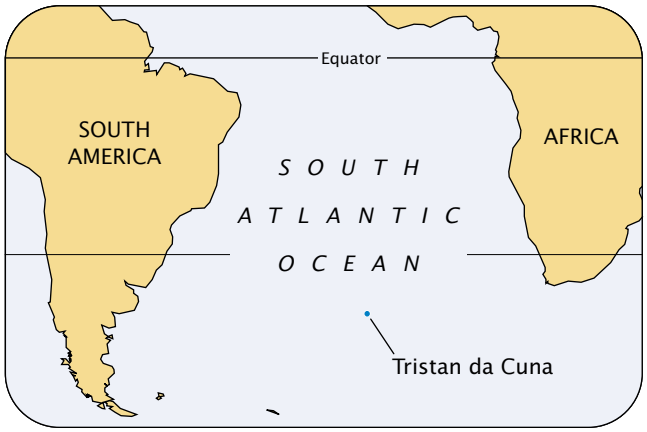
23 Population and Evolutionary Genetics

The inhabitants of the island of Tristan da Cuna have one of the highest incidences of asthma in the world due to the population's unique genetic history. (John Eckwall.)

The Genetic History of Tristan da Cuna

In the fall of 1993, geneticist Noé Zamel arrived at Tristan da Cuna, a small remote island in the South Atlantic (◖ FIGURE 23.1). It had taken Zamel 9 days to make the trip from his home in Canada, first by plane from Toronto to South Africa and then aboard a small research vessel to the island. Because of its remote location, the people of Tristan da Cuna call their home "the loneliest island," but isolation was not what attracted Zamel to Tristan da Cuna. Zamel was looking for a gene that causes asthma, and the inhabitants of Tristan da Cuna have one of the world's highest incidences of hereditary asthma: more than half of the islanders display some symptoms of the disease.

The high frequency of asthma on Tristan da Cuna derives from the unique history of the island's gene pool. The population traces its origin to William Glass, a Scot who moved his family there in 1817. They were joined by some shipwrecked sailors and a few women who migrated from the island of St. Helena but, owing to its remote location and lack of a deep harbor, the island population remained largely isolated. The descendants of Glass and the other settlers intermarried, and slowly the island population increased in number; by 1855, about 100 people inhabited the island. However, Tristan da Cuna's population dropped markedly when, after William Glass's death in 1856, many islanders migrated to South America and South Africa. By 1857, only 33 people remained, and the population grew slowly afterward. It was reduced again in 1885 when a small



23.1 Tristan de Cuna is a small island in the South Atlantic.

boat carrying 15 men was capsized by a huge wave, drowning all on board. Many of the widows and their children left the island, and the population dropped from 106 to 59. In 1961, a volcanic eruption threatened the main village. Fortunately, all of the islanders were rescued and transported to England, where they spent 2 years before returning to Tristan da Cuna.

Today, just a little more than 300 people permanently inhabit the island. These islanders have many genes in common and, in fact, all the island's inhabitants are no less closely related than cousins. Because the founders of the colony were few in number and many were already related, many of the genes in today's population can be traced to just a few original settlers. The population has always been small, which also gives rise to inbreeding and allows chance factors to have a large effect on the frequencies of the alleles in the population. The abrupt population reductions in 1856 and 1885 eliminated some alleles from the population and elevated the frequencies of others. As will be discussed in this chapter, the events affecting these islanders (small number of founders, limited population size, inbreeding, and population reduction) affect the proportions of alleles in a population. All of these factors have contributed to the high proportion of alleles that cause asthma among the inhabitants of Tristan da Cuna.

Tristan da Cuna illustrates how the history of a population shapes its genetic makeup. *Population genetics* is the branch of genetics that studies the genetic makeup of *groups* of individuals and how a group's genetic composition changes with time. Population geneticists usually focus their attention on a **Mendelian population,** which is a group of interbreeding, sexually reproducing individuals that have a common set of genes, the **gene pool.** A population evolves through changes in its gene pool; so population genetics is therefore also the study of evolution. Population geneticists study the variation in alleles within and between groups and the evolutionary forces responsible for shaping the patterns of genetic variation found in nature. In this chapter, we will learn how the gene pool of a population is measured and what factors are responsible for shaping it. In the later part of the chapter, we will examine molecular studies of genetic variation and evolution.

Genetic Variation

An obvious and pervasive feature of life is variability. Consider a group of students in a typical college class, the members of which vary in eye color, hair color, skin pigmentation, height, weight, facial features, blood type, and susceptibility to numerous diseases and disorders. No two students in the class are likely to be even remotely similar in appearance (◖ FIGURE 23.2a).

Humans are not unique in their extensive variability; almost all organisms exhibit variation in phenotype. For instance, lady beetles are highly variable in their patterns of spots (◖ FIGURE 23.2b), mice vary in body size, snails have different numbers of stripes on their shells, and plants vary in their susceptibility to pests. Much of this phenotypic variation is hereditary. Recognition of the extent of phenotypic variation and its genetic basis led Charles Darwin to the idea of evolution through natural selection.

In fact, even more genetic variation exists in populations than is visible in the phenotype. Much variation exists at the molecular level owing to the redundancy of the genetic code, which allows different codons to specify the same amino acids. Thus two individuals can produce the same protein even if their DNA sequences are different. DNA sequences between the genes and introns within genes do not encode proteins; so much of the variation in these sequences also has little effect on the phenotype.

The amount of genetic variation within natural populations and the forces that limit and shape it are of primary interest to population geneticists. Genetic variation is the basis of all evolution, and the extent of genetic variation within a population affects its potential to adapt to environmental change.

An important, but frequently misunderstood, tool used in population genetics is the mathematical model. Let's take a moment to consider what a model is and how it can be used. A mathematical model usually describes a process in terms of an equation. Factors that may influence the process are represented by variables in the equation; the equation defines the way in which the variables influence the process. Most models are simplified representations of a process, because it is impossible to simultaneously consider all of the influencing factors; some must be ignored in order to examine the effects of others. At first, a model might consider only one or a few factors, but, after their effects are understood, the model can be improved by the addition of

23.2 All organisms exhibit genetic variation. (a) Extensive variation among humans. (b) Variation in spotting patterns of Asian lady beetles. (Part a, Paul Warner/AP)

more details. It is important to realize that even a simple model can be a source of valuable insight into how a process is influenced by key variables.

www.whfreeman.com/pierce More information on genetic diversity within the human species

Before we can explore the evolutionary processes that shape genetic variation, we must be able to describe the genetic structure of a population. The usual way of doing so is to enumerate the types and frequencies of genotypes and alleles in a population.

Calculation of Genotypic Frequencies

A frequency is simply a proportion or a percentage, usually expressed as a decimal fraction. For example, if 20% of the alleles at a particular locus in a population are *A*, we would say that the frequency of the *A* allele in the population is .20. For large populations, where it is not practical to determine the genes of all individuals, a sample of individuals from the population is usually taken and the genotypic and allelic frequencies are calculated for this sample (see Chapter 22 for a discussion of samples.) The genotypic and allelic frequencies of the sample are then used to represent the gene pool of the population.

To calculate a **genotypic frequency,** we simply add up the number of individuals possessing the genotype and divide by the total number of individuals in the sample (N). For a locus with three genotypes *AA*, *Aa*, and *aa*, the frequency (f) of each genotype is:

$$f(AA) = \frac{\text{number of } AA \text{ individuals}}{N} \qquad (23.1)$$

$$f(Aa) = \frac{\text{number of } Aa \text{ individuals}}{N}$$

$$f(aa) = \frac{\text{number of } aa \text{ individuals}}{N}$$

The sum of all the genotypic frequencies always equals 1.

Calculation of Allelic Frequencies

The gene pool of a population can also be described in terms of the allelic frequencies. There are always fewer alleles than genotypes; so the gene pool of a population can be described in fewer terms when the allelic frequencies are used. In a sexually reproducing population, the genotypes are only temporary assemblages of the alleles: the genotypes

break down each generation when individual alleles are passed to the next generation through the gametes, and so it is the types and numbers of alleles, not genotypes, that have real continuity from one generation to the next and that make up the gene pool of a population.

Allelic frequencies can be calculated from (1) the numbers or (2) the frequencies of the genotypes. To calculate the **allelic frequency** from the numbers of genotypes, we count the number of copies of a particular allele present in a sample and divide by the total number of all alleles in the sample:

$$\text{frequency of an allele} = \frac{\text{number of copies of the allele}}{\text{number of copies of all alleles at the locus}} \qquad (23.2)$$

For a locus with only two alleles (A and a), the frequencies of the alleles are usually represented by the symbols p and q, and can be calculated as follows:

$$p = f(A) = \frac{2n_{AA} + n_{Aa}}{2N} \qquad (23.3)$$

$$q = f(a) = \frac{2n_{aa} + n_{Aa}}{2N}$$

where n_{AA}, n_{Aa}, and n_{aa} represent the numbers of *AA*, *Aa*, and *aa* individuals, and N represents the total number of individuals in the sample. We divide by $2N$ because each diploid individual has two alleles at a locus. The sum of the allelic frequencies always equals 1 ($p + q = 1$); so after p has been obtained, q can be determined by subtraction: $q = 1 - p$.

Alternatively, allelic frequencies can also be calculated from the genotypic frequencies. To do so, we add the frequency of the homozygote for each allele to half the frequency of the heterozygote (because half of the heterozygote's alleles are of each type):

$$p = f(A) = f(AA) + \tfrac{1}{2} f(Aa) \qquad (23.4)$$

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

We obtain the same values of p and q whether we calculate the allelic frequencies from the numbers of genotypes (Equation 23.3) or from the genotypic frequencies (Equation 23.4).

Loci with multiple alleles We can use the same principles to determine the frequencies of alleles for loci with more than two alleles. To calculate the allelic frequencies from the numbers of genotypes, we count up the number of copies of an allele by adding twice the number of homozygotes to the number of heterozygotes that possess the allele and divide this sum by twice the number of individuals in the sample. For a locus with three alleles (A^1, A^2, and A^3) and six genotypes (A^1A^1, A^1A^2, A^1A^3, A^2A^2, A^2A^3, and A^3A^3), the frequencies (p, q, and r) of the alleles are:

$$p = f(A^1) = \frac{2n_{A^1A^1} + n_{A^1A^2} + n_{A^1A^3}}{2N} \qquad (23.5)$$

$$q = f(A^2) = \frac{2n_{A^2A^2} + n_{A^1A^2} + n_{A^2A^3}}{2N}$$

$$r = f(A^3) = \frac{2n_{A^3A^3} + n_{A^1A^3} + n_{A^2A^3}}{2N}$$

Alternatively, we can calculate the frequencies of multiple alleles from the genotypic frequencies by extending Equation 23.4. Once again, we add the frequency of the homozygote to half the frequency of each heterozygous genotype that possesses the allele:

$$p = f(A^1) = f(A^1A^1) + \tfrac{1}{2} f(A^1A^2) + \tfrac{1}{2} f(A^1A^3) \qquad (23.6)$$

$$q = f(A^2) = f(A^2A^2) + \tfrac{1}{2} f(A^1A^2) + \tfrac{1}{2} f(A^2A^3)$$

$$r = f(A^3) = f(A^3A^3) + \tfrac{1}{2} f(A^1A^3) + \tfrac{1}{2} f(A^2A^3)$$

X-linked loci To calculate allelic frequencies for genes at X-linked loci, we apply these same principles. However, we must remember that a female possesses two X chromosomes and therefore has two X-linked alleles, whereas a male has only a single X chromosome and has one X-linked allele.

Suppose there are two alleles at an X-linked locus, X^A and X^a. Females may be either homozygous (X^AX^A or X^aX^a) or heterozygous (X^AX^a). All males are hemizygous (X^AY or X^aY). To determine the frequency of the X^A allele (p), we first count the number of copies of X^A: we multiply the number of X^AX^A females by two and add the number of X^AX^a females and the number of X^AY males. We then divide the sum by the total number of alleles at the locus, which is twice the total number of females plus the number of males:

$$p = f(X^A) = \frac{2n_{X^AX^A} + n_{X^AX^a} + n_{X^AY}}{2n_{\text{females}} + n_{\text{males}}} \qquad (23.7a)$$

Similarly, the frequency of the X^a allele is:

$$q = f(X^a) = \frac{2n_{X^aX^a} + n_{X^AX^a} + n_{X^aY}}{2n_{\text{females}} + n_{\text{males}}} \qquad (23.7b)$$

The frequencies of X-linked alleles can also be calculated from genotypic frequencies by adding the frequency of the females that are homozygous for the allele, half the frequency of the females that are heterozygous for the allele, and the frequency of males hemizygous for the allele:

$$p = f(X^A) = f(X^AX^A) + \tfrac{1}{2} f(X^AX^a) + f(X^AY) \qquad (23.8)$$

$$q = f(X^a) = f(X^aX^a) + \tfrac{1}{2} f(X^AX^a) + f(X^aY)$$

If you remember the logic behind all of these calculations, you can determine allelic frequencies for any set of genotypes, and it will not be necessary to memorize the formulas.

Concepts

Population genetics is concerned with the genetic composition of a population and how it changes with time. The gene pool of a population can be described by the frequencies of genotypes and alleles in the population.

Worked Problem

The human MN blood type antigens are determined by two codominant alleles, L^M and L^N (see p. 000 in Chapter 5). The MN blood types and corresponding genotypes of 398 Finns from Karjala are tabulated here.

Phenotype	Genotype	Number
MM	L^ML^M	182
MN	L^ML^N	172
NN	L^NL^N	44

Source: W. C. Boyd, *Genetics and the Races of Man* (Boston: Little, Brown, 1950.)

Calculate the allelic and genotypic frequencies at the MN locus for the Karjala population.

• Solution

The genotypic frequencies for the population are calculated with the following formula:

$$\text{genotypic frequency} = \frac{\text{number of individuals with genotype}}{\text{total number of individudals in sample}(N)}$$

$$f(L^ML^M) = \frac{\text{number of } L^ML^M \text{ individuals}}{N} = \frac{182}{398} = .457$$

$$f(L^ML^N) = \frac{\text{number of } L^ML^N \text{ individuals}}{N} = \frac{172}{398} = .432$$

$$f(L^NL^N) = \frac{\text{number of } L^NL^N \text{ individuals}}{N} = \frac{44}{398} = .111$$

The allelic frequencies can be calculated from either the numbers or the frequencies of the genotypes. To calculate allelic frequencies from numbers of genotypes, we add the number of copies of the allele and divide by the number of copies of all alleles at that locus.

$$\text{frequency of an allele} = \frac{\text{number of copies of the allele}}{\text{number of copies of all alleles}}$$

$$p = f(L^M) = \frac{(2n_{L^ML^M}) + (n_{L^ML^N})}{2N} = \frac{2(182) + 172}{2(398)} = \frac{536}{796} = .673$$

$$q = f(L^N) = \frac{(2n_{L^NL^N}) + (n_{L^ML^N})}{2N} = \frac{2(44) + 172}{2(398)} = \frac{260}{796} = .327$$

To calculate the allelic frequencies from genotypic frequencies, we add the frequency of the homozygote for that genotype to half the frequency of each heterozygote that contains that allele:

$$p = f(L^M) = f(L^ML^M) + \tfrac{1}{2} f(L^ML^N) = .457 + \tfrac{1}{2}(.432) = .673$$

$$p = f(L^N) = f(L^NL^N) + \tfrac{1}{2} f(L^ML^N) = .111 + \tfrac{1}{2}(.432) = .327$$

The Hardy-Weinberg Law

The primary goal of population genetics is to understand the processes that shape a population's gene pool. First, we must ask what effects reproduction and Mendelian principles have on the genotypic and allelic frequencies: How do the segregation of alleles in gamete formation and the combining of alleles in fertilization influence the gene pool? The answer to this question lies in the **Hardy-Weinberg law,** one of the most important principles of population genetics.

