they are rarely exactly the same in two individuals. The main drawback is that these

TABLE 2.1 Summary of genetic markers used most often in ecological genetics

Marker	Codominant	Amount of variation revealed
Visible polymorphisms	Sometimes	Low
Allozymes	Yes	Low-moderate
RFLP	Yes	Moderate
RAPD	Rarely	High
AFLP	Rarely	High
VNTR	Yes	High

markers are difficult to develop for new species, because the VNTR regions must be found and the sequences of the flanking regions determined.

The ultimate level of resolution of genetic variation is **gene sequencing**. Sequencing is usually done by automated sequencing machines, and the entire genomes of a number of organisms (including humans) have been or are being sequenced. However, it is still too time consuming and expensive to use sequencing for examining genetic variation within most populations, since this usually requires sequencing a number of genes in a number of individuals. Another drawback for the application of sequencing to ecological genetics is that in most cases little is known about how specific sequences affect the phenotype. Sequencing has become the method of choice for reconstruction of phylogenies, however, since only one or a few individuals need to be sequenced for each taxon sampled.

Table 2.1 summarizes the different types of markers discussed above. Currently, microsatellites and AFLP are the most commonly used in ecological genetics, although many examples in later chapters are based on visible polymorphisms and allozymes, because there is a larger literature on these and they provide simpler examples. Microsatellites are preferred if a sufficient number of primer sequences are known or can be generated for the organism under study; they are codominant and can have many alleles per locus in a population. AFLP markers are based on random sequences so, unlike microsatellites, they take very little time to develop for each species, a key advantage for studies of nonmodel organisms. Another advantage for AFLP is that many loci can be scored on each gel. The disadvantages of AFLP are that they are dominant and have only two alleles per locus.

Measuring Genetic Variation: Simple Summary Statistics

The genetic variation uncovered by genetic markers can be quantified in a number of ways. The most fundamental measures in population genetics are

allele and genotypic frequencies. The word *frequency* in population genetics is used to mean proportion, so the **frequency** of a given allele or genotype is simply its proportion out of all the alleles or genotypes present at the locus in the population. The term *gene frequency* is often used interchangeably with *allele frequency*; this is unfortunate, because it is inaccurate and leads to confusion. Genes do not differ in frequency, alleles do.

If we have only two alleles at a locus, designated *A* and *a*, then the usual symbols for their frequencies are *p* and *q*, respectively. Since frequencies (proportions) must sum to 1, then

$$p + q = 1$$

$$q = 1 - p$$

2.1

These equations are very simple, but they are extremely important to keep in mind when thinking about allele frequencies. Genotypic frequencies are usually denoted by uppercase letters, and they also must add to 1. In the case of a locus with just two alleles, *P* is the frequency of *AA*, *H* is the frequency of *Aa*, and *Q* is the frequency of *aa*. These must also sum to 1; that is, *P* + *H* + *Q* = 1. To calculate allele frequencies from genetic marker data, divide the number of each allele by the total number of alleles in the sample:

$$p = \frac{2n_{AA} + n_{Aa}}{2n}$$

2.2

where *n* = the number of individuals in the sample and *n_{AA}* and *n_{Aa}* are the numbers of *AA* and *Aa* genotypes in the sample, respectively. In Equation 2.2 the 2 in the numerator is present because there are two *A* alleles in each homozygous individual, and the 2 in the denominator is present because there are two alleles at each locus in each diploid individual.

We can rearrange Equation 2.2 to obtain the allele frequency in terms of the genotype frequencies:

$$\begin{aligned} p &= \frac{2n_{AA}}{2n} + \frac{n_{Aa}}{2n} \\ p &= \frac{n_{AA}}{n} + \frac{1}{2} \frac{n_{Aa}}{n} \\ p &= f(AA) + \frac{1}{2} f(Aa) \\ &= P + \frac{H}{2} \end{aligned}$$

2.3

Here *f(AA)* stands for the frequency (proportion) of homozygotes.

EXERCISE: Write the equations for q in terms of both number and frequency of genotypes.

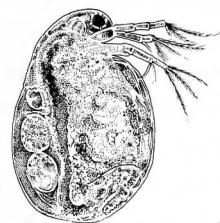
ANSWER:

$$q = \frac{2n_{aa} + n_{Aa}}{2n} \quad 2.4$$

$$q = f(aa) + \frac{1}{2} f(Aa)$$

$$= Q + \frac{H}{2} \quad 2.5$$

As an example, we can calculate allele frequencies in a natural population of water fleas (*Daphnia obtusa*; Spitze 1993). In the Ojibway Pond population, the following genotypes were scored at the allozyme locus *PGM*: 57 individuals were *MM*, 53 *MS*, and 18 *SS* (*M* and *S* stand for medium and slow). Using Equations 2.2 and 2.4, and p for the frequency of *M*, we find:



$$p = \frac{(2 \times 57) + 53}{2 \times 128} = 0.652$$

$$q = \frac{(2 \times 18) + 53}{2 \times 128} = 0.348$$

Note that p and q sum to 1, and that 128 is the total number of individuals in the population (57+53+18).

EXERCISE: Calculate the allele and genotype frequencies in Figure 2.4.

ANSWER: There are five *FF* homozygotes, two *SS* homozygotes, and four *FS* heterozygotes. Using Equations 2.2 and 2.4, $p = 0.636$ and $q = 0.364$. Always check that $p + q = 1$, which it does here. Genotypic frequencies $P = 5/11 = 0.454$, $H = 2/11 = 0.182$, $Q = 4/11 = 0.364$, which also sum to 1.

Organization of Genetic Variation within Populations

In this section we discuss the different ways that mating systems organize alleles into diploid genotypes (Figure 2.8). Under different mating systems (e.g., random mating vs. inbreeding) different genotypic frequencies are generated from the same allele frequencies.