The Hardy-Weinberg law was formulated independently by both Godfrey H. Hardy and Wilhelm Weinberg in 1908. (Similar conclusions were reached by several other geneticists about the same time.) The law is actually a mathematical model that evaluates the effect of reproduction on the genotypic and allelic frequencies of a population. It makes several simplifying assumptions about the population and provides two key predictions if these assumptions are met. For an autosomal locus with two alleles, the Hardy-Weinberg law can be stated as follows:

Assumptions—If a population is large, randomly mating, and not affected by mutation, migration, or natural selection, then:

Prediction 1—the allelic frequencies of a population do not change; and

Prediction 2—the genotypic frequencies stabilize (will not change) after one generation in the proportions p^2 (the frequency of *AA*), $2pq$ (the frequency of *Aa*), and q^2 (the frequency of *aa*), where p equals the frequency of allele *A* and q equals the frequency of allele *a*.

The Hardy-Weinberg law indicates that, when the assumptions are met, reproduction alone does not alter allelic or genotypic frequencies and the allelic frequencies determine the frequencies of genotypes.

The statement that genotypic frequencies stabilize after one generation means that they may change in the first generation after random mating, because one generation of random mating is required to produce Hardy-Weinberg proportions of the genotypes. Afterward, the genotypic frequencies, like allelic frequencies, do not change as long as the population continues to meet the assumptions of the Hardy-Weinberg law. When genotypes are in the expected proportions of p^2, $2pq$, and q^2, the population is said to be in **Hardy-Weinberg equilibrium.**

Concepts

The Hardy-Weinberg law describes how reproduction and Mendelian principles affect the allelic and genotypic frequencies of a population.

Closer Examination of the Assumptions of the Hardy-Weinberg Law

Before we consider the implications of the Hardy-Weinberg law, we need to take a closer look at the three assumptions that it makes about a population. First, it assumes that the population is large. How big is "large"? Theoretically, the Hardy-Weinberg law requires that a population be infinitely large in size, but this requirement is obviously unrealistic. In practice, a population need only be large enough that chance deviations from expected ratios do not cause significant changes in allelic frequencies. Later in the chapter, we will examine the effects of small population size on allelic frequencies.

A second assumption of the Hardy-Weinberg law is that individuals in the population mate randomly, which means that each genotype mates in proportion to its frequency. For example, suppose that three genotypes are present in a population in the following proportions: $f(AA) = .6$, $f(Aa) = .3$, and $f(aa) = .1$. With random mating, the frequency of mating between two *AA* homozygotes ($AA \times AA$) will be equal to the multiplication of their frequencies: $.6 \times .6 = .36$, whereas the frequency of mating between two *aa* homozygotes ($aa \times aa$) will be only $.1 \times .1 = .01$.

A third assumption of the Hardy-Weinberg law is that the allelic frequencies of the population are not affected by natural selection, migration, and mutation. Although mutation occurs in every population, its rate is so low that it has little effect on the predictions of the Hardy-Weinberg law. Although natural selection and migration are significant factors in real populations, we must remember that the purpose of the Hardy-Weinberg law is to examine only the effect of reproduction on the gene pool. When this effect is known, the effects of other factors (such as migration and natural selection) can be examined.

A final point that should be mentioned is that the assumptions of the Hardy-Weinberg law apply to a *single* locus. No real population mates randomly for all traits; nor is a population completely free of natural selection for all traits. The Hardy-Weinberg law, however, does not require random mating and the absence of selection, migration, and mutation for all traits; it requires these conditions only for the locus under consideration. A population may be in Hardy-Weinberg equilibrium for one locus but not for others.

Implications of the Hardy-Weinberg Law

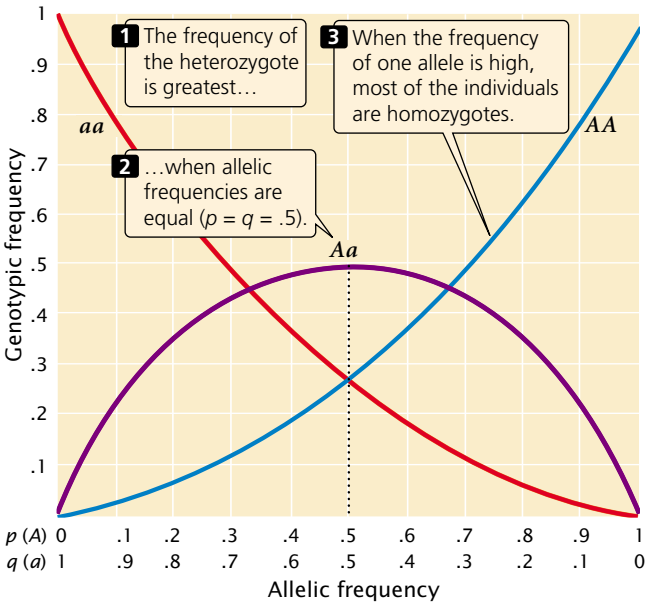
The Hardy-Weinberg law has several important implications for the genetic structure of a population. One implication is that a population cannot evolve if it meets the Hardy-Weinberg assumptions, because evolution consists of change in the allelic frequencies of a population. Therefore the Hardy-Weinberg law tells us that reproduction alone will not bring about evolution. Other processes such as natural selection, mutation, migration, or chance in small populations are required for populations to evolve.

A second important implication is that, when a population is in Hardy-Weinberg equilibrium, the genotypic frequencies are determined by the allelic frequencies. For a locus with two alleles, the frequency of the heterozygote is greatest when allelic frequencies are between .33 and .66 and is at a maximum when allelic frequencies are each .5 (◖ FIGURE 23.3). The heterozygote frequency also never exceeds .5. Furthermore, when the frequency of one allele is low, homozygotes for that allele will be rare, and most of the copies of a rare allele will be present in heterozygotes. As you can see from Figure 23.3, when the frequency of allele *a* is .2, the frequency of the *aa* homozygote is only .04 (q^2), but the frequency of *Aa* heterozygotes is .32 ($2pq$); 80% of the *a* alleles are in heterozygotes.

A third implication of the Hardy-Weinberg law is that a single generation of random mating produces the equilibrium frequencies of p^2, $2pq$, and q^2. The fact that genotypes are in Hardy-Weinberg proportions does not prove that the population is free from natural selection, mutation, and migration. It means only that these forces have not acted since the last time random mating took place.

Extensions of the Hardy-Weinberg Law

The Hardy-Weinberg law can also be applied to multiple alleles and X-linked alleles. The genotypic frequencies expected under Hardy-Weinberg equilibrium will differ according to the situation.



23.3 When a population is in Hardy-Weinberg equilibrium, the proportions of genotypes are determined by frequencies of alleles.

Hardy-Weinberg expectations for loci with multiple alleles In general, the genotypic frequencies expected at equilibrium are the square of the allelic frequencies. For an autosomal locus with two alleles, these frequencies are $(p + q)^2 = p^2 + 2pq + q^2$. We can also use the square of the allelic frequencies to calculate the equilibrium frequencies for a locus with multiple alleles. An autosomal locus with three alleles, A^1, A^2, and A^3, has six genotypes: A^1A^1, A^1A^2, A^2A^2, A^1A^3, A^2A^3, and A^3A^3. According to the Hardy-Weinberg law, the frequencies of the genotypes at equilibrium depend on the frequencies of the alleles. If the frequencies of alleles A^1, A^2, and A^3 are p, q, and r, respectively, then the equilibrium genotypic frequencies will be the square of the allelic frequencies $(p + q + r)^2 = p^2 + 2pq + q^2 + 2pr + 2qr + r^2$, where:

$$p^2 = f(A^1A^1) \qquad (23.9)$$

$$2pq = f(A^1A^2)$$

$$q^2 = f(A^2A^2)$$

$$2pr = f(A^1A^3)$$

$$2qr = f(A^2A^3)$$

$$r^2 = f(A^3A^3)$$

The square of the allelic frequencies can also be used to calculate the expected genotypic frequencies for loci with four or more alleles.

Hardy-Weinberg expectations for X-linked loci For an X-linked locus with two alleles, X^A and X^a, there are five possible genotypes: X^AX^A, X^AX^a, X^aX^a, X^AY, and X^aY. Females possess two X-linked alleles, and the expected proportions of the female genotypes can be calculated by using the square of the allelic frequencies. If the frequencies of X^A and X^a are p and q, respectively, then the equilibrium frequencies of the female genotypes are $(p + q)^2 = p^2$ (frequency of X^AX^A) + $2pq$ (frequency of X^AX^a) + q^2 (frequency of X^aX^a). Males have only a single X-linked allele, and so the frequencies of the male genotypes are p (frequency of X^AY) and q (frequency of X^aY). Notice that these expected frequencies are the proportions of the genotypes among males and females rather than the proportions among the entire population. Thus, p^2 is the expected proportion of females with the genotype X^AX^A; if females make up 50% of the population, then the expected proportion of this genotype in the entire population is $.5 \times p^2$.

The frequency of an X-linked recessive trait among males is q, whereas the frequency among females is q^2. When an X-linked allele is uncommon, the trait will therefore be much more frequent in males than in females. Consider hemophilia A, a clotting disorder caused by an X-linked recessive allele with a frequency (q) of approximately 1 in 10,000, or .0001. At Hardy-Weinberg equilibrium, this frequency will also be the frequency of the disease among males. The frequency of the disease among females, however, will be $q^2 = (.0001)^2 = .00000001$, which is only 1 in 10 million. Hemophilia is 1000 times as frequent in males as in females.

Testing for Hardy-Weinberg Proportions

If a population is in equilibrium, then it is randomly mating for the locus in question, and selection, migration, mutation, and small population size have not significantly influenced the genotypic frequencies since random mating last took place. To determine whether these conditions are met, the genotypic proportions expected under the Hardy-Weinberg law must be compared with the observed genotypic frequencies. To do so, we first calculate the allelic frequencies, then find the expected genotypic frequencies by using the square of the allelic frequencies, and finally compare the observed and expected genotypic frequencies by using a chi-square test.

Worked Problem

Jeffrey Mitton and his colleagues found three genotypes (R^2R^2, R^2R^3, and R^3R^3) at a locus encoding the enzyme peroxidase in ponderosa pine trees growing in Colorado. The observed numbers of these genotypes at Glacier Lake, Colorado, were:

Genotypes	Number observed
R^2R^2	135
R^2R^3	44
R^3R^3	11

Are the ponderosa pine trees at Glacier Lake, Colorado, in Hardy-Weinberg equilibrium at the peroxidase locus?

• Solution

If the frequency of the R^2 allele equals p and the frequency of the R^3 allele equals q, the frequency of the R^2 allele is:

$$p = f(R^2) = \frac{(2n_{R^2R^2}) + (n_{R^2R^3})}{2N} = \frac{135 + 44}{2(190)} = .826$$

The frequency of the R^3 allele is obtained by subtraction:

$$q = f(R^3) = 1 - p = .174$$

The frequencies of the genotypes expected under Hardy-Weinberg equilibrium are then calculated by using p^2, $2pq$, and q^2:

$$R^2R^2 = p^2 = (.826)^2 = .683$$

$$R^2R^3 = 2pq = 2(.826)(.174) = .287$$

$$R^3R^3 = q^2 = (.174)^2 = .03$$

Multiplying each of these expected genotypic frequencies by the total number of observed individuals in the sample (190), we obtain the *numbers* expected for each genotype:

$$R^2R^2 = .683 \times 190 = 129.7$$

$$R^2R^3 = .287 \times 190 = 54.5$$

$$R^3R^3 = .03 \times 190 = 5.7$$

Comparing these expected numbers with the observed numbers of each genotype, we see that there are more R^2R^2 homozygotes and fewer R^2R^3 heterozygotes and R^3R^3 homozygotes in the population than we expect at equilibrium.

A goodness-of-fit chi-square test is used to determine whether the differences between the observed and the expected numbers of each genotype are due to chance:

$$\chi^2 = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$$
$$= \frac{(135 - 129.7)^2}{129.7} + \frac{(44 - 54.5)^2}{54.5} + \frac{(11 - 5.7)^2}{5.7}$$
$$= 0.22 + 2.02 + 4.93 = 7.17$$

The calculated chi-square value is 7.17; to obtain the probability associated with this chi-square value, we determine the appropriate degrees of freedom.

Up to this point, the chi-square test for assessing Hardy-Weinberg equilibrium has been identical with the chi-square tests that we used in Chapter 3 to assess progeny ratios in a genetic cross, where the degrees of freedom were $n - 1$ and n equaled the number of expected genotypes. For the Hardy-Weinberg test, however, we must subtract an additional degree of freedom, because the expected numbers are based on the observed allelic frequencies; therefore, the observed numbers are not completely free to vary. In general, the degrees of freedom for a chi-square test of Hardy-Weinberg equilibrium equal the number of expected genotypic classes minus the number of associated alleles. For this particular Hardy-Weinberg test, the degrees of freedom are $3 - 2 = 1$.

Once we have calculated both the chi-square value and degrees of freedom, the probability associated with this value can be sought in a chi-square table (Table 3.4). With one degree of freedom, a chi-square value of 7.17 has a probability between .01 and .001. It is very unlikely that the peroxidase genotypes observed at Glacier Lake are in Hardy-Weinberg proportions.

Concepts

The observed number of genotypes in a population can be compared to the Hardy-Weinberg expected proportions by using a goodness of fit chi-square test.

Estimating Allelic Frequencies with the Hardy-Weinberg Law

A practical use of the Hardy-Weinberg law is that it allows us to calculate allelic frequencies when dominance is present. For example, cystic fibrosis is an autosomal recessive disorder characterized by respiratory infections, incomplete digestion, and abnormal sweating (see p. 000 in Chapter 6). Among North American Caucasians, the incidence of the disease is approximately 1 person in 2000. The formula for calculating allelic frequency (Equation 23.3) requires that we know the numbers of homozygotes and heterozygotes, but cystic fibrosis is a recessive disease, and so we cannot easily distinguish between homozygous normal persons and heterozygous carriers. Although molecular tests are available for identifying heterozygous carriers of the cystic fibrosis gene, the low frequency of the disease makes widespread screening impractical. In such situations, the Hardy-Weinberg law can be used to estimate the allelic frequencies.

If we assume that a population is in Hardy-Weinberg equilibrium with regard to this locus, then the frequency of the recessive genotype (aa) will be q^2, and the allelic frequency is the square root of the genotypic frequency:

$$q = \sqrt{f(aa)} \qquad (23.10)$$

The frequency of cystic fibrosis in North American Caucasians is approximately 1 in 2000, or .0005; so $q = \sqrt{0.0005} = .02$. Thus, about 2% of the alleles in the Caucasian population encode cystic fibrosis. We can calculate the frequency of the normal allele by subtracting: $p = 1 - q = 1 - .02 = .98$. After we have calculated p and q, we can use the Hardy-Weinberg law to determine the frequencies of homozygous normal people and heterozygous carriers of the gene:

$$f(AA) = p^2 = (.98)^2 = .960$$

$$f(Aa) = 2pq = 2(.02)(.98) = .0392$$

Thus about 4% (1 of 25) of Caucasians are heterozygous carriers of the allele that causes cystic fibrosis.