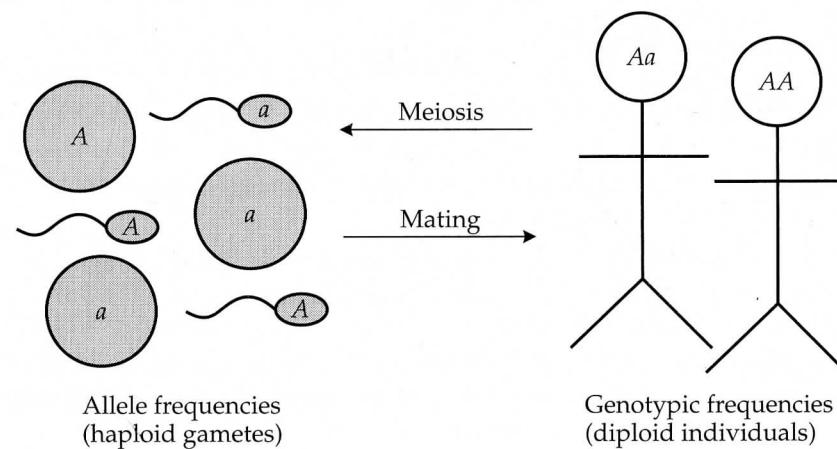


Figure 2.8 A schematic showing how mating combines alleles (in the haploid pool of gametes) into genotypes (in the diploid individuals). The genotypic frequencies produced by a given set of allele frequencies depends on the mating system. The cycle repeats itself when the diploid individuals produce haploid gametes through meiosis.

Random mating

Our starting point will be the simplest case, **random mating**. When mating is random, the chance that an individual mates with another individual of a given genotype is equal to the frequency of that genotype. A more precise definition comes from probability: Under random mating, the frequency of matings between two genotypes is equal to the product of their individual frequencies. When mating is random and a number of other assumptions are met, then the genotypic frequencies for a given set of allele frequencies can be calculated with the **Hardy-Weinberg model**. Godfrey Hardy and Wilhelm Weinberg each independently published the model in 1908, laying the foundation for the field of population genetics. An introduction to models is presented in Box 2.1.

Assumptions of the Hardy-Weinberg model

The Hardy-Weinberg model has eight assumptions. They are listed here along with some brief explanations.

1. The organism is **diploid**, **sexual**, and has **discrete generations**. Discrete generations refer to a life history like that of an annual plant, in which the parental generation has died by the time the offspring generation reproduces.
2. Allele frequencies are the same in both sexes.
3. **Mendelian segregation** occurs, which means that heterozygotes produce equal numbers of gametes containing each allele. For example, an

BOX 2.1 What Is a Model?

A **model** is a theoretical abstraction of the real world that has two principal uses:

- To reduce complexity, allowing us to see important underlying patterns; that is, to see the forest in spite of the trees.
- To make specific predictions to test with experiments or observations. Thus, models can guide **empirical** studies by suggesting which data are most important to gather.

However, it is important to remember that models cannot provide direct information about what is actually occurring in the real world. The predictions made by models need to be supported or refuted by empirical data.

There are three main types of models—verbal, simulation, and analytical. **Verbal models** tend to take the form of “if this condition holds, then logically this should happen.” Although some people have a low opinion of verbal models because they lack mathematical rigor, these models can be extraordinarily powerful. Darwin’s theory of natural selection was a verbal model, yet it revolutionized biology. Darwin’s verbal model stated that if there is variation in a trait that is related to fitness, and this variation is heritable, then the trait will change across generations; that is, it will evolve.

In **simulation models** the system to be modeled is simulated in a computer. For example, the mating system diagrammed in Figure 2.8 could be simulated in the computer, starting with frequencies of different alleles in the gametes. The computer could then be instructed to combine gametes randomly for random mating and calculate the resulting diploid genotypic frequencies.

Analytical models define the entire system with equations that can then be solved for different values of the input variables to make predictions about the behavior of the system. Analytical models are the most difficult to construct, but can also be the most powerful. For example, the Hardy-Weinberg model is an analytical model of allele and genotypic frequencies under random mating.

All models make starting **assumptions**, either explicitly or implicitly, to simplify the system. These assumptions are partly to make the model mathematically or computationally tractable, especially in the case of analytical models, but also to make the models easier to understand—remember that simplification of the real world is one of the primary goals of models. It is important to consider the assumptions of any model carefully, because they determine how well the model fits the real world. This does not mean that a model is worthless for understanding an organism that violates some of the model’s assumptions, because very often the predictions of a model will be **robust** to these violations. To be *robust* means that the predictions do not change much when some of the assumptions are violated.

Aa individual produces equal numbers of *A* and *a* gametes. There are a few genes that violate this assumption; this condition is known as **meiotic drive** or **segregation distortion**. When meiotic drive occurs, one allele in heterozygous individuals is overrepresented in the gametes.

Examples include the *T* allele in mice (*Mus*) and the *segregation distorter* locus in *Drosophila*.

4. **Random mating** occurs, meaning that mating is random with respect to the genotypes under consideration (it may be nonrandom with respect to genotypes at other loci). Later in this chapter we will discuss two types of nonrandom mating, assortative mating and inbreeding.
5. **No mutation**
6. **No migration**
7. **No random genetic drift (large population size)**
8. **No natural selection**

The Hardy-Weinberg model is robust to many violations, especially of the last four assumptions. After considering allele and genotype frequencies under all these assumptions, we will **relax** (remove) one or more of the last five assumptions and examine the consequences of each of these in turn. In fact, much of population genetics involves study of the consequences of relaxing these last five assumptions, to better understand how mating systems, mutation, migration, drift, and natural selection affect the genetics and evolution of populations.

Deriving the Hardy-Weinberg equations

Recall our previous comment that we will often want to calculate genotypic frequencies from allele frequencies and vice versa. Given the assumptions above, this conversion back and forth is simple, and is the essence of the **Hardy-Weinberg principle**: Genotypic frequencies in the offspring can be calculated directly from allele frequencies in the parents after only one generation of random mating. There are two principal ways of deriving the Hardy-Weinberg equations, and each provides insight into how genotypic frequencies arise under random mating.

The first derivation starts with allele frequencies in the gametes produced by parents. Given the assumptions above, the frequencies of alleles in the pool of gametes produced by the parents are the same as the allele frequencies in the parental population. With random mating, these gametes unite randomly during fertilization, so that the genotypic frequencies of the offspring are the products of the gametic allele frequencies. Figure 2.9 shows cross-multiplication squares drawn with the sides of the squares proportional to the allele frequencies. Because the areas of the squares are the products of the lengths of their sides, the areas are proportional to the offspring genotypic frequencies. Note that there are two ways to make a heterozygote—an *A* sperm and an *a* egg or an *a* sperm and an *A* egg—so that the genotype frequencies of those two boxes are added together.