Concepts

Although allelic frequencies cannot be calculated directly for traits that exhibit dominance, the Hardy-Weinberg law can be used to estimate the allelic frequencies if the population is in Hardy-Weinberg equilibrium for that locus. The frequency of the recessive allele will be equal to the square root of the frequency of the recessive trait.

Nonrandom Mating

An assumption of the Hardy-Weinberg law is that mating is random with respect to genotype. Nonrandom mating affects the way in which alleles combine to form genotypes and alters the genotypic frequencies of a population.

We can distinguish between two types of nonrandom mating. **Positive assortative mating** refers to a tendency for like individuals to mate. For example, humans exhibit positive assortative mating for height: tall people mate preferentially with other tall people; short people mate preferentially with other short people. **Negative assortative mating** refers to a tendency for unlike individuals to mate. If people engaged in negative assortative mating for height, tall and short people would preferentially mate.

One form of nonrandom mating is **inbreeding,** which is preferential mating between related individuals. Inbreeding is actually positive assortative mating for relatedness, but it differs from other types of assortative mating because it affects all genes, not just those that determine the trait for which the mating preference occurs. Inbreeding causes a departure from the Hardy-Weinberg equilibrium frequencies of p^2, $2pq$, and q^2. More specifically, it leads to an increase in the proportion of homozygotes and a decrease in the proportion of heterozygotes in a population. **Outcrossing** is the avoidance of mating between related individuals.

Inbreeding is usually measured by the **inbreeding coefficient,** designated F, which is a measure of the probability that two alleles are "identical by descent." In a diploid organism, homozygous individual has two copies of the same allele. These two copies may be the same in *state,* which means that the two alleles are alike in structure and function but do not have a common origin. Alternatively, the two alleles in a homozygous individual may be the same because they are identical by *descent*—the copies are descended from a single allele that was present in an ancestor (◖ FIGURE 23.4). If we go back far enough in time, many alleles are likely to be identical by descent but, for calculating the effects of inbreeding, we consider identity by descent by going back only a few generations.

Inbreeding coefficients can range from 0 to 1. A value of 0 indicates that mating in a large population is random; a value of 1 indicates that all alleles are identical by descent. Inbreeding coefficients can be calculated from analyses of pedigrees or they can be determined from the reduction in the heterozygosity of a population. Although we will not go into the details of how F is calculated, it's important to understand how inbreeding affects genotypic frequencies.

When inbreeding occurs, the frequency of the genotypes will be:

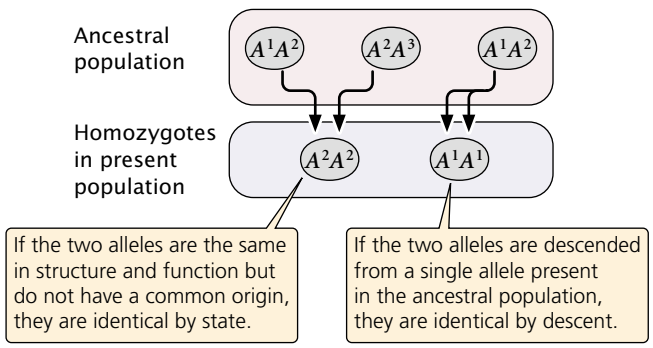
$$f(AA) = p^2 + Fpq \qquad (23.11)$$

$$f(Aa) = 2pq - 2Fpq$$

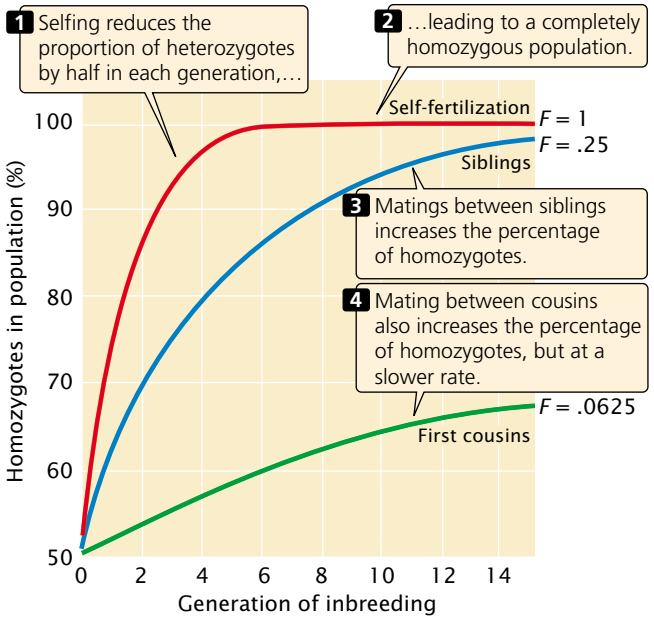
$$f(aa) = q^2 + Fpq$$

With inbreeding, the proportion of heterozygotes *decreases* by $2Fpq$, and half of this value (Fpq) is *added* to the proportion of each homozygote.

Consider a population that reproduces by self-fertilization (so $F = 1$). We will assume that this population begins with genotypic frequencies in Hardy-Weinberg proportions (p^2, $2pq$, and q^2). With selfing, each homozygote produces



◖ 23.4 **Individuals may be homozygous by state or by descent.** Inbreeding is a measure of the probability that two alleles are identical by descent.



23.5 Inbreeding increases the percentage of homozygous individuals in a population.

progeny only of the same homozygous genotype ($AA \times AA$ produces all AA; and $aa \times aa$ produces all aa), whereas only half the progeny of a heterozygote will be like the parent ($Aa \times Aa$ produces $^1/_4$ AA, $^1/_2$ Aa, and $^1/_4$ aa). Selfing therefore reduces the proportion of heterozygotes in the population by half with each generation, until all genotypes in the population are homozygous (Table 23.1 and Figure 23.5).

For most outcrossing species, close inbreeding is harmful because it increases the proportion of homozygotes and thereby boosts the probability that deleterious and lethal recessive alleles will combine to produce homozygotes with a harmful trait. Assume that a recessive allele (a) that causes a genetic disease has a frequency (q) of .01. If the population mates randomly ($F = 0$), the frequency of individuals affected with the disease (aa) will be $q^2 = .01^2 = .0001$; so only 1 in 10,000 individuals will have the disease. However, if $F = .25$ (the equivalent of brother–sister mating), then the expected frequency of the homozygote genotype is $q^2 + 2pqF = (.01)^2 + 2(.99)(.01)(.25) = .0026$; thus, the genetic disease is 26 times as frequent at this level of inbreeding. This increased appearance of lethal and deleterious traits with inbreeding is termed **inbreeding depression;** the more intense the inbreeding, the more severe the inbreeding depression.

The harmful effects of inbreeding have been recognized by humans for thousands of years and are the basis of cultural taboos against mating between close relatives. William Schull and James Neel found that, for each 10% increase in F, the mean IQ of Japanese children dropped six points. Child mortality also increases with close inbreeding (Table 23.2); children of first cousins have a 40% increase in mortality over that seen among the children of randomly mated people. Inbreeding also has deleterious effects on crops (Figure 23.6) and domestic animals.

Inbreeding depression is most often studied in humans, as well as in plants and animals reared in captivity, but the negative effects of inbreeding may be more severe in natural populations. Julie Jimenez and her colleagues collected wild mice from a natural population in Illinois and bred them in the laboratory for three to four generations. Laboratory matings were chosen so that some mice had no inbreeding, whereas others had an inbreeding coefficient of .25. When both types of mice were released back into the wild, the weekly survival of the inbred mice was only 56% of that of the noninbred mice. Inbred male mice also continously lost weight after release into the wild, whereas noninbred male mice regained their body weight within a few days after release.

In spite of the fact that inbreeding is generally harmful for outcrossing species, a number of plants and animals regularly inbreed and are successful (Figure 23.7). Inbreeding is commonly used to produce domesticated plants and animals having desirable traits. As stated earlier, inbreeding increases homozygosity, and eventually all individuals in the population become homozygous for the same allele. If a species undergoes inbreeding for a number of generations, many deleterious recessive alleles are weeded out by natural or artificial selection so that the population becomes homozygous for beneficial alleles. In this way, the harmful effects of inbreeding may eventually be eliminated, leaving a population that is homozygous for beneficial traits.

Table 23.1 Generational increase in frequency of homozygotes in a self-fertilizing population starting with $p = q = .5$

	Genotypic Frequencies		
Generation	***AA***	***Aa***	***aa***
1	$^1/_4$	$^1/_2$	$^1/_4$
2	$^1/_4 + ^1/_8 = ^3/_8$	$^1/_4$	$^1/_4 + ^1/_8 = ^3/_8$
3	$^3/_8 + ^1/_{16} = ^7/_{16}$	$^1/_8$	$^3/_8 + ^1/_{16} = ^7/_{16}$
4	$^7/_{16} + ^1/_{32} = ^{15}/_{32}$	$^1/_{16}$	$^7/_{16} + ^1/_{32} = ^{15}/_{32}$
n	$\frac{1 - (^1/_2)^n}{2}$	$(^1/_2)^n$	$\frac{1 - (^1/_2)^n}{2}$
∞	$^1/_2$	0	$^1/_2$

Concepts

Nonrandom mating alters the frequencies of the genotypes but not the frequencies of the alleles. Inbreeding is preferential mating between related individuals. With inbreeding, the frequency of homozygotes increases while the frequency of heterozygotes decreases.

Table 23.2 Effects of inbreeding on Japanese children

Genetic Relationship of Parents	*F*	Mortality of Children (Through 12 Years of Age)
Unrelated	0	.082
Second cousins	0.016 ($^1/_{64}$)	.108
First cousins	.0625 ($^1/_{16}$)	.114

Source: After D. L. Hartl, and A. G. Clark, *Principles of Population Genetics*. 2d ed. (Sunderland, MA: Sinauer, 1989), Table 2. Original data from W. J. Schull, and J. V. Neel, *The Effects of Inbreeding on Japanese Children*. (New York: Harper & Row, 1965).

Changes in Allelic Frequencies

The Hardy-Weinberg law indicates that allelic frequencies do not change as a result of reproduction; thus, other processes must cause alleles to increase or decrease in frequency. Processes that bring about change in allelic frequency include mutation, migration, genetic drift (random effects due to small population size), and natural selection.

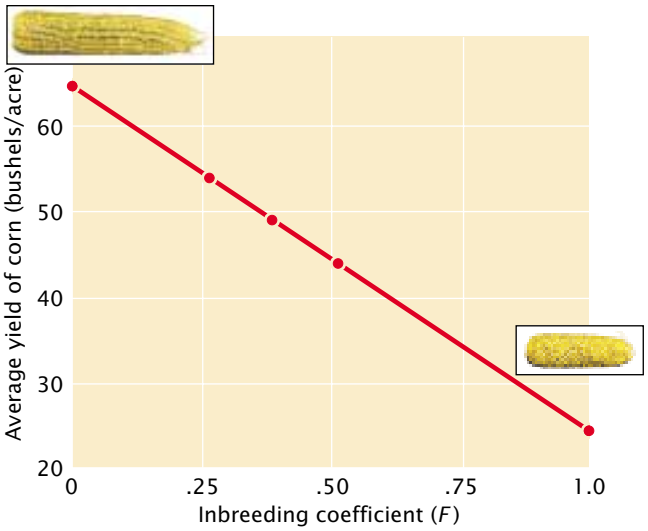
Mutation

Before evolution can occur, genetic variation must exist within a population; consequently, all evolution depends on processes that generate genetic variation. Although new *combinations* of existing genes may arise through recombination in meiosis, all genetic variants ultimately arise through mutation.

The effect of mutation on allelic frequencies Mutation can influence the rate at which one genetic variant increases at the expense of another. Consider a single locus in a population of 25 diploid individuals. Each individual possesses two alleles at the locus under consideration; so the gene pool of the population consists of 50 allelic copies. Let us assume that there are two different alleles, designated G^1 and G^2 with frequencies p and q, respectively. If there are 45 copies of G^1 and 5 copies of G^2 in the population, $p = .90$ and $q = .10$. Now suppose that a mutation changes a G^1 allele into a G^2 allele. After this mutation, there are 44 copies of G^1 and 6 copies of G^2, and the frequency of G^2 has increased from .10 to .12. Mutation has changed the allelic frequency.

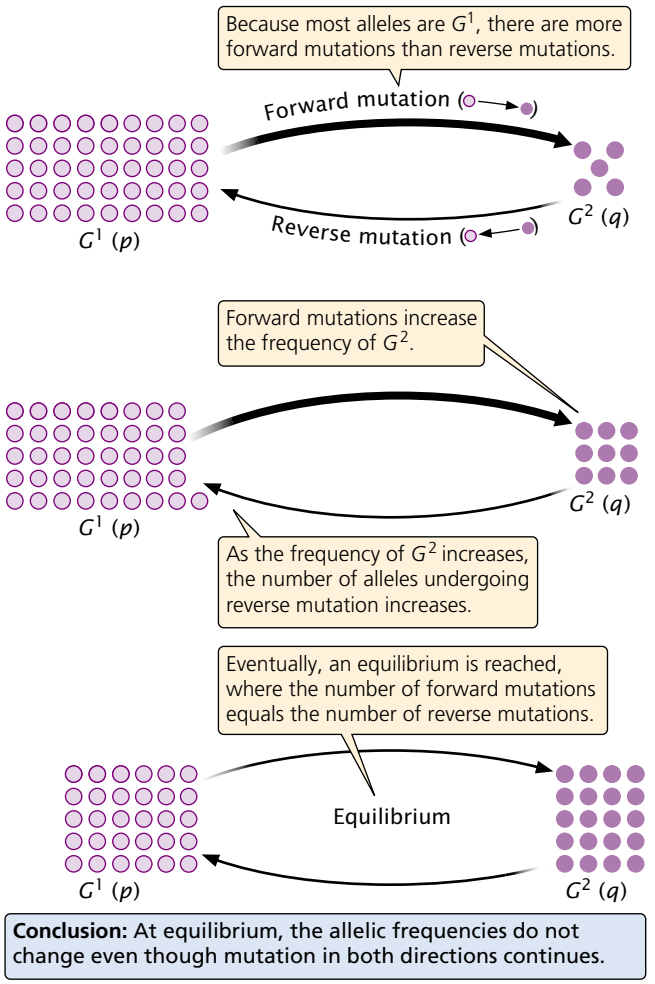
If copies of G^1 continue to mutate to G^2, the frequency of G^2 will increase and the frequency of G^1 will decrease (◖ FIGURE 23.8). The amount that G^2 will change (Δq) as a result of mutation depends on: (1) the rate of G^1-to-G^2 mutation (μ); and (2) p, the frequency of G^1 in the population When p is large, there are many copies of G^1 available to mutate to G^2, and the amount of change will be relatively large. As more mutations occur and p decreases, there will be fewer copies of G^1 available to mutate to G^2. The change in G^2 as a result of mutation equals the mutation rate times the allelic frequency:

$$\Delta q = \mu p \qquad (23.12)$$



◖ 23.6 **Inbreeding often has deleterious effects on crops.** As inbreeding increases, the average yield of corn, for example, decreases.

◖ 23.7 **Although inbreeding is generally harmful, a number inbreeding organisms are successful.**



23.8 Recurrent mutation changes allelic frequencies. Forward and reserve mutations eventually lead to a stable equilibrium.