Therefore, under Hardy-Weinberg assumptions the frequency of *AA* homozygotes (P' , where the prime denotes the offspring generation) is p^2 ,

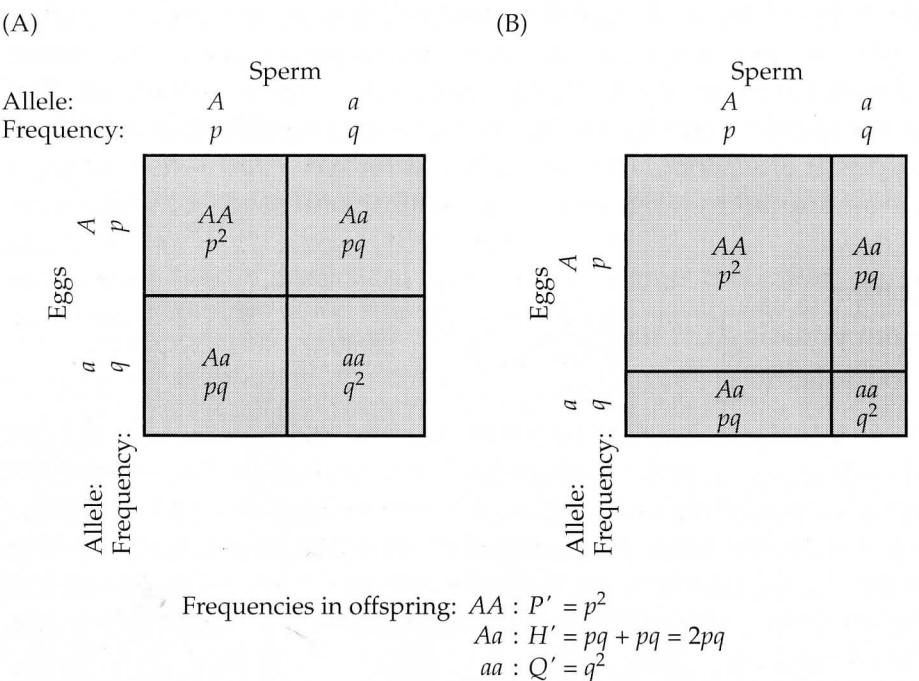


Figure 2.9 Cross-multiplication squares showing graphically how genotypic frequencies (inside boxes) are products of allele frequencies in gametes under Hardy-Weinberg assumptions. In the square on the left, the allele frequencies are equal (i.e., $p = q = 0.5$), while in the right-hand square the A allele is three times as common as a ($p = 0.75$; $q = 0.25$).

the frequency of aa (Q') is q^2 , and the frequency of heterozygotes (H') is $2pq$. Figure 2.9A shows the case where the two alleles are at equal frequency ($p = q = 0.5$), so that the frequencies of the homozygotes are each 0.25 and the frequency of the heterozygote is 0.5.

Now we can determine the allele frequencies in the offspring using equation 2.3:

$$\begin{aligned}
 p' &= P' + \frac{H'}{2} \\
 &= p^2 + \frac{(2pq)}{2} \quad \text{substitute from Hardy-Weinberg equation} \\
 &= p^2 + pq \quad \text{cancel} \\
 &= p(p+q) \quad \text{factor} \\
 &= p \quad \text{remember } p+q=1
 \end{aligned}$$

EXERCISE: Derive q' in terms of q .

These derivations show that under Hardy-Weinberg assumptions, allele and genotypic frequencies remain the same across generations. This means that Mendelian inheritance, by itself, does not change allele frequencies. This is why geneticists often refer to the **Hardy-Weinberg equilibrium (HWE)**. If the genotypic frequencies are changed without changing the allele frequencies, then the genotypic frequencies will return to the Hardy-Weinberg values (p^2 , $2pq$, q^2) after one generation of random mating. If some other force changes the allele frequencies (see Chapter 3), however, a new HWE occurs with genotypic frequencies corresponding to the new allele frequencies (i.e., new p and q produce new p^2 , $2pq$, q^2), again after one generation of random mating.

An example of this is given in Figure 2.9B. Here $p = 0.75$ and $q = 0.25$. As long as the assumptions of Hardy-Weinberg are met, the genotypic frequencies are still p^2 , $2pq$, and q^2 , which with the new allele frequencies are approximately 0.562, 0.375, and 0.062. This illustrates a key point: Hardy-Weinberg equilibria can occur for any allele frequency. It is a common misconception that HWE means equal allele frequencies, $p = q = 0.5$, but even if one allele is fixed the population is in HWE (p , p^2 , and P all equal 1; q , $2pq$, H , q^2 , and Q all equal 0).

Figure 2.10 shows the frequencies of all three genotypes under HWE at a locus with two alleles for all possible allele frequencies. This graph helps us understand a wide variety of fundamental concepts in population genetics;

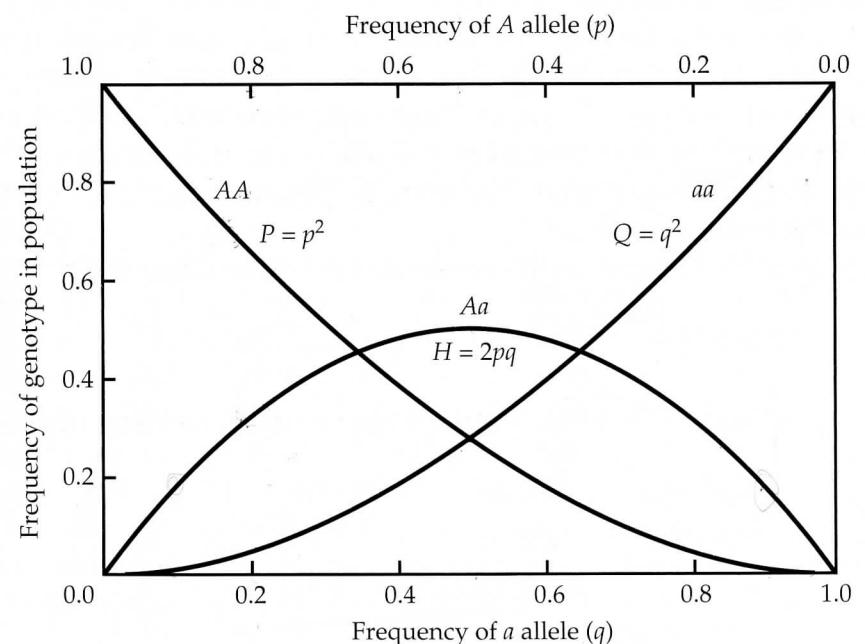


Figure 2.10 Genotypic frequencies across all allele frequencies for two alleles under Hardy-Weinberg equilibrium. Note that when an allele is rare, the frequency of homozygotes for that allele is much lower than the frequency of heterozygotes.