As the frequency of p decreases as a result of mutation, the change in frequency due to mutation will be less and less

So far we have considered only the effects of $G^1 \rightarrow G^2$ forward mutations. Reverse $G^2 \rightarrow G^1$ mutations also occur at rate ν, which will probably be different from the forward mutation rate, μ. Whenever a reverse mutation occurs, the frequency of G^2 decreases and the frequency of G^1 increases (see Figure 23.8). The rate of change due to reverse mutations equals the reverse mutation rate times the allelic frequency of G^2 ($\Delta q = \nu q$). The overall change in allelic frequency is a balance between the opposing forces of forward mutation and reverse mutation:

$$\Delta q = \mu p - \nu q \qquad (23.13)$$

Reaching equilibrium of allelic frequencies Consider an allele that begins with a high frequency of G^1 and a low frequency of G^2. In this population, many copies of G^1 are initially available to mutate to G^2, and the increase in G^2 due to forward mutation will be relatively large. However, as the frequency of G^2 increases as a result of forward mutations, fewer copies of G^1 are available to mutate; so the number of forward mutations decreases. On the other hand, few copies of G^2 are initially available to undergo a reverse mutation to G^1 but, as the frequency of G^2 increases, the number of copies of G^2 available to undergo reverse mutation to G^1 increases; so the number of genes undergoing reverse mutation will increase. Eventually, the number of genes undergoing forward mutation will be counterbalanced by the number of genes undergoing reverse mutation. At this point, the increase in q due to forward mutation will be equal to the decrease in q due to reverse mutation, and there will be no net change in allelic frequency ($q = 0$), in spite of the fact that forward and reserve mutations continue to occur. The point at which there is no change in the allelic frequency of a population is referred to as **equilibrium** (see Figure 23.8).

Factors determining allelic frequencies at equilibrium We can determine the allelic frequencies at equilibrium by manipulating Equation 23.13. Recall that $p = 1 - q$. Substituting $1 - q$ for p in Equation 23.13, we get:

$$\begin{aligned} \Delta q &= \mu(1 - q) - \nu q \\ &= \mu - \mu q - \nu q \\ &= \mu - q(\mu + \nu) \end{aligned} \qquad (23.14)$$

At equilibrium, Δq will be 0; so:

$$0 = \mu - q(\mu + \nu) \qquad (23.15)$$

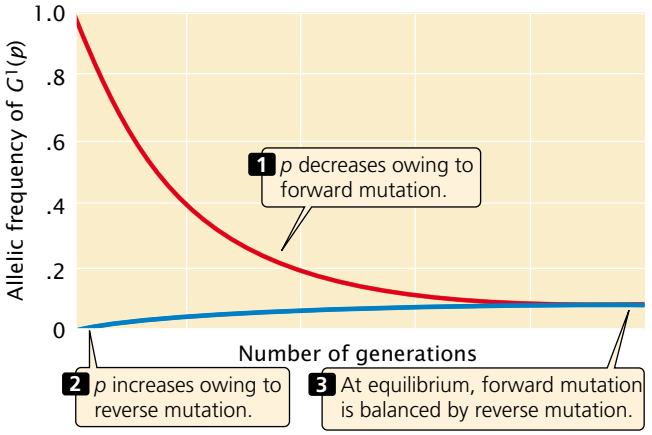
$$q(\mu + \nu) = \mu$$

$$\hat{q} = \frac{\mu}{\mu + \nu}$$

where $\hat{q}$ equals the frequency of G^2 at equilibrium. This final equation tells us that the allelic frequency at equilibrium is determined solely by the forward and reverse mutation rates.

Summary of effects When the only evolutionary force acting on a population is mutation, allelic frequencies change with the passage of time because some alleles mutate into others. Eventually, these allelic frequencies reach equilibrium and are determined only by the forward and reverse mutation rates. When the allelic frequencies reach equilibrium, the Hardy-Weinberg law tells us that genotypic frequencies also will remain the same.

The mutation rates for most genes are low; so change in allelic frequency due to mutation in one generation is very small, and long periods of time are required for a population to reach mutational equilibrium. For example, if the forward



23.9 Change due to recurrent mutation slows as the frequency of *p* drops. Allelic frequencies are approaching mutational equilibrium at typical low mutation rates. The allelic frequency of G^1 decreases as a result of forward ($G^1 \rightarrow G^2$) mutation at rate μ (.0001) and increases as a result of reverse ($G^2 \rightarrow G^1$) mutation at rate ν (.00001). Owing to the low rate of mutations, eventual equilibrium takes many generations to be reached.

and reverse mutation rates for alleles at a locus are 1×10^{-5} and 0.3×10^{-5} per generation, respectively (rates that have actually been measured at several loci in mice), and the allelic frequencies are $p = .9$ and $q = .1$, then the net change in allelic frequency per generation due to mutation is:

$$\begin{aligned}\Delta q &= \mu p - \nu q \\ &= (1 \times 10^{-5})(.9) - (.3 \times 10^{-5})(.1) \\ &= 8.7 \times 10^{-6} = .0000087\end{aligned}$$

Therefore, change due to mutation in a single generation is extremely small and, as the frequency of p drops as a result of mutation, the amount of change will become even smaller (◖ FIGURE 23.9). The effect of typical mutation rates on Hardy-Weinberg equilibrium is negligible, and many generations are required for a population to reach mutational equilibrium. Nevertheless, if mutation is the only force acting on a population for long periods of time, mutation rates will determine allelic frequencies.

Concepts

Recurrent mutation causes changes in the frequencies of alleles. At equilibrium, the allelic frequencies are determined by the forward and reverse mutation rates. Because mutation rates are low, the effect of mutation per generation is very small.

Migration

Another process that may bring about change in the allelic frequencies is the influx of genes from other populations, commonly called **migration** or **gene flow.** One of the assumptions of the Hardy-Weinberg law is that migration does not take place, but many natural populations do experience migration from other populations. The overall effect of migration is twofold: (1) it prevents genetic divergence *between* populations and (2) it increases genetic variation *within* populations.

The effect of migration on allelic frequencies Let us consider the effects of migration by looking at a simple, unidirectional model of migration between two populations that differ in the frequency of an allele *a.* Say the frequency of this allele in population I is q_I and in population II is q_{II} (◖ FIGURE 23.10a and b). In each generation, a representative sample of the individuals in population I migrates to population II (◖ FIGURE 23.10c) and reproduces, adding its genes to population II's gene pool. Migration is only from population I to population II (is unidirectional), and all the conditions of the Hardy-Weinberg law apply, except the absence of migration.

After migration, population II consists of two types of individuals (◖ FIGURE 23.10d). Some are migrants; they make up proportion m of population II, and they carry genes from population I; so the frequency of allele *a* in the migrants is q_I. The other individuals in population II are the original residents. If the migrants make up proportion m of population II, then the residents make up $1 - m$; because the residents originated in population II, the frequency of allele *a* in this group is q_{II}. After migration, the frequency of allele *a* in the merged population II (q'_{II}) is:

$$q'_{II} = q_I(m) + q_{II}(1 - m) \qquad (23.16)$$

where $q_I(m)$ is the contribution to q made by the copies of allele *a* in the migrants and $q_{II}(1 - m)$ is the contribution to q made by copies of allele *a* in the residents. The change in the allelic frequency due to migration (Δq) will be equal to the new frequency of allele *a* (q'_{II}) minus the original frequency of the allele (q_{II}):

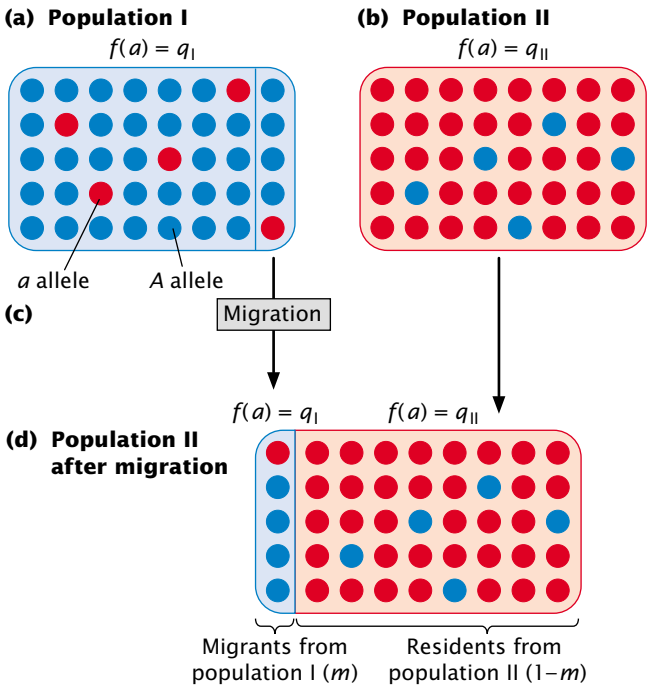
$$q_{II} = q'_{II} - q_{II}$$

In Equation 23.16, we determined that q'_{II} equals $q_I(m) + q_{II}(1 - m)$. Substituting this value for q'_{II} into the preceding equation, we get:

$$q = q_I(m) + q_{II}(1 - m) - q_{II}$$

Expanding the term $q_{II}(1 - m)$, we get:

$$q = q_I m + q_{II} - q_{II} m - q_{II}$$



Conclusion: The frequency of allele *a* in population II after migration is $q'_{II} = q_I m + q_{II}(1-m)$.

23.10 The amount of change in allelic frequency due to migration between populations depends on the difference in allelic frequency and the extent of migration. Shown here is a model of the effect of unidirectional migration on allelic frequencies. (a) The frequency of allele *a* in the source population (population I) is q_I. (b) The frequency of this allele in the recipient population (population II) is q_{II}. (c) Each generation, a random sample of individuals migrate from population I to population II. (d) After migration, population II consists of migrants and residents. The migrants constitute proportion m and have a frequency of *a* equal to q'_I; the residents constitute proportion $1 - m$ and have a frequency of *a* equal to q_{II}.

In this last equation, we are subtracting q_{II} from q_{II}, which gives us zero; so the equation simplifies to:

$$\begin{aligned} q &= q_I m - q_{II} m \\ &= m(q_I - q_{II}) \end{aligned} \qquad (23.17)$$

Equation 23.17 summarizes the factors that determine the amount of change in allelic frequency due to migration. The amount of change in q is directly proportional to the migration (m); as the amount of migration increases, the change in allelic frequency increases. The magnitude of change is also affected by the differences in allelic frequencies of the two populations ($q_I - q_{II}$); when the difference is large, the change in allelic frequency will be large.

With each generation of migration, the frequencies of the two populations become more and more similar until, eventually, the allelic frequency of population II equals that of population I. When $q_I - q_{II} = 0$, there will be no further change in the allelic frequency of population II, in spite of the fact that migration continues. If migration between two populations takes place for a number of generations with no other evolutionary forces present, an equilibrium is reached at which the allelic frequency of the recipient population equals that of the source population.

The simple model of unidirectional migration between two populations just outlined can be expanded to accommodate multidirectional migration between several populations (FIGURE 23.11).

The overall effect of migration Migration has two major effects. First, it causes the gene pools of populations to become more similar. Later, we will see how genetic drift and natural selection lead to genetic differences between populations; migration counteracts this tendency and tends to keep populations homogeneous in their allelic frequencies. Second, migration adds genetic variation to populations. Different alleles may arise in different populations owing to rare mutational events, and these alleles can be spread to new populations by migration, increasing the genetic variation within the recipient population.

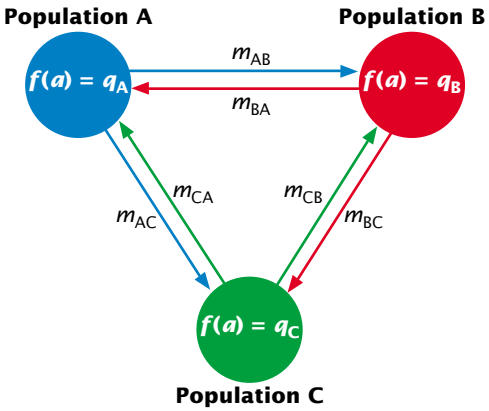
Concepts

Migration causes changes in the allelic frequency of a population by introducing alleles from other populations. The magnitude of change due to migration depends on both the extent of migration and the difference in allelic frequencies between the source and the recipient populations. Migration decreases genetic differences between populations and increases genetic variation within populations.

Genetic Drift

The Hardy-Weinberg law assumes random mating in an infinitely large population; only when population size is infinite will the gametes carry genes that perfectly represent the parental gene pool. But no real population is infinitely large, and when population size is limited, the gametes that unite to form individuals of the next generation carry a sample of alleles present in the parental gene pool. Just by chance, the composition of this sample may deviate from that of the parental gene pool, and this deviation may cause allelic frequencies to change. The smaller the gametic sample, the greater the chance that its composition will deviate from that of the entire gene pool.

The role of chance in altering allelic frequencies is analogous to flipping a coin. Each time we flip a coin, we have a 50% chance of getting a head and a 50% chance of getting a



Frequency of *a* after migration

$$q'_A = q_B m_{BA} + q_C m_{CA} + q_A(1 - m_{BA} - m_{CA})$$
$$q'_B = q_A m_{AB} + q_C m_{CB} + q_B(1 - m_{AB} - m_{CB})$$
$$q'_C = q_A m_{AC} + q_B m_{BC} + q_C(1 - m_{AC} - m_{BC})$$

23.11 Model of multidirectional migration among three populations, A, B, and C, with initial frequency of allele *a* equal to q_A, q_B, and q_C, respectively. The proportion of a population made up of migrants from other populations is designated by *m*, where the subscripts represent the source and recipient populations. For example, m_{AC} represents the proportion of population C that consists of individuals that moved from A to C. The allelic frequencies in populations A, B, and C after migration are represented by q'_A, q'_B, and q'_C.

tail. If we flip a coin 1000 times, the observed ratio of heads to tails will be very close to the expected 50:50 ratio. If, however, we flip a coin only 10 times, there is a good chance that we will obtain not exactly 5 heads and 5 tails, but rather maybe 7 heads and 3 tails or 8 tails and 2 heads. This kind of deviation from an expected ratio due to limited sample size is referred to as **sampling error.**

Sampling error occurs when gametes unite to produce progeny. Many organisms produce a large number of gametes but, when population size is small, a limited number of gametes unite to produce the individuals of the next generation. Chance influences which alleles are present in this limited sample and, in this way, sampling error may lead to changes in allelic frequency, which is called **genetic drift.** Because the deviations from the expected ratios are random, the direction of change is unpredictable. We can nevertheless predict the magnitude of the changes.

The magnitude of genetic drift The amount of sampling error resulting from genetic drift can be estimated from the variance in allelic frequency. Variance is a statistical measure that describes the degree of variability in a trait (see p. 000 in Chapter 22). Suppose that we observe a large number of separate populations, each with N individuals and allelic frequencies of p and q. After one generation of random mating, genetic drift expressed in terms of the variance in allelic frequency among the populations (s_p^2) will be:

$$s_p^2 = \frac{pq}{2N} \qquad (23.18)$$

The amount of change resulting from genetic drift (the variance in allelic frequency) is determined by two parameters: the allelic frequencies (p and q) and the population size (N). Genetic drift will be maximal when p and q are equal (each .5) and when the population size is small.

The effect of population size on genetic drift is illustrated by a study conducted by Luca Cavalli-Sforza and his colleagues. They studied variation in blood types among villagers in the Parm Valley of Italy, where the amount of migration between villages was limited. They found that variation in allelic frequency was greatest between small isolated villages in the upper valley but decreased between larger villages and towns farther down the valley. This result is exactly what we expect with genetic drift: there should be more genetic drift and thus more variation among villages when population size is small.

For ecological and demographic studies, population size is usually defined as the number of individuals in a group. The evolution of a gene pool depends, however, only on those individuals who contribute genes to the next generation. Population geneticists usually define population size as the equivalent number of breeding adults, the **effective population size** (N_e). Several factors determine the equivalent number of breeding adults. One factor is the sex ratio. When the numbers of males and females in the population are equal, the effective population size is simply the sum of reproducing males and females. When they are unequal, then the effective population size is:

$$N_e = \frac{4 \times n_{males} \times n_{females}}{n_{males} + n_{females}} \qquad (23.19)$$

Table 23.3 gives the effective population size for a theoretical population of 100 individuals with different proportions of males and females. Notice that, when the number of males and females is unequal, the effective population size is smaller than it is when the number of males and females is the same. For example, when a population consists of 90 males and 10 females, the effective population size is only 36, and genetic drift will occur as though the actual population consisted of only 36 individuals, equally divided between males and females. A population with 90 males and 10 females has the same effective population size as a population with 10 males and 90 females—it makes no difference which sex is in excess.

Table 23.3 Effective population size (N_e) in theoretical populations of 100 individuals, each with a different sex ratio

Sex Ratio*	Number of Males	Number of Females	N_e
1.00	50	50	100
3.00	75	25	75
0.33	25	75	75
9.00	90	10	36
0.10	10	90	36
99.00	99	1	3.96
0.01	1	99	3.96

*The sex ratio is the ratio of the number of males to the number of females.

The reason that the sex ratio influences genetic drift is that half the genes in the gene pool come from males and half come from females. When one sex is present in low numbers, genetic drift increases because half of the genes are coming from a small number of individuals. In a population consisting of 10 males and 90 females, the overall population size is relatively large (100), but only 10 males contribute half the genes to the next generation. Sampling error therefore affects the range of genes present in the male gametes, and chance will have a major effect on the allelic frequencies of the next generation.

Other factors that influence effective population size include variation between individuals in reproductive success, fluctuations in population size, the age structure of the population, and whether mating is random.

Concepts

Genetic drift is change in allelic frequency due to chance factors. The amount of change in allelic frequency due to genetic drift is inversely related to the effective population size (the equivalent number of breeding adults in a population). Effective population size decreases when there are unequal numbers of breeding males and females.

Causes of genetic drift All genetic drift arises from sampling error, but there are several different ways in which sampling error can arise. First, a population may be reduced in size for a number of generations because of limitations in space, food, or some other critical resource. Genetic drift in a small population for multiple generations can significantly affect the composition of a population's gene pool.

A second way that sampling error can arise is through the **founder effect,** which is due to the establishment of a population by a small number of individuals; the population of Tristan da Cuna, discussed in the introduction to this chapter, underwent a founder effect. Although a population may increase and become quite large, the genes carried by all its members are derived from the few genes originally present in the founders (assuming no migration or mutation). Chance events affecting which genes were present in the founders will have an important influence on the makeup of the entire population. The small number of founders of Tristan da Cuna included two sisters and a daughter who suffered from asthma; the high incidence of asthma on the island today can be traced to alleles carried by these founders.

A third way that genetic drift arises is through a **genetic bottleneck,** which develops when a population undergoes a drastic reduction in population size. A genetic bottleneck developed in northern elephant seals (◖ FIGURE 23.12). Before 1800, thousands of elephant seals were found along the California coast, but the population was devastated by hunting between 1820 and 1880. By 1884, as few as 20 seals survived on a remote beach of Isla de Guadelupe west of Baja, California. Restrictions on hunting enacted by the United States and Mexico allowed the seals to recover, and there are now more than 30,000 seals in the population. All seals in the population today are genetically similar, because they have genes that were carried by the few survivors of the population bottleneck.

The effects of genetic drift Genetic drift has several important effects on the genetic composition of a population. First, it produces change in allelic frequencies within a population. Because drift is random, allelic frequency is just as likely to increase as it is to decrease and will wander with the passage of time (hence the name genetic drift). ◖ FIGURE 23.13 illustrates a computer simulation of genetic drift in five populations over 30 generations, starting with $q = .5$ and maintaining a constant population size of 10 males and 10 females. These allelic frequencies change randomly from generation to generation.

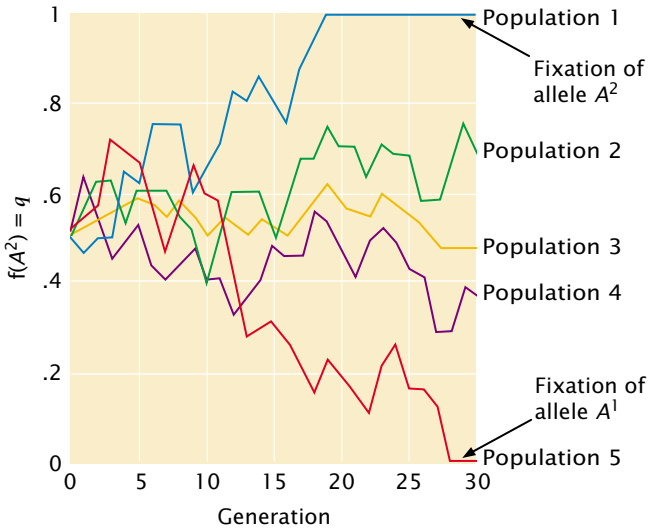
A second effect of genetic drift is to reduce genetic variation within populations. Through random change, an allele may eventually reach a frequency of either 1 or 0, at which point all individuals in the population are homozygous for one allele. When an allele has reached a frequency of 1, we say that it has reached **fixation.** Other alleles are lost (reach a frequency of 0) and can be restored only by migration from another population or by mutation. Fixation, then, leads to a loss of genetic variation within a population. This loss can be seen in northern elephant seals. Today, these seals have low levels of genetic variation; a study of 24 protein-encoding genes found no individual or population differences in these genes.

Given enough time, all small populations will become fixed for one allele. Which allele becomes fixed is random and is determined by the initial frequency of the allele. If

23.12 Northern elephant seals underwent a severe genetic bottleneck between 1820 and 1880. Today, these seals have low levels of genetic variation. (Lisa Husar/DRK Photo.)

the population begins with two alleles, each with a frequency of .5, both alleles have an equal probability of fixation. However, if one allele is initially common, it is more likely to become fixed.

A third effect of genetic drift is that different populations diverge genetically with time. In Figure 23.13, all five populations begin with the same allelic frequency ($q = .5$) but, because drift occurs randomly, the frequencies in different populations do not change in the same way, and so



23.13 Genetic drift changes allelic frequencies within populations, leading to a reduction in genetic variation through fixation and genetic divergence among populations. Shown here is a computer simulation of changes in the frequency of allele A^2 (q) in five different populations due to random genetic drift. Each population consists of 10 males and 10 females and begins with $q = .5$.

populations gradually acquire genetic differences. Notice that, although the variance in allelic frequency among the populations increases, the average allelic frequency remains basically the same. Eventually, all the populations reach fixation; some will become fixed for one allele and others will become fixed for the alternative allele. This divergence of populations through genetic drift is strikingly illustrated in the results of an experiment carried out by Peter Buri on fruit flies (FIGURE 23.14).

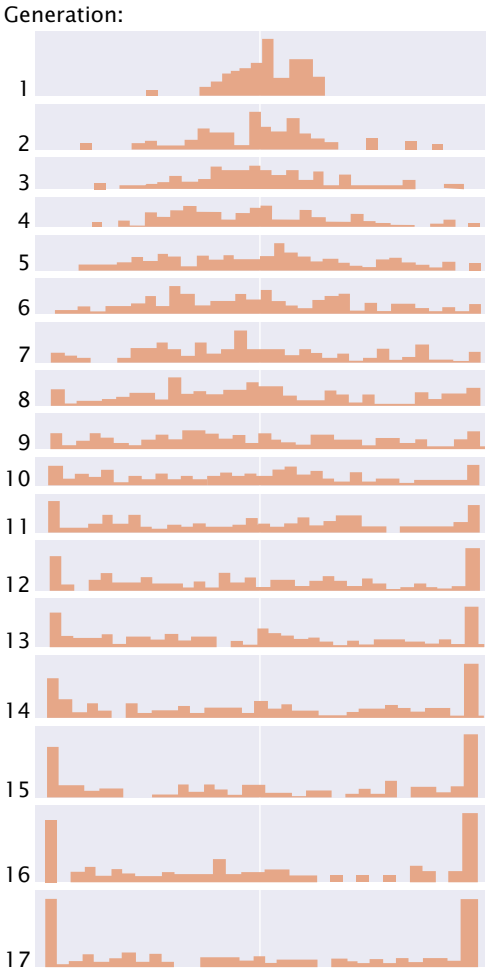
The three results of genetic drift (allelic frequency change, loss of variation within populations, and genetic divergence between populations) occur simultaneously, and all result from sampling error. The first two results occur *within* populations, whereas the third occurs *between* populations.

Concepts

Genetic drift results from continuous small population size, founder effect (establishment of a population by a few founders), and bottleneck effect (population reduction). Genetic drift causes change in allelic frequencies within a population, loss of genetic variation through fixation of alleles, and genetic divergence between populations.

Natural Selection

A final process that brings about changes in allelic frequencies is natural selection, the differential reproduction of genotypes (see p. 000 in Chapter 22). Natural selection takes place when individuals with adaptive traits produce more offspring. If the adaptive traits have a genetic basis, they are inherited by the offspring and appear with greater frequency



23.14 Populations diverge in allelic frequency and become fixed for one allele as a result of genetic drift. In this experiment, Buri examined the frequency of two alleles (bw^{75} and bw) that affect *Drosophila* eye color in 107 replicate populations. Each population consisted of 8 males and 8 females; each population began with the frequency of bw^{75} equal to .5.

in the next generation. A trait that provides a reproductive advantage thereby increases over time, enabling populations to become better suited to their environments—to become better adapted. Natural selection is unique among evolutionary forces in that it promotes adaptation (◖ FIGURE 23.15).

Fitness and selection coefficient The effect of natural selection on the gene pool of a population depends on the fitness values of the genotypes in the population. **Fitness** is defined as the relative reproductive success of a genotype. Here the term *relative* is critically important: fitness is the reproductive success of one genotype compared with the reproductive successes of other genotypes in the population.

Fitness (W) ranges from 0 to 1. Suppose the number of viable offspring produced by three genotypes is:

Genotypes:	A^1A^1	A^1A^2	A^2A^2
Mean number of offspring produced:	10	5	2

To calculate fitness for each genotype, we take the average number of offspring produced by a genotype and divide it by the mean number of offspring produced by the most prolific genotype:

$$\text{Fitness } (W): \quad \underset{W_{11} = \frac{10}{10} = 1.0}{\boldsymbol{A^1A^1}} \qquad \underset{W_{12} = \frac{5}{10} = .5}{\boldsymbol{A^1A^2}}$$

$$\underset{W_{22} = \frac{2}{10} = .2}{\boldsymbol{A^2A^2}} \tag{23.20}$$

The fitness of the genotype A^1A^1 is designated W_{11}, that of A^1A^2 is W_{12}, and that of A^2A^2 is W_{22}. A related variable is the **selection coefficient** (s), which is the relative intensity of selection against a genotype. The selection coefficient is equal to $1 - W$; so the selection coefficients for the preceding three genotypes are:

23.15 Natural selection produces adaptations, such as those seen in polar bears that inhabit the extreme Arctic environment. These bears blend into the snowy background, which helps them in hunting seals. The hairs of their fur stay erect even when wet, and thick layers of blubber provide insulation, which protects against subzero temperatures. Their digestive tracts are adapted to a seal-based carnivorous diet. (Tom and Pat Leeson/DRK Photo.)

	A^1A^1	A^1A^2	A^2A^2
Selection coefficient $(1 - W)$:	$s_{11} = 0$	$s_{12} = .5$	$s_{22} = .8$

We usually speak of selection for a particular genotype, but keep in mind that, when selection is *for* one genotype, selection is automatically *against* at least one other genotype.

Concepts

Natural selection is the differential reproduction of genotypes. It is measured as fitness, which is the reproductive success of a genotype compared with other genotypes in a population.

The general selection model Differential fitness among genotypes over time leads to changes in the frequencies of the genotypes, which, in turn, lead to changes in the frequencies of the alleles that make up the genotypes. We can predict the effect of natural selection on allelic frequencies by using a general selection model, which is outlined in Table 23.4. Use of this model requires knowledge of both the initial allelic frequencies and the fitness values of the genotypes. It assumes that mating is random and the only force acting on a population is natural selection.

We have defined fitness in terms of relative reproduction, but it will be easier to understand the logic behind the general selection model if we think of the fitness of the genotypes as differences in survival. It applies equally to fitnesses representing differential reproduction.

Let's apply the general selection model outlined in Table 23.4. Imagine a flock of sparrows overwintering in Rochester, New York. Assume that we can determine the genotypes for a locus that affects the ability of the birds to survive the winter; perhaps the genes at this locus determine the amount of fat that a bird accumulates before the onset of winter. For genotypes A^1A^1, A^1A^2, and A^2A^2, p represents the frequency of A^1 and q represents the frequency of A^2. On the first line of the table, we record the initial genotypic frequencies before selection has acted, before the onset of winter. If mating has been random (an assumption of the model), the genotypes will have the Hardy-Weinberg equilibrium frequencies of p^2, $2pq$, and q^2. On the second row of the table, we put the fitness values of the corresponding genotypes. Some of the birds die in the winter; so here the fitness values represent the relative survival of the three genotypes. The proportion of the population represented by each genotype after selection is obtained by multiplying the initial genotypic frequency times its fitness (third row of Table 23.4). Now the genotypes are no longer in Hardy-Weinberg equilibrium.

The mean fitness ($\overline{W}$) of the population is the sum of the proportionate contributions of the three genotypes:

$$\overline{W} = p^2W_{11} + 2pqW_{12} + q^2W_{22} \qquad (23.21)$$

The mean fitness $\overline{W}$ is the average fitness of all individuals in the population and allows the frequencies of the genotypes after selection to be obtained. In our flock of birds, these frequencies will be those of the three genotypes after the winter mortality. The frequency of a genotype after selection will be equal to its proportionate contribution divided by the mean fitness of the population ($p^2W_{11}/\overline{W}$ for genotype A^1A^1, $2pqW_{12}/\overline{W}$ for genotype A^1A^2, and $q^2W_{22}/\overline{W}$ for genotype A^2A^2), as shown in the fourth line of Table 23.4. When the new genotypic frequencies have been calculated, the new allelic frequency of A^1 (p') can be determined by using the now-familiar formula (Equation 23.4):

$$p' = f(A^1) = f(A^1A^1) + \tfrac{1}{2}f(A^1A^2)$$

and that of q' can be obtained by subtraction:

$$q' = 1 - p'$$

Table 23.4 Method for determining changes in allelic frequency due to selection

	A^1A^1	A^1A^2	A^2A^2
Initial genotypic frequencies	p^2	$2pq$	q^2
Fitnesses	W_{11}	W_{12}	W_{22}
Proportionate contribution of genotypes to population	p^2W_{11}	$2pqW_{12}$	q^2W_{22}
Relative genotypic frequency after selection	$\frac{p^2W_{11}}{\overline{W}}$	$\frac{2pqW_{22}}{\overline{W}}$	$\frac{q^2W_{22}}{\overline{W}}$

Note: $\overline{W} = p^2W_{11} + 2pqW_{12} + q^2W_{22}$
Allelic frequencies after selection: $p' = f(A^1) = f(A^1A^1) + \frac{1}{2}f(A^1A^2)$
$q' = 1 - p$

The last step is to determine the genotypic frequencies in the *next* generation. In regard to our birds, these genotypic frequencies are those of the offspring of the birds that survived the winter. If the survivors mate randomly, the genotypic frequencies in the next generation will be p'^2, $2p'q'$, and q'^2.