we will refer to it often. There are three important points to note. First, these are the genotypic frequencies only at HWE; factors like inbreeding that cause deviations from HWE will change these genotypic frequencies. Second, at intermediate allele frequencies, heterozygotes are more common than either homozygote. Third, when the frequency of an allele is less than 0.1, virtually all of these alleles are contained in heterozygotes and there are very few homozygotes for the allele. As we will see in more detail in Chapter 3, this is how deleterious recessive alleles (for example, lethal disease alleles in humans) are maintained in the population. Because selection maintains the harmful recessive alleles at a low frequency, almost all of the harmful alleles are contained in heterozygous individuals, in which the recessive disease is not expressed.

Another way to derive the Hardy-Weinberg equations is to start with the matings between the diploid parents rather than with the gametes they produce. With three genotypes there are six possible types of matings. For example, the mating $AA \times AA$ occurs only when an AA male mates with an AA female, and this occurs a proportion $P \times P$ (or P^2) of the time. Similarly, an $AA \times Aa$ mating occurs when an AA female mates with an Aa male (proportion $P \times H$), or when an Aa female mates with an AA male (proportion $H \times P$)—so the overall proportion of $AA \times Aa$ matings is $PH + HP = 2PH$. The frequencies of these and the other types of matings are given in the second column of Table 2.2.

The offspring genotypes produced by the matings are given in the last three columns of Table 2.2. The offspring frequencies follow from the assumption of Mendelian segregation, so that an Aa heterozygote produces

TABLE 2.2 Demonstration of the Hardy-Weinberg principle

Mating	Frequency of mating	Offspring genotype frequencies		
		AA	Aa	aa
$AA \times AA$	P^2	1	0	0
$AA \times Aa$	$2PH$	$\frac{1}{2}$	$\frac{1}{2}$	0
$AA \times aa$	$2PQ$	0	1	0
$Aa \times Aa$	H^2	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$
$Aa \times aa$	$2HQ$	0	$\frac{1}{2}$	$\frac{1}{2}$
$aa \times aa$	Q^2	0	0	1
Totals (next generation)		P'	H'	Q'

where: $P' = P^2 + 2PH/2 + H^2/4 = (P + H/2)^2 = p^2$
 $H' = 2PH/2 + 2PQ + H^2/2 + 2HQ/2 = 2(P + H/2)(Q + H/2) = 2pq$
 $Q' = H^2/4 + 2HQ/2 + Q^2 = (Q + H/2)^2 = q^2$

an equal number of A -bearing and a -bearing gametes. Homozygous AA genotypes produce only A -bearing gametes, and homozygous aa genotypes produce only a -bearing gametes. Therefore, a mating of AA with aa produces all Aa offspring, a mating of AA with Aa produces $\frac{1}{2} AA$ and $\frac{1}{2} Aa$ offspring, a mating of Aa with Aa produces $\frac{1}{4} AA$, $\frac{1}{2} Aa$, and $\frac{1}{4} aa$ offspring, and so forth.

The genotype frequencies of AA , Aa , and aa after one generation of random mating are denoted in Table 2.2 as P' , H' , and Q' respectively. The new genotype frequencies are calculated as the sum of the products shown at the bottom of the table. For each genotype, the frequency of each mating producing the genotype is multiplied by the fraction of that genotype produced by that mating. The new genotype frequencies P' , H' , and Q' simplify to p^2 , $2pq$, and q^2 , the Hardy-Weinberg frequencies.

Box 2.2 introduces some of the basics of statistical analysis, which we will use throughout the rest of the book.

Uses of the Hardy-Weinberg Model: Tests for Departure from HWE

It is often useful to test whether genotypic frequencies in a real population are in HWE, because if they differ significantly, this suggests that one or more assumptions of HWE have been violated and therefore a potentially important evolutionary or ecological process is causing the deviation. The converse is not true, however, because HWE is robust to violations of assumptions and the chi-square test is not particularly powerful. Therefore, finding no significant deviation from HWE does not mean that none of the assumptions of HWE have been violated.

As an example, we will use the chi-square test to determine whether Spitzer's (1993) PGM allozyme genotype frequencies from *Daphnia*, discussed on page 24, deviate significantly from HWE. The first step is to determine the numbers of each genotype *expected* under HWE given the allele frequencies that we calculated above. We will then use the chi-square test to see how closely the *observed* numbers of the three genotypes fit this expectation. With allele frequencies $p = 0.652$ and $q = 0.348$, the expected genotypic frequencies are $p^2 = 0.425$, $2pq = 0.454$, and $q^2 = 0.121$ (which sum to 1). Multiplying each of these by the sample size (128 individuals) gives the expected numbers as 54.4, 58.1, and 15.5. This conversion is necessary because the chi-square test must be based on the observed *numbers*, not frequencies. (Using genotypic frequencies instead of numbers in the chi-square test is a common error.) The comparison is thus between the observed (*obs*) and expected (*exp*) numbers:

	MM	MS	SS	Totals
<i>obs</i>	57	53	18	128
<i>exp</i>	54.4	58.1	15.5	128

BOX 2.2 A Brief Introduction to Statistics

Ecological geneticists use statistics to find patterns among the sometimes bewildering variability of the natural world. Thus, the goals of statistical analysis are similar to the goals of modeling, but while modeling is based on theoretical abstractions, statistics is a way of understanding empirical data collected in the real world. These data are usually collected from a random sample of a population, and statistical tests are used to determine if patterns in the sample are due to random chance or some real biological phenomenon. The larger the sample size (often denoted by n or N), the more confident we can be that the patterns are real.