The general selection model can be used to calculate the allelic frequencies after any type of selection. It is also possible to work out formulas for determining the change in allelic frequency when selection is against recessive, dominant, and codominant traits, as well as traits in which the heterozygote has highest fitness (Table 23.5).

Concepts

The change in allelic frequency due to selection can be determined for any type of genetic trait by using the general selection model.

The results of selection The results of selection depend on the relative fitnesses of the genotypes. If we have three genotypes (A^1A^1, A^1A^2, and A^2A^2) with fitnesses W_{11}, W_{12}, and W_{22}, we can identify six different types of natural selection (Table 23.6). In type 1 selection, a dominant allele A^1 confers a fitness advantage; in this case, the fitnesses of genotypes A^1A^1 and A^1A^2 are equal and higher than the fitness of A^2A^2 ($W_{11} = W_{12} > W_{22}$). Because the heterozygote and the A^1A^1 homozygote both have copies of the A^1 allele and produce more offspring than the A^2A^2 homozygote does, the frequency of the A^1 allele will increase over time, whereas the frequency of the A^2 allele will decrease. This form of selection, in which one allele or trait is favored over another, is termed **directional selection.**

Type 2 selection (Table 23.6) is directional selection against a dominant allele A^1 ($W_{11} = W_{12} < W_{22}$). In this case, the A^2 allele increases and the A^1 allele decreases. Type 3 and type 4 selection also are directional selection, but in these cases there is incomplete dominance and the heterozygote has a fitness that is intermediate between the two homozygotes ($W_{11} < W_{12} < W_{22}$ for type 3; $W_{11} > W_{12} > W_{22}$ for type 4). When A^1A^1 has the highest fitness (type 3), over time the A^1 allele increases and the A^2 allele decreases. When A^2A^2 has the highest fitness (type 4), over time the A^2 allele increases and the A^1 allele decreases. Eventually, directional selection leads to fixation of the favored allele and elimination of the other allele, as long as no other evolutionary forces act on the population.

Two types of selection (types 5 and 6) are special situations that lead to equilibrium, where there is no further change in allelic frequency. Type 5 selection is referred to as **overdominance** or heterozygote advantage. Here, the heterozygote has higher fitness than the fitnesses of the two homozygotes ($W_{11} < W_{12} > W_{22}$). With overdominance, both alleles are favored in the heterozygote, and neither allele is eliminated from the population. Initially, the allelic frequencies may change because one homozygote has higher fitness than the other; the direction of change will depend on the relative fitness values of the two homozygotes. The allelic frequencies change with overdominant selection until a stable equilibrium is reached, at which point there is no further change. The allelic frequency at equilibrium ($\hat{q}$) depends on the relative fitnesses (usually expressed as selection coefficients) of the two homozygotes:

$$\hat{q} = f(A^2) = \frac{s_{11}}{s_{11} + s_{22}} \qquad (23.22)$$

where s_{11} represents the selection coefficient of the A^1A^1 homozygote and s_{22} represents the selection coefficient of the A^2A^2 homozygote.

The last type of selection (type 6) is **underdominance,** in which the heterozygote has lower fitness than both

Table 23.5 Formulas for calculating change in allelic frequencies with different types of selection

	Fitness Values			
Type of Selection	A^1A^1	A^1A^2	A^2A^2	**Change in q**
Selection against a recessive trait	1	1	$1 - s$	$\frac{-spq^2}{1 - sp^2}$
Selection against a dominant trait	1	$1 - s$	$1 - s$	$\frac{-spq^2}{1 - s + sq^2}$
Selection against a trait with no dominance	1	$1 - \frac{1}{2}s$	$1 - s$	$\frac{-\frac{1}{2}spq}{1 - sq}$
Selection against both homozygotes (overdominance)	$1 - s_{11}$	1	$1 - s_{22}$	$\frac{pq(s_{11}p - s_{22}q)}{1 - s_{11}p^2 - s_{22}q^2}$

Table 23.6 Types of natural selection

Type	Fitness Relation	Form of Selection	Result
1	$W_{11} = W_{12} > W_{22}$	Directional selection against recessive allele A^2	A^1 increases, A^2 decreases
2	$W_{11} = W_{12} < W_{22}$	Directional selection against dominant allele A^1	A^2 increases, A^1 decreases
3	$W_{11} > W_{12} > W_{22}$	Directional selection against incompletely dominant allele A^2	A^1 increases, A^2 decreases
4	$W_{11} < W_{12} < W_{22}$	Directional selection against incompletely dominant allele A^1	A^2 increases, A^1 decreases
5	$W_{11} < W_{12} > W_{22}$	Overdominance	Stable equilibrium, both alleles maintained
6	$W_{11} > W_{12} < W_{22}$	Underdominance	Unstable equilibrium

Note: W_{11}, W_{12}, and W_{22} represent the fitnesses of genotypes A^1A^1, A^1A^2, and A^2A^2, respectively.

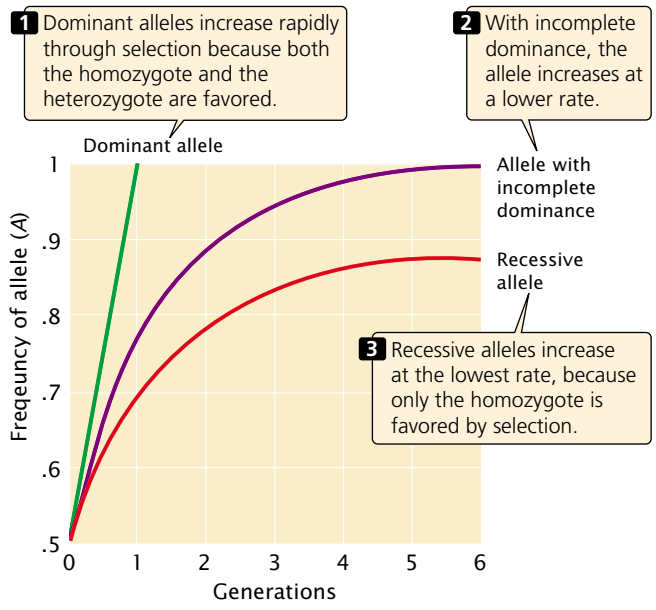
homozygotes ($W_{11} > W_{12} < W_{22}$). Underdominance leads to an *unstable* equilibrium; here allelic frequencies will not change as long as they are at equilibrium but, if they are disturbed from the equilibrium point by some other evolutionary force, they will move away from equilibrium until one allele eventually becomes fixed.

Concepts

Natural selection changes allelic frequencies; the direction and magnitude of change depends on the intensity of selection, the dominance relations of the alleles, and the allelic frequencies. Directional selection favors one allele over another and eventually leads to fixation of the favored allele. Overdominance leads to a stable equilibrium with maintenance of both alleles in the population. Underdominance produces an unstable equilibrium because the heterozygote has lower fitness than those of the two homozygotes.

The rate of change in allelic frequency due to natural selection The rate at which an allele changes in frequency owing to selection depends on the intensity of selection and the dominance relations among the genotypes (FIGURE 23.16). Under directional selection, dominant alleles will increase much more rapidly than recessive alleles, because homozygotes and heterozygotes are favored. With incomplete dominance, the heterozygote has a selective advantage, but not as much as the homozygote; so incompletely dominant alleles increase in frequency at a lower rate than that of dominant alleles. Recessive alleles increase at the lowest rate, because only the homozygote is favored by selection.

The rate at which selection changes allelic frequencies also depends on the allelic frequency itself. If an allele (A^2) is lethal and recessive, $W_{11} = W_{12} = 1$, whereas $W_{22} = 0$. The



23.16 The rate of change in allelic frequency due to selection depends on the dominance relations among the genotypes. Here, change in the frequency of an allele is shown for different types of dominance with a constant selection coefficient.

frequency of the A^2 allele will decrease over time (because the A^2A^2 homozygote produces no offspring), and the rate of decrease will be proportional to the frequency of the recessive allele. When the frequency of the allele is high, the change in each generation is relatively large but, as the frequency of the allele drops, a higher proportion of the alleles are in the heterozygous genotypes, where they are immune to the action of natural selection (the heterozygotes have the same phenotype as the favored homozygote). Thus, selection against a rare recessive allele is very inefficient and its removal from the population is slow.

The relation between the frequency of a recessive allele and its rate of change under natural selection has an important implication. Some people believe that the medical treatment of patients with rare recessive diseases will cause the disease gene to increase, eventually leading to degeneration of the human gene pool. This mistaken belief was the basis of eugenic laws that were passed in the early part of the twentieth century prohibiting the marriage of persons with certain genetic conditions and allowing the involuntary sterilization of others. However, most copies of rare recessive alleles are present in heterozygotes, and selection against the homozygotes will have little effect on the frequency of a recessive allele. Thus whether the homozygotes reproduce or not has little effect on the frequency of the disorder.

Mutation and natural selection Recurrent mutation and natural selection act as opposing forces on detrimental alleles; mutation increases their frequency and natural selection decreases their frequency. Eventually, these two forces reach an equilibrium, in which the number of alleles added by mutation is balanced by the number of alleles removed by selection.

Table 23.5 shows that the change in allelic frequency due to selection against a recessive allele is $-spq^2/(1 - sq^2)$. When q is very low, q^2 is near zero; so $1 - sq^2$ will be approximately $1 - 0 = 1$. Thus, when q is very low, the decrease in frequency due to selection is approximately $-spq^2$. The increase in frequency of an allele due to forward mutations is μp (Equation 23.12). At equilibrium, the effects of mutation and selection are balanced; so

$$spq^2 = \mu p$$

This equation can be rearranged:

$$q^2 = \frac{\mu p}{sp} = \frac{\mu}{s}$$

Taking the square root of each side, we get

$$\hat{q} = \frac{\mu}{s} \qquad (23.23)$$

The frequency of the allele at equilibrium ($\hat{q}$) is therefore equal to the square root of the mutation rate divided by the selection coefficient. With the use of the equation for selection acting on a dominant allele (see Table 23.5) and similar reasoning, the frequency of a dominant allele at equilibrium can be shown to be

$$\hat{q} = \frac{\mu}{s} \qquad (23.24)$$

Achondroplasia (discussed in Chapter 17) is a common type of human dwarfism that results from a dominant gene. People with this condition are fertile, although they produce only about 74% as many children as are produced by people without achondroplasia. The fitness of people with achondroplasia therefore averages .74, and the selection coefficient (s) is $1 - W$, or .26. If we assume that the mutation rate for achondroplasia is about 3×10^{-5} (a typical mutation rate in humans), then we can predict that the equilibrium frequency for the achondroplasia allele will be $\hat{q} = (.00003/.26) = .0001153$. This frequency is close to the actual frequency of the disease.

Concepts

Mutation and natural selection act as opposing forces on detrimental alleles: mutation tends to increase their frequency and natural selection tends to decrease their frequency, eventually producing an equilibrium.

Connecting Concepts

The General Effects of Evolutionary Forces

You now know that four processes bring about change in the allelic frequencies of a population: mutation, migration, genetic drift, and natural selection. Their short- and long-term effects on allelic frequencies are summarized in Table 23.7. In some cases, these changes continue until one allele is eliminated and the other becomes fixed in the population. Genetic drift and directional selection will eventually result in fixation, provided these forces are the only ones acting on a population. With the other evolutionary forces, allelic frequencies change until an equilibrium point is reached, and then there is no additional change in allelic frequency. Mutation, migration, and some forms of natural selection can lead to stable equilibria (see Table 23.7).

Table 23.7 Effects of different evolutionary forces on allelic frequencies within populations

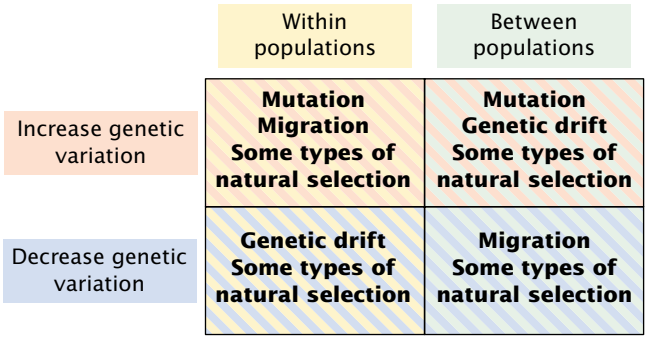
Force	Short-Term Effect	Long-Term Effect
Mutation	Change in allelic frequency	Equilibrium reached between forward and reverse mutations
Migration	Change in allelic frequency	Equilibrium reached when allelic frequencies of source and recipient population are equal
Genetic drift	Change in allelic frequency	Fixation of one allele
Natural selection	Change in allelic frequency	Directional selection: fixation of one allele
		Overdominant selection: equilibrium reached

The different evolutionary forces affect both genetic variation within populations and genetic divergence between populations. Evolutionary forces that maintain or increase genetic variation within populations are listed in the upper-left quadrant of Figure 23.17. These forces include some types of natural selection, such as overdominance in which both alleles are favored. Mutation and migration also increase genetic variation within populations because they introduce new alleles to the population. Evolutionary forces that decrease genetic variation within populations are listed in the lower-left quadrant of Figure 23.17. These forces include genetic drift, which decreases variation through fixation of alleles, and some forms of natural selection such as directional selection.

The various evolutionary forces also affect the amount of genetic divergence between populations. Natural selection increases divergence among populations if different alleles are favored in the different populations, but it can also *decrease* divergence between populations by favoring the same allele in the different populations. Mutation almost always increases divergence between populations because different mutations arise in each population. Genetic drift also increases divergence between populations because changes in allelic frequencies due to drift are random and are likely to change in different directions in separate populations. Migration, on the other hand, decreases divergence between populations because it makes populations similar in their genetic composition.

Migration and genetic drift act in opposite directions: migration increases genetic variation within populations and decreases divergence between populations, whereas genetic drift decreases genetic variation within populations and increases divergence among populations. Mutation increases both variation within populations and divergence between populations. Natural selection can either increase or decrease variation within populations, and it can increase or decrease divergence between populations.

It is important to keep in mind that real populations are simultaneously affected by many evolutionary forces. This discussion has examined the effects of mutation, migration, genetic drift, and natural selection in isolation so that the influence of each process would be clear. However, in the real world, populations are commonly affected by several evolutionary forces at the same time, and evolution results from the complex interplay of numerous processes.



23.17 Mutation, migration, genetic drift, and natural selection have different effects on genetic variation within populations and on genetic divergence between populations.

www.whfreeman.com/pierce Web addresses for a number of resources on evolution

Molecular Evolution

For many years, it was not possible to examine genes directly, and evolutionary biology was confined largely to the study of how phenotypes change with the passage of time. The tremendous advances in molecular genetics in recent years

have made it possible to investigate evolutionary change directly by analyzing protein and nucleic acid sequences. These molecular data offer a number of advantages for studying the process and pattern of evolution:

1. **Molecular data are genetic.** Evolution results from genetic change over time. Anatomical, behavioral, and physiological traits often have a genetic basis, but the relation between the underlying genes and the trait may be complex. Protein and nucleic acid sequence variation has a clear genetic basis that is easy to interpret.
2. **Molecular methods can be used with all organisms.** Early studies of population genetics relied on simple genetic traits such as human blood types or banding patterns in snails, which are restricted to a small group of organisms. However, all living organisms have proteins and nucleic acids; so molecular data can be collected from any organism.
3. **Molecular methods can be applied to a huge amount of genetic variation.** An enormous amount of data can be accessed by molecular methods. The human genome, for example, contains more than 3 billion base pairs of DNA, which constitutes a large pool of information about our evolution.
4. **All organisms can be compared with the use of some molecular data.** Trying to assess the evolutionary history of distantly related organisms is often difficult because they have few characteristics in common. The evolutionary relationships between angiosperms were traditionally assessed by comparing floral anatomy, whereas the evolutionary relationships of bacteria were determined by their nutritional and staining properties. Because plants and bacteria have so few structural characteristics in common, evaluating how they are related to one another was difficult in the past. All organisms have certain molecular traits in common, such as ribosomal RNA sequences and some fundamental proteins. These molecules offer a valid basis for comparisons among all organisms.
5. **Molecular data are quantifiable.** Protein and nucleic acid sequence data are precise, accurate, and easy to quantify, which facilitates the objective assessment of evolutionary relationships.
6. **Molecular data often provide information about the process of evolution.** Molecular data can reveal important clues about the process of evolution. For example, the results of a study of DNA sequences have revealed that one type of insecticide resistance in mosquitoes probably arose from a single mutation that subsequently spread throughout the world.
7. **The database of molecular information is large and growing.** Today, this database of DNA and protein sequences can be used for making evolutionary comparisons and inferring mechanisms of evolution.

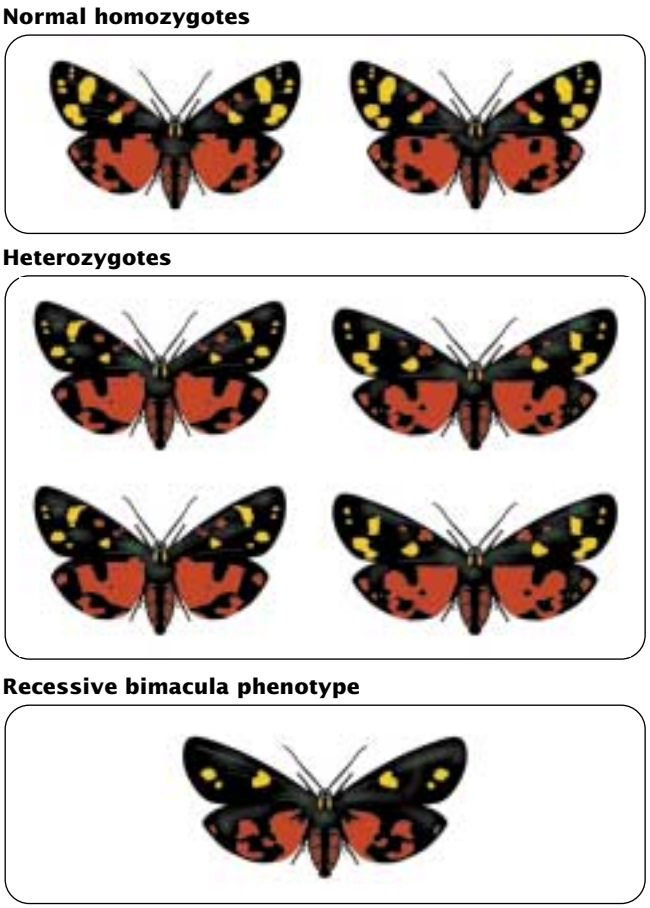
Studies of molecular evolution fall into three primary areas. First, much past research has focused on determining the extent and causes of genetic variation in natural populations. Molecular techniques allow these matters to be addressed directly by examining sequence variation in proteins and DNA. A second area of research examines molecular processes that influence evolutionary events, and the results of these studies have elucidated new mechanisms and processes of evolution that were not suspected before the application of molecular techniques to evolutionary biology. A third area of research in molecular evolution applies molecular techniques to constructing **phylogenies** (evolutionary trees) of various groups of organisms. A detailed evolutionary history is found in the DNA sequences of every organism, and molecular techniques allow this history to be read.

Concepts

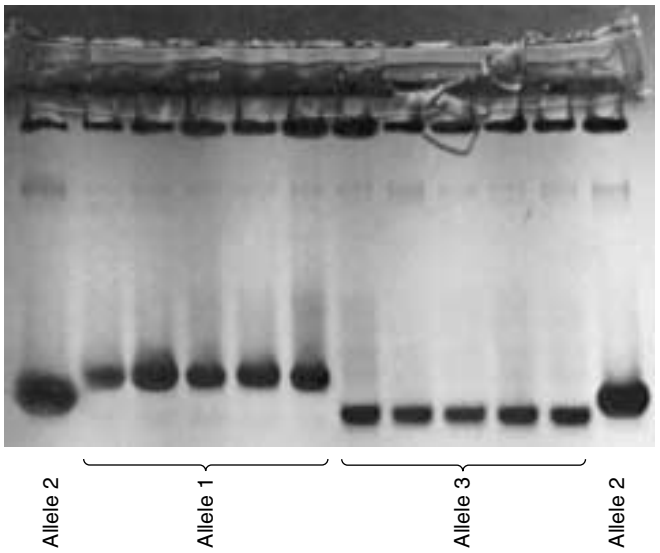
Molecular techniques and data offer a number of advantages for evolutionary studies. Molecular data (1) are genetic in nature and can be investigated in all organisms; (2) provide potentially large data sets; (3) allow all organisms to be compared, by using the same characteristics; (4) are easily quantifiable; and (5) provide information about the process of evolution.

Protein Variation

The study of the amounts and kinds of genetic variation in natural populations is central to the study of evolution. For many traits, a complex interaction of many genes and environmental factors determines the phenotype, and assessing the amount of genetic variation by examining phenotypic variation was difficult. Early population geneticists were forced to rely on the phenotypic traits that had a simple genetic basis, such as human blood types or spotting patterns in butterflies (◖ FIGURE 23.18). The initial breakthrough that first allowed the direct examination of molecular evolution was the application of electrophoresis (see Figure 18.4) to population studies. This technique separates macromolecules, such as proteins or nucleic acids, on the basis of their size and charge. In 1966, Richard Lewontin and John Hubby extracted proteins from wild fruit flies, separated the proteins by electrophoresis, and stained for specific enzymes. Examining the pattern of bands on gels enabled them to assign genotypes to individual flies and to quantify the amount of genetic variation in natural populations. In the same year, Harry Harris quantified genetic variation in human populations by using the same technique. Protein variation has now been examined in hundreds of different species by using protein electrophoresis (◖ FIGURE 23.19).



23.18 Early population geneticists were forced to rely on the phenotypic traits that had a simple genetic basis. Variation in the spotting patterns of the butterfly *Panaxia dominula* is an example.



23.19 Molecular variation in proteins is revealed by electrophoresis. Tissue samples from three fruit flies have been subjected to electrophoresis and stained for malate dehydrogenase. Homozygotes are represented as single bands; heterozygotes as triple bands. The genotype of each fly is given below each sample.

Measures of genetic variation The amount of genetic variation in populations is commonly measured by two parameters. The **proportion of polymorphic loci** is the proportion of examined loci in which more than one allele is present in a population. If we examined 30 different loci and found two or more alleles present at 15 of these loci, the percentage of polymorphic loci would be $15/30 = 0.5$. The **expected heterozygosity** is the proportion of individuals that are expected to be heterozygous at a locus under the Hardy-Weinberg conditions, which is $2pq$ when there are two alleles present in the population. The expected heterozygosity is often preferred to the observed heterozygosity because expected heterozygosity is independent of the breeding system of an organism. For example, if a species self-fertilizes, it may have little or no heterozygosity but still have considerable genetic variation, which will be detected by the expected heterozygosity. Expected heterozygosity is typically calculated for a number of loci and is then averaged over all the loci examined.

The percentage of polymorphic loci and the expected heterozygosity have been determined by protein electrophoresis for a number of species (Table 23.8). About one-third of all protein loci are polymorphic, and expected heterozygosity averages about 10%, although there is considerable diversity among species. These measures actually underestimate the true amount of genetic variation, though, because protein electrophoresis does not detect some amino acid substitutions; nor does it detect genetic variation in DNA that does not alter the amino acids of a protein (synonymous codons and variation in noncoding regions of the DNA).

Explanations for protein variation By the late 1970s, geneticists recognized that most populations possess large amounts of genetic variation, although the evolutionary significance of this fact was not at all clear. Two opposing hypotheses arose to account for the presence of the extensive molecular variation in proteins. The **neutral-mutation hypothesis** proposed that the molecular variation revealed by protein electrophoresis is adaptively neutral; that is, individuals with different molecular variants have equal fitness. This hypothesis does not propose that the proteins are functionless; rather, it suggests that most variants revealed by protein electrophoresis are functionally equivalent. Because these variants are functionally equivalent, natural selection does not differentiate between them, and their evolution is shaped largely by the random processes of genetic drift and mutation. The neutral-mutation hypothesis accepts that

Table 23.8 Proportion of polymorphic loci and heterozygosity for different organisms, as determined by protein electrophoresis

		Proportion of Polymorphic Loci		Heterozygosity	
Group	**Number of Species**	**Mean**	**SD***	**Mean**	**SD***
Plants	15	0.26	0.17	0.07	0.07
Invertebrates (excluding insects)	28	0.40	0.28	0.10	0.07
Insects (excluding *Drosophila*)	23	0.33	0.20	0.07	0.08
Drosophila	32	0.43	0.13	0.14	0.05
Fish	61	0.15	0.01	0.05	0.04
Amphibians	12	0.27	0.13	0.08	0.04
Reptiles	15	0.22	0.13	0.05	0.02
Birds	10	0.15	0.11	0.05	0.04
Mammals	46	0.15	0.10	0.04	0.02

* SD, standard deviation from the mean.
Source: After L. E. Mettler, T. G. Gregg, and H. E. Schaffer, *Population Genetics and Evolution,* 2d ed. (Englewood Cliffs, NJ: Prentice Hall, 1988), Table 9.2. Original data from E. Nevo, Genetic variation in natural populations: patterns and theory, *Theoretical Population Biology* 13(1978):121–177.

natural selection is an important force in evolution, but views selection as a process that favors the "best" allele while eliminating others. It proposes that, when selection is important, there will be *little* genetic variation.

The **balance hypothesis** proposes, on the other hand, that the genetic variation in natural populations is maintained by selection that favors variation (balancing selection). Overdominance, in which the heterozygote has higher fitness than that of either homozygote, is one type of balancing selection. Under this hypothesis, the molecular variants are not physiologically equivalent and do not have the same fitness. Instead, genetic variation within natural populations is shaped largely by selection, and, when selection is important, there will be *much* variation.

Many attempts to prove one hypothesis or the other failed, because precisely how much variation was actually present was not clear (remember that protein electrophoresis detects only *some* genetic variation) and because both hypotheses are capable of explaining many different patterns of genetic variation. The controversy over the forces that control variation revealed by protein electrophoresis continues today, but the results of more-recent studies that provide direct information about DNA sequence variation demonstrate that much variation at the level of DNA has little obvious effect on the phenotype and therefore is likely to be neutral.

Concepts

The application of electrophoresis to the study of protein variation in natural populations revealed that most organisms possess large amounts of genetic variation. The neutral-mutation hypothesis proposes that most molecular variation is neutral with regard to natural selection and is shaped largely by mutation and genetic drift. The balance hypothesis proposes that genetic variation is maintained by balancing selection.

www.whfreeman.com/pierce Information on genetic variation in natural populations

DNA Sequence Variation

The development of techniques for isolating, restricting, and sequencing DNA in the 1970s and 1980s provided powerful tools for detecting, quantifying, and investigating genetic variation. The application of these techniques has provided a detailed view of molecular variation.

Restriction enzymes are one tool that can be used to detect genetic variation in DNA and examine patterns of genetic variation in nature. Each restriction enzyme recognizes and cuts a particular sequence of DNA nucleotides, known as

that enzyme's restriction site (see Chapter 18). Variation in the presence of a restriction site is called a restriction fragment length polymorphism (RFLP; see Figure 18.26). Each restriction enzyme recognizes a limited number of nucleotide sites in a particular piece of DNA but, if a number of different restriction enzymes are used and the sites recognized by the enzymes are assumed to be random sequences, RFLPs can be used to estimate the amount of variation in the DNA and the proportion of nucleotides that differ between organisms.

Methods for determining the complete nucleotide sequences of DNA fragments (see p. 000 in Chapter 19) provide the most detailed evolutionary information, although they are both time consuming and expensive. DNA sequencing in evolutionary studies is therefore usually limited to a few individuals or to short sequences. Nevertheless, the high resolution of information provided by sequencing is often invaluable for understanding molecular processes that influence evolution and for determining phylogenies of closely related organisms. For example, DNA sequencing has been used to study the evolution of human immunodeficiency virus (HIV), the virus that causes AIDS. Like many other RNA viruses, HIV evolves rapidly, often changing its sequences within a single host over a period of several years. Evolutionary comparisons of HIV sequences in a dentist and seven of his patients who had AIDS demonstrated that five of the patients contracted AIDS from the dentist, whereas the other two patients probably acquired their HIV infection elsewhere.

Concepts

Restriction fragment length polymorphisms and DNA sequencing can be used to directly examine genetic variation.

Molecular Evolution of HIV in a Florida Dental Practice

In July 1990, the U.S. Center for Disease Control (CDC) reported that a young woman in Florida (later identified as Kimberly Bergalis) had become HIV positive after undergoing an invasive dental procedure performed by a dentist who had AIDS. Bergalis had no known risk factors for HIV infection and no known contact with other HIV-positive persons. The CDC acknowledged that Bergalis might have acquired the infection from her dentist. Subsequently, the dentist wrote to all of his patients, suggesting that they be tested for HIV infection. By 1992, 7 of the dentist's patients had tested positive for HIV, and this number eventually increased to 10.

Originally diagnosed with HIV infection in 1986, the dentist began to develop symptoms of AIDS in 1987 but continued to practice dentistry for another 2 years. All of his HIV-positive patients had received invasive dental procedures, such as root canals and tooth extractions, in the period when the dentist was infected. Among the seven patients originally studied by the CDC (patients A–G, Table 23.9), two had known risk factors for HIV infection (intravenous drug use, homosexual behavior, or sexual relations with HIV-infected persons), and a third had possible but unconfirmed risk factors.

To determine whether the dentist had infected his patients, the CDC conducted a study of the molecular evolution of HIV isolates from the dentist and the patients. HIV undergoes rapid evolution, making it possible to trace the path of its transmission. This rapid evolution also allows HIV to develop drug resistance quickly, making the development of a treatment for AIDS difficult.