We will discuss three basic kinds of statistical analyses in this book, which are designed to answer three different types of questions. In part, these types of questions depend on the kind of data that are collected. Data can be either **discrete** (also called **qualitative** or **categorical**) or **continuous (quantitative)**. Discrete data can be easily grouped into distinct groups or classes. All the genetic markers we discussed above, both visible and molecular, produce distinct genotypes that are good examples of discrete data. Most phenotypic traits, however, are continuously distributed; that is, there are not a few distinct types. Examples include virtually any measure of size of most traits of any organism, such as weight, height, length, or width. These traits are called quantitative because they need to be measured; they will be discussed further in Chapters 4–6. The three types of questions and corresponding statistical analyses are:

1. Is there a difference between two or more discrete groups in some continuously variable character? For example, are there differences between male beetles in the mass of the offspring they father? Are there differences in flower shape between several populations of plants when grown in a common environment? Statistical tests used for these kinds of questions include *t*-tests and analysis of variance (ANOVA); we will discuss these techniques in Chapter 4.
2. Is there a relationship between two continuous variables? Examples of these kinds of questions are: Is the tarsus length of a bird related to how many offspring it produces? Are the leg lengths of parent and offspring frogs correlated? These types of questions are often addressed using the statistical tools of correlation and regression (see Chapters 4 and 5).
3. Does a pattern in a population deviate from that expected from a given model? This is often called a goodness-of-fit test. An example of this is a test of whether genotypic frequencies in a real population deviate from Hardy-Weinberg expectations. A common way to do this is with a chi-square test, described in the main text below.

All statistical methods produce a test statistic. Typically, the larger the test statistic, the more confidence we have that the difference or relationship we are testing was not caused merely by a chance occurrence in our sample, but rather reflects a real pattern in the population. This level of confidence is quantified by the *P*-value, which is the probability that this difference or relationship could

Box 2.2 continued

have been produced by chance. Therefore, the smaller the *P*-value is, the more confident we are that we have a real difference or relationship.

At what *P*-value do we conclude that a pattern is real? By convention, tests with *P*-values less than 0.05 are judged statistically significant, meaning we have confidence that they are real and not just due to chance. However, a *P*-value of 0.05 means we have a 5% chance of being wrong when we conclude there is a real pattern. Also, there is nothing magic about 0.05, and in reality we have almost the same confidence in a result with a *P*-value of 0.06 as one with a *P*-value of 0.04. The bottom line is that the lower the *P*-value is, the more confidence we have that the result reflects a real difference or relationship. The larger the sample size, the lower the *P*-value will be on average for the same magnitude of pattern, because we have higher confidence when we have sampled more of the population.

(Calculating the totals for the observed and expected numbers is a useful crosscheck.) Thus, the fit to HWE in the observed is close but not exact. The *P*-value from the chi-square test will tell us how likely it is that this amount of deviation from HWE could be due to chance alone. In comparisons of this type, the value of the chi-square test statistic is calculated as:

$$\chi^2 = \sum \frac{(obs - exp)^2}{exp} \quad 2.6$$

where the summation sign Σ means summation over all classes of data, in this case all three genotypes. The resulting value of

$$\chi^2 = \frac{(57 - 54.4)^2}{54.4} + \frac{(53 - 58.1)^2}{58.1} + \frac{(18 - 15.5)^2}{15.5} = 0.98$$

is the test statistic.

Associated with any χ^2 value is a second number called the **degrees of freedom (d.f.)**. In general, the number of degrees of freedom associated with a χ^2 equals the number of classes of data (in this case, 3 genotypes) minus 1 (because the observed and expected totals must be equal), minus the number of parameters estimated from the data (in this case, 1, because the parameter p was estimated from the data). Thus, the number of degrees of freedom for our chi-square value is $3 - 1 - 1 = 1$. (Note: A degree of freedom is not deducted for estimating q because of the relation $q = 1 - p$; that is, once p has been estimated, the estimate of q is automatically fixed, so we deduct just the one degree of freedom corresponding to p .)

The actual assessment of goodness of fit is determined from Figure 2.11. To use the chart, find the value of χ^2 along the horizontal axis; then move

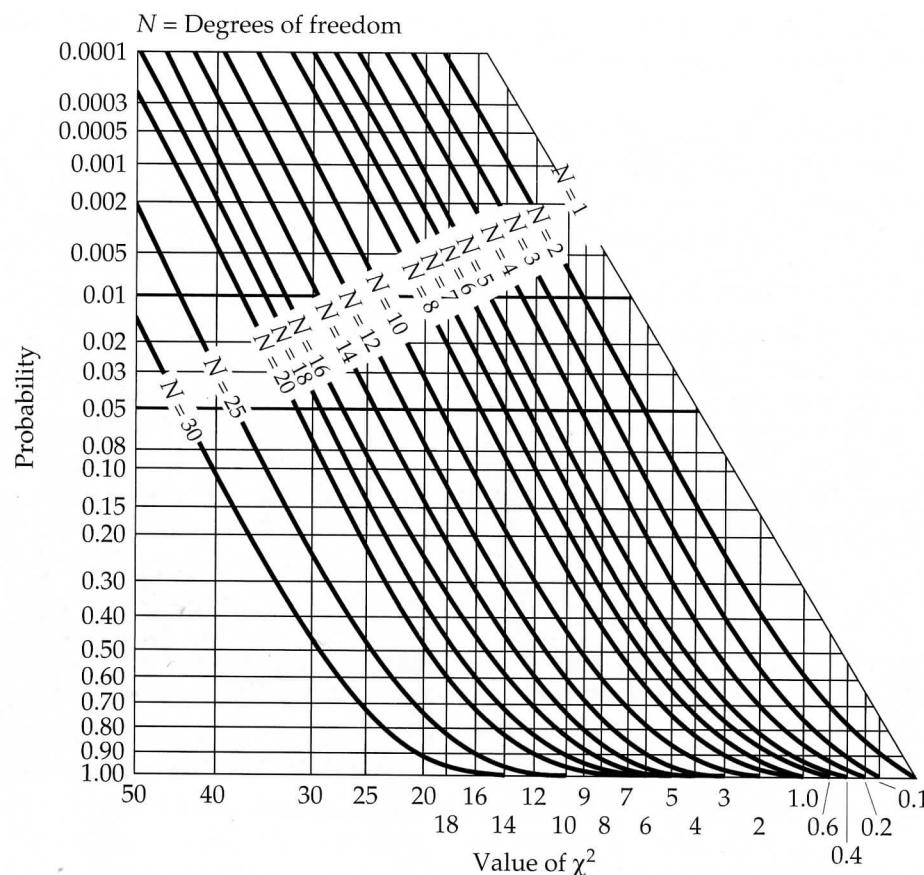


Figure 2.11 Graph of χ^2 . To use the graph, find the value of χ^2 along the x (horizontal) axis, then move straight up until you intersect the diagonal line corresponding to the appropriate number of degrees of freedom. The P -value is read from the y (vertical) axis at that intersection point. (Courtesy of James F. Crow.)

vertically from this value until the proper degrees-of-freedom line is intersected; then move horizontally from this point of intersection to the vertical axis and read the corresponding probability value. In the present case, with $\chi^2 = 0.98$ and one degree of freedom, the corresponding probability value is about $P = 0.36$. This P -value represents the probability that chance alone could produce a deviation between the observed and expected values at least as great as the deviation actually obtained. Thus, if the probability is large, it means that chance alone could account for the deviation, and it strengthens our confidence in the validity of the model used to obtain the expectations—in this case HWE. On the other hand, if the probability associated with the χ^2 is small, it means that chance alone is not likely to lead to a deviation as large as actually obtained, and it undermines our confidence in the validity of the model. Because the probability in the *Daphnia* example is 0.36, which means there is a 36% probability that the deviations from HWE

could have been due to chance, we have no reason to reject the hypothesis that the genotype frequencies are in HWE for this gene locus.

EXERCISE: At the *PGI* locus in the Ojibway Pond population of *Daphnia*, Spizte found two alleles, *S* and *S-*, and the number of individuals with each genotype was 42 *SS*, 48 *SS-*, and 38 *S-S-*. Test these for deviation from HWE.

ANSWER: Allele frequencies (from Equations 2.2 and 2.4) are $p = 0.516$ and $q = 0.484$ (which sum to 1). Expected genotypic frequencies are $p^2 = 0.266$, $2pq = 0.500$, and $q^2 = 0.234$ (which sum to 1).

Multiplying by the sample size gives the expected numbers: 34.0, 64.0, and 30.0. The χ^2 test statistic calculated from the observed and expected numbers using Equation 2.6 is 8.0. With one *d.f.*, the P -value for this chi-square is 0.005, which gives us confidence that there is a real deviation from HWE. There is a deficit of heterozygotes, and an equal excess of homozygotes; inbreeding, as well as other factors, can cause this pattern. (See the section on Nonrandom Mating later in this chapter.)

Recessive alleles hidden in heterozygotes

HWE sometimes helps solve the dilemma that arises when studying dominant markers, such as many visible polymorphisms, RAPDs, and AFLPs. The dilemma is that dominant homozygous and heterozygous genotypes cannot be distinguished, and so the genotypic and allele frequencies cannot be estimated directly. The solution is that, if one is willing to assume HWE, then the allele frequencies can be estimated anyway. The trick is to use the observed frequency of recessive homozygotes to estimate the q^2 term of the HWE. That is, if Q is the frequency of homozygous recessive genotypes in a sample, then q is estimated as the square root of Q because the expected value of q^2 equals Q . Since HWE is robust to violations of many of the assumptions, this estimate can work in real populations even though some of the assumptions are likely to be violated. However, many loci in natural populations are not in HWE, and there is no way to know this with a dominant marker.

An example of using HWE to estimate allele frequencies with a dominant marker is in **industrial melanism**, which may be the best-known example of evolution in action and is a classic example in ecological genetics as well (Majerus 1998). The term *industrial melanism* refers to an observed increase in the frequency of dark-pigmented (melanic) morphs of a number of insect

species, resulting from coal-burning during the industrial revolution that blackened insect resting places with soot. In most cases, the melanic color pattern is due to a single dominant allele. In one study of a heavily polluted area near Birmingham, England, Kettlewell (1956) observed a frequency of 87% melanic peppered moths (*Biston betularia*). Therefore, the frequency of recessive homozygotes is 0.13, and assuming HWE, the frequency q of the recessive allele is estimated as the square root of 0.13, or 0.36.

EXERCISE: Assuming HWE, calculate the frequency p of the dominant allele and the frequencies of the three genotypes.

ANSWER: $p = 1 - q = 0.64$, and the frequencies of the genotypes are $P = 0.41$, $H = 0.46$, and $Q = 0.13$ (p^2 , $2pq$, q^2). It does not make sense to try to test the fit of these genotypic frequencies to Hardy-Weinberg frequencies, because they were calculated using the Hardy-Weinberg formula, so the fit must be perfect.

Nonrandom Mating

For the rest of this Chapter and the beginning of the next, we will be relaxing one assumption of Hardy-Weinberg at a time, while keeping all the other assumptions in place. This will allow us to examine the genetic results of each process in isolation. In Chapter 3 we will relax more than one assumption simultaneously, to see how the different genetic processes interact with each other. In the next two sections we discuss two types of nonrandom mating, assortative mating and inbreeding. Except for negative assortative mating (discussed in the next section), nonrandom mating does not change allele frequencies, but rather changes how the alleles are distributed into diploid genotypes.

Assortative mating

Assortative mating means nonrandom mating based on some phenotypic trait, and can be either positive or negative. In **positive assortative mating** (often simply called assortative mating) mates are phenotypically more similar to each other than expected by chance. Positive assortative mating is very common. You may have noticed positive assortative mating for height in humans—tall people often date and marry other tall people, and shorter people do the same. Humans also date and marry assortatively for a number of measures of socioeconomic status, including the number of rooms in the couples' parents' houses! Positive assortative mating for size occurs in a wide variety of animals, especially arthropods (Crespi 1989). Plants mate assortatively for flowering time, because early flowering plants are often no longer flowering when late flowering plants are in bloom (Waser 1993).

The genetic effect of positive assortative mating is to increase **homozygosity** (the frequency of homozygous genotypes) and decrease **heterozygosity**. Think of a locus without dominance in a moth, where AA moths tend to be larger than Aa moths, which in turn are larger than aa moths. With positive assortative mating, AA moths will tend to mate with other AA , and aa with aa ; all these matings will produce only homozygous offspring. Even matings between Aa moths will produce half homozygous offspring. There will be very few of the $AA \times aa$ matings that produce all heterozygotes, so the frequency of heterozygotes decreases.

The genetic effects of assortative mating are only at those loci that affect the phenotypic trait by which the organisms are mating assortatively (and at loci linked to these). Thus, positive assortative mating for flowering time will increase homozygosity and decrease heterozygosity only at gene loci that affect flowering time. Loci that do not affect the mating trait are not affected, because mating is random with respect to genotypes at those loci. This fact distinguishes assortative mating from inbreeding, which affects all loci equally (discussed in the next section). Like inbreeding, positive assortative mating changes genotypic frequencies without changing allele frequencies.

Negative assortative mating, also called disassortative mating, occurs when mates are phenotypically less similar to each other than expected by chance. A very common and familiar example of disassortative mating is the two sexes—males tend to mate with females and vice versa. Otherwise, disassortative mating is far less common than positive assortative mating, especially in animals. An important and well-studied example of disassortative mating in plants is self-incompatibility. A number of large plant families have specific self-incompatibility loci. Pollen that shares alleles at this locus in common with the plant it lands on cannot successfully fertilize that plant (Barrett 1988). This prevents self-fertilization and sometimes prevents matings between close relatives, since close relatives often share identical alleles.

Not surprisingly, the genetic effect of negative assortative mating is the opposite of the effect of positive assortative mating—it increases the frequency of heterozygotes and decreases homozygosity. This is caused by AA and aa homozygotes mating with each other more often than expected by chance because they are phenotypically dissimilar. Note that, as with positive assortative mating, the genetic effect is only at loci that affect the phenotypic trait used in mating as well as loci linked to these. Unlike all the other factors we discuss in this chapter, negative assortative mating can alter allele frequencies through a process called negative frequency-dependent selection, which we will discuss further in Chapter 3.

Inbreeding

There are many different ways in which the word *inbreeding* is used and defined (see Chapter 3; Keller and Waller 2002; Templeton and Read 1994).

In this book we define **inbreeding** as mating between individuals in a population that are more closely related than expected by random chance. *The importance of inbreeding is due to its genetic effect, which is to increase homozygosity at all loci.* The reason that inbreeding affects all loci equally is that related individuals are genetically similar by common ancestry, and therefore they are more likely than unrelated individuals to share alleles throughout the genome. This contrasts with assortative mating, in which mating is between individuals that are *phenotypically* similar, and thus tend to share alleles only at loci affecting the phenotypic traits used in mating.

To understand how inbreeding increases homozygosity, it is simplest to examine the most extreme case of inbreeding, **self-fertilization**. Self-fertilization (or **selfing**) can occur only in hermaphrodites. Selfing is common in plants and in internal parasites such as tapeworms. In selfing, male and female gametes are produced normally through meiosis; the difference is that pollen or sperm fertilizes ovules or eggs on the same individual. Because there is recombination, selfed progeny are not genetically identical to the parent. If we start with a population in HWE with equal allele frequencies at a given locus ($p = q = 0.5$), then the genotypic frequencies are 0.25 for the two homozygotes and 0.5 heterozygotes (Figure 2.12). With self-fertilization, homozygotes can only produce homozygous offspring, whereas heterozygotes produce 25% of each homozygote and 50% heterozygotes. (Self-fertilization in a heterozygote is the same as a mating between two heterozygotes.) The arrows in Figure 2.12 show why the fre-

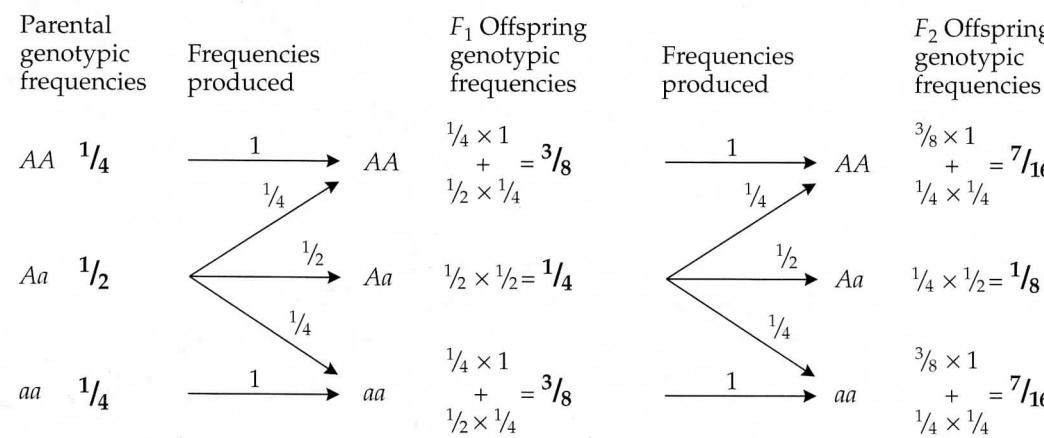


Figure 2.12 Genetic effects of inbreeding illustrated with self-fertilization. To calculate the offspring genotypic frequencies, multiply the parental frequency by the frequency of each offspring genotype produced by that parental genotype (above each arrow). Both homozygous and heterozygous parents produce homozygous offspring, so the frequencies need to be summed to get the total frequencies of homozygous offspring.

quency of homozygotes increases with self-fertilization: Whereas all offspring of homozygous parents are homozygous, only half of the offspring of heterozygotes are heterozygous, with the other half divided equally between the two homozygotes. A similar process occurs, but at a slower rate, with less severe forms of inbreeding (e.g., sibling or first-cousin mating); that is, the frequency of homozygotes increases at the expense of the frequency of heterozygotes.

Inbreeding by itself does not change allele frequencies. One way to explain this is to think about alleles as being shuffled into homozygotes without changing the frequencies of these alleles. To check this principle, use Equations 2.3 and 2.5 to calculate the allele frequencies in the offspring generation. The increase in homozygosity and lack of change in allele frequencies with inbreeding occur regardless of what the allele frequencies are—these results do not depend on the allele frequencies being 0.5 as in our example. This fact is illustrated in Figure 2.13, which shows the genotypic frequencies at all allele frequencies after one generation of selfing, as well as those under HWE from Figure 2.10.

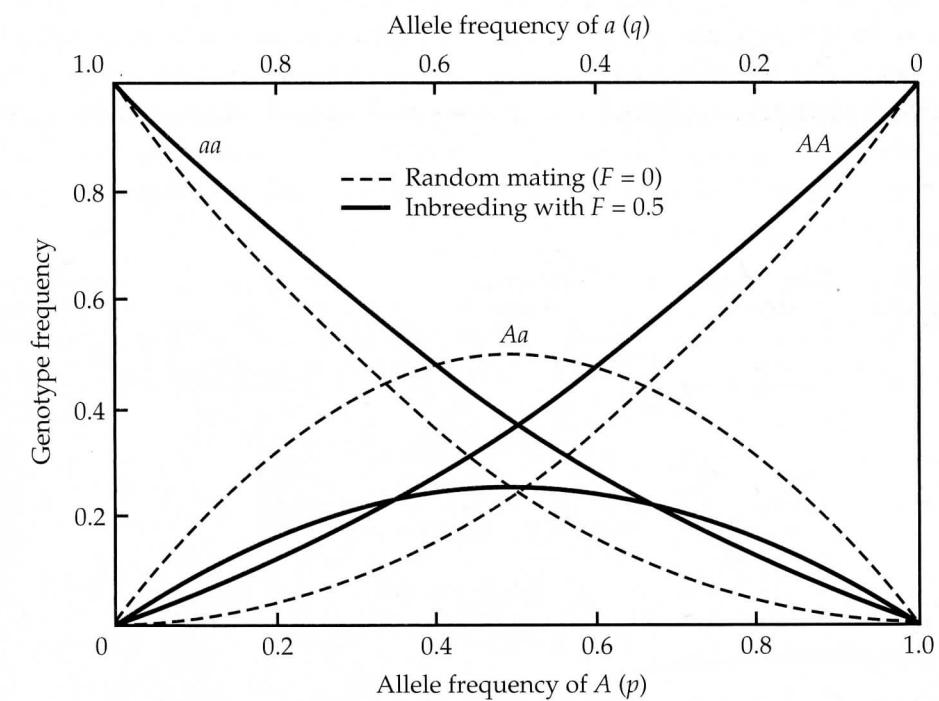


Figure 2.13 Changes in genotypic frequencies across all allele frequencies due to inbreeding. The dashed lines are genotypic frequencies under HWE from Figure 2.10, and the solid lines are those after one generation of selfing ($F = 0.5$). Note that with inbreeding the curves for homozygotes move up and the curve for heterozygotes moves down.

EXERCISE: Work through the effects of one generation of selfing starting from a Hardy-Weinberg population in which $p = 0.75$ and $q = 0.25$. Calculate genotypic and allele frequencies in the offspring.

ANSWER: Genotypic frequencies in the parental generation are 0.562, 0.375, and 0.062. After one generation of selfing, the heterozygote frequency is halved, with one quarter of the heterozygotes being converted to each homozygote. Therefore, the new genotypic frequencies are 0.656, 0.188, and 0.156, and the allele frequencies are unchanged (by Equations 2.3 and 2.5).

We can use the predictable increase in homozygosity and decrease in heterozygosity resulting from inbreeding to measure the effects of inbreeding. In the following equation, the **inbreeding coefficient**, F , measures the amount of inbreeding by comparing the frequency of heterozygotes in the population to the frequency expected under random mating:

$$F = \frac{H_0 - H}{H_0} \quad 2.7$$

As before, H equals the frequency of heterozygotes in the population, which can be measured using a variety of genetic markers, and H_0 is the expected frequency under HWE, that is, $2pq$. So the inbreeding coefficient equals the

EXERCISE: What is F after one generation of self-fertilization with $p = q = 0.5$ and with $p = 0.75$ and $q = 0.25$ (our two examples above)?

ANSWER: $F = \frac{0.5 - 0.25}{0.5} = 0.5$

for $p = q = 0.5$, and

$$F = \frac{0.375 - 0.188}{0.375} = 0.5$$

for $p = 0.75$ and $q = 0.25$;

It is the same for any starting allele frequency, since $H = H_0/2$ after one generation of selfing regardless of allele frequency.

fractional reduction in heterozygosity due to inbreeding, relative to a random mating population with the same allele frequencies.

F can range between -1 and 1 . When the population is in HWE, H_0 equals H and $F = 0$; that is, no inbreeding. If inbreeding continues, the population can become entirely homozygous, that is, $H = 0$ so $F = 1$. This happens very rapidly with selfing (Figure 2.14), and less rapidly with less severe forms of inbreeding, but note that with half-sibling mating (mating between two individuals that share only one of their parents in common) the inbreeding coefficient is more than 0.8 after only 20 generations. Figure 2.14 shows that there is no one value of F for a given mating system, because F also depends on the number of generations over which inbreeding has taken place. When F is negative, it means that the population is more outbred than expected by chance, that is, active avoidance of mating with relatives.

EXERCISE: Repeat the selfing exercise on the bottom of page 40 for two more generations, and calculate F , genotypic, and allele frequencies for both offspring generations.

Figure 2.14 shows that the inbreeding coefficient increases more rapidly with more closely related mates. The degree of relatedness is quantified as the coefficient of relatedness (r_{IJ}), which is the proportion of genes shared between two individuals I and J due to common descent (often referred to as *identical by descent* or IBD; Maynard Smith 1998). For example, r_{IJ} for a parent and each of their offspring is 0.5 in a sexual species, because half of each

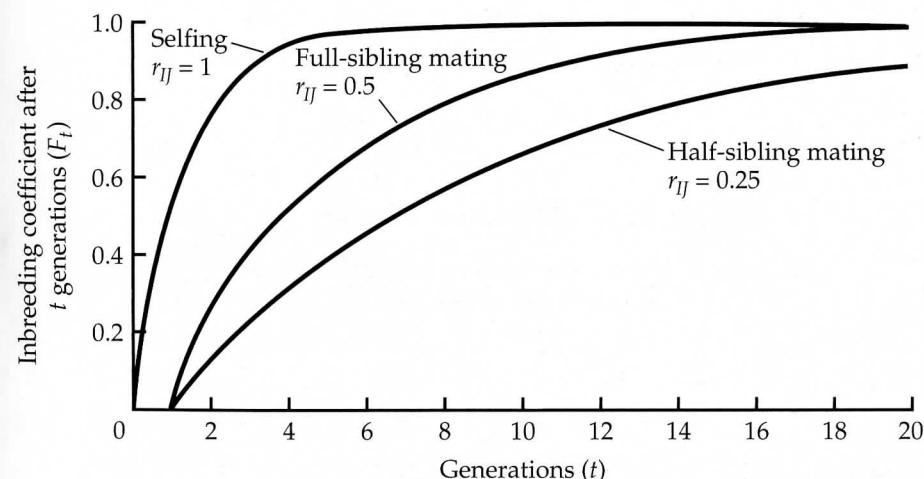


Figure 2.14 Theoretical increase in the inbreeding coefficient F for regular systems of mating: selfing, full-sibling mating, and half-sibling mating.