Blood specimens were collected from the dentist, the patients, and a group of 35 local controls (other HIV-infected

Table 23.9 HIV-positive persons included in study of HIV isolates from a Florida dental practice

Person	Sex	Known Risk Factors	Average Differences in DNA Sequences (%) From HIV from Dentist	Average Differences in DNA Sequences (%) From HIV from Controls
Dentist	M	Yes		11.0
Patient A	F	No	3.4	10.9
Patient B	F	No	4.4	11.2
Patient C	M	No	3.4	11.1
Patient E	F	No	3.4	10.8
Patient G	M	No	4.9	11.8
Patient D	M	Yes	13.6	13.1
Patient F	M	Yes	10.7	11.9

Source: After C. Ou, et al., *Science* 256(1992):1165–1171, Table 1.

people who lived within 90 miles of the dental practice but who had no known contact with the dentist). DNA was extracted from white blood cells, and a 680-bp fragment of the *envelope* gene of the virus was amplified by PCR (see p. 000 in Chapter 16). The fragments from the dentist, the patients, and the local controls were then sequenced and compared.

The divergence between the viral sequences taken from the dentist, the seven patients, and the controls is shown Table 23.9. Viral DNA taken from patients with no confirmed risk factors (patients A, B, C, E, and G) differed from the dentist's viral DNA by 3.4% to 4.9%, whereas the viral DNA from the controls differed from the dentist's by an average of 11%. The viral sequences collected from five patients (A, B, C, E, and G) were more closely related to the viral sequences collected from the dentist than to viral sequences from the general population, strongly suggesting that these patients acquired their HIV infection from the dentist. The viral isolates from patients D and F (patients with confirmed risk factors), however, differed from that of the dentist by 10.7% and 13.6%, suggesting that these two patients did not acquire their infection from the dentist.

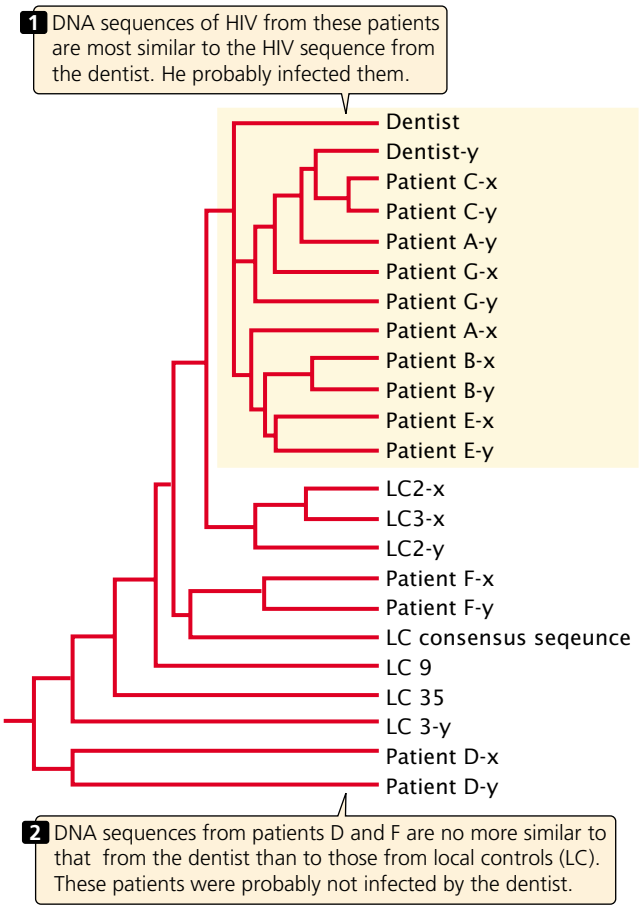
A phylogenetic tree depicting the evolutionary relationships of the viral sequences (FIGURE 23.20) confirmed that the virus taken from the dentist had a close evolutionary relationship to viruses taken from patients A, B, C, E, and G. The viruses from patients D and F, with known risk factors, were no more similar to the virus from the dentist than to viruses from local controls, indicating that the dentist most likely infected five of his patients, whereas the other two patients probably acquired their infections elsewhere. Of three additional HIV-positive patients that have been identified since 1992, only one has viral sequences that are closely related to those from the dentist.

The study of HIV isolates from the dentist and his patients provides an excellent example of the relevance of molecular evolutionary studies to real-world problems. How the dentist infected his patients during their visits to his office remains a mystery, but this case is clearly unusual. A study of almost 16,000 patients treated by HIV-positive health-care workers failed to find a single case of confirmed transmission of HIV from the health-care worker to the patient.

www.whfreeman.com/pierce Web site of the U.S. Center for Disease Control

Patterns of Molecular Variation

The results of molecular studies of numerous genes have demonstrated that different genes and different parts of the same gene often evolve at different rates. Rates of evolutionary change in nucleotide sequences are usually measured as the rate of nucleotide substitution, which is the number of substitutions taking place per nucleotide site per year. To calculate the rate of nucleotide substitution, we begin by looking at homologous sequences from different organisms. We compare the homologous sequences and determine the number of nucleotides that differ between the two sequences. We might compare the growth-hormone sequences for mice and rats, which diverged from a common ancestor some 15 million years ago. From the number of different nucleotides in their growth-hormone genes, we compute the number of nucleotide substitutions that must have taken place since they diverged. Because the same site may have mutated more than once, the number of nucleotide substitutions is larger than the number of nucleotide differences in two sequences; so special mathematical methods have been developed for inferring the actual number of substitutions likely to have taken place.



23.20 Evolutionary tree showing the relationships of HIV isolates from a dentist, seven of his patients (A through G), and other HIV-positive persons from the same region (local controls, LC). The letters x and y represent different isolates from the same patient. The phylogeny is based on DNA sequences taken from the *envelope* gene of the virus. Viral sequences from patients A, B, C, E, and G cluster with those of the dentist, indicating a close evolutionary relationship. Sequences from patients D and F, along with those of local controls, are more distantly related. [C. Ou et al. Molecular epidemiology of HIV transmission in a dental practice, *Science* 256(1992): 1167.]

When we have the number of nucleotide substitutions per nucleotide site, we divide by the amount of evolutionary time that separates the two organisms (usually obtained from the fossil record) to obtain an overall rate of nucleotide substitution. For the mouse and rat growth-hormone gene, the overall rate of nucleotide substitution is approximately 8×10^{-9} substitutions per site per year.

Nucleotide changes in a gene that alter the amino acid sequence of a protein are referred to as nonsynonymous substitutions. Nucleotide changes, particularly those at the third position of the codon, that do not alter the amino acid sequence are called synonymous substitutions. The rate of nonsynonymous substitution varies widely among mammalian genes. The rate for the α-actin protein is only 0.01×10^{-9} substitutions per site per year, whereas the rate for interferon γ is 2.79×10^{-9}, about 1000 times as high. The rate of synonymous substitution also varies among genes, but not to the extent of variation in the nonsynonymous rate. For most protein-encoding genes, the synonymous rate of change is considerably higher than the nonsynonymous rate because synonymous mutations are tolerated by natural selection (Table 23.10). Nonsynonymous mutations, on the other hand, alter the amino acid sequence of the protein and in many cases are detrimental to the fitness of the organism, so most of these mutations are eliminated by natural selection.

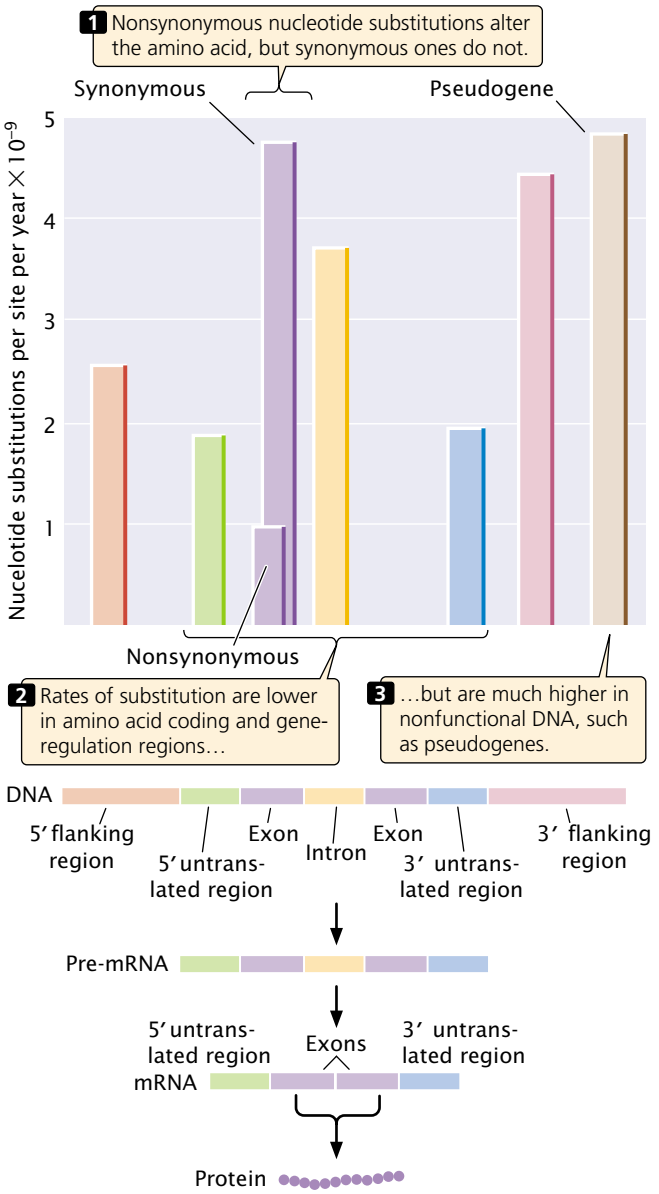
Different parts of a gene also evolve at different rates, with the highest rates of substitutions in regions of the gene that have the least effect on function, such as the third position of a codon, flanking regions, and introns (◖ FIGURE 23.21). The 5′ and 3′ flanking regions of genes are not transcribed into RNA, and therefore substitutions in these regions do not alter the amino acid sequence of the protein, although they may affect gene expression (see Chapter 16). Rates of substitution in introns are nearly as high. Although these nucleotides do not encode amino acids, introns must be spliced out of the pre-mRNA for a functional protein to be produced, and particular sequences are required at the 5′ splice site, 3′ splice site, and branch point for correct splicing (see Chapter 14).

Substitution rates are somewhat lower in the 5′ and 3′ untranslated regions of a gene. These regions are transcribed into RNA but do not encode amino acids. The 5′ untranslated region contains the ribosome-binding site, which is essential for translation, and the 3′ untranslated region contains sequences that may function in regulating mRNA stability and translation; so substitutions in these regions may have deleterious effects on organismal fitness and will not be tolerated.

Table 23.10 Rates of nonsynonymous and synonymous substitutions in mammalian genes based on human–rodent comparisons

Gene	Nonsynonymous Rate (per Site per 10^9 Years)	Synonymous Rate (per Site per 10^9 Years)
α-Actin	0.01	3.68
β-Actin	0.03	3.13
Albumin	0.91	6.63
Aldolase A	0.07	3.59
Apoprotein E	0.98	4.04
Creatine kinase	0.15	3.08
Erythropoietin	0.72	4.34
α-Globin	0.55	5.14
β-Globin	0.80	3.05
Growth hormone	1.23	4.95
Histone 3	0.00	6.38
Immunoglobulin heavy chain (variable region)	1.07	5.66
Insulin	0.13	4.02
Interferon α1	1.41	3.53
Interferon γ	2.79	8.59
Luteinizing hormone	1.02	3.29
Somatostatin-28	0.00	3.97

Source: After W. Li and D. Graur, *Fundamentals of Molecular Evolution* (Sunderland, MA: Sinauer, 1991), p. 69, Table 1.



23.21 Different parts of genes evolve at different rates. The highest rates of nucleotide substitution are in sequences that have the least effect on protein function.

The lowest rates of substitution are seen in nonsynonymous changes in the coding region, because these substitutions always alter the amino acid sequence of the protein and are often deleterious. The highest rates of substitution are in pseudogenes, which are duplicated nonfunctional copies of genes that have acquired mutations. Such a gene no longer produces a functional product; so mutations in pseudogenes have no effect on the fitness of the organism.

In summary, there is a relation between the function of a sequence and its rate of evolution; higher rates are found where they have the least effect on function. This observation fits with the neutral-mutation hypothesis, which predicts that molecular variation is not affected by natural selection.

The Molecular Clock

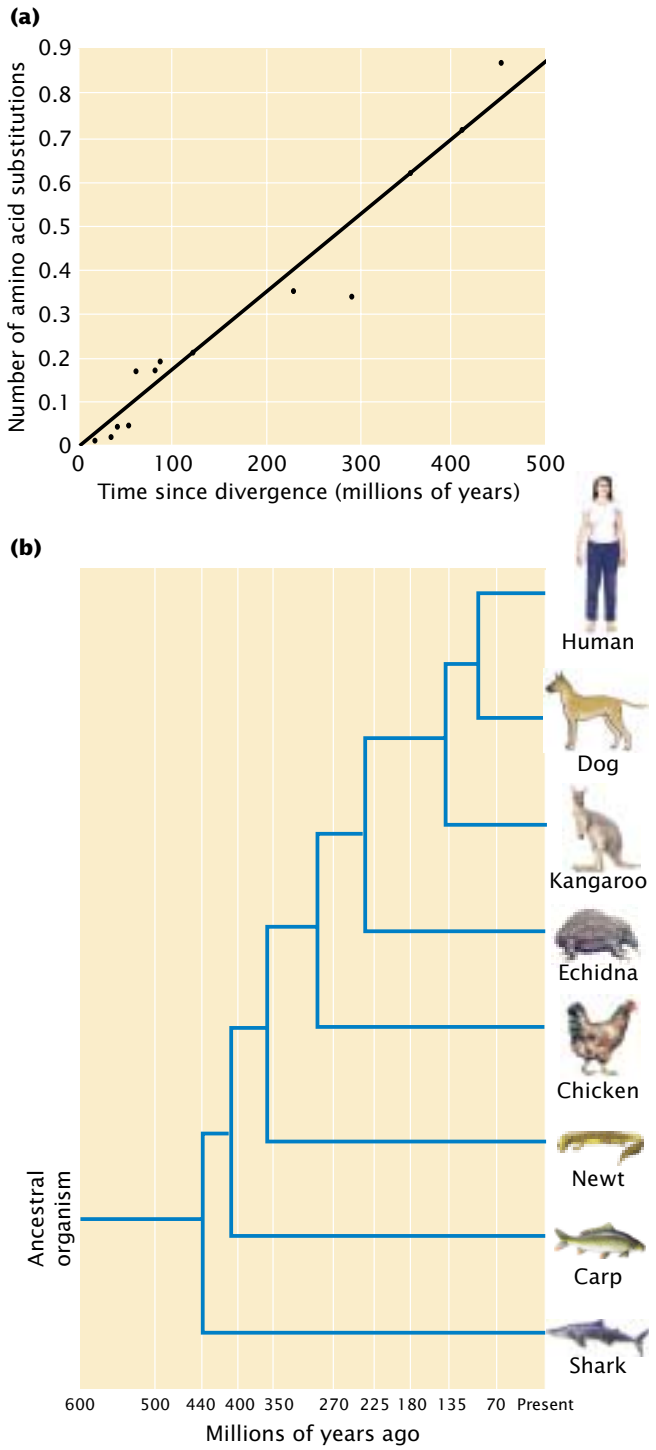
The neutral-mutation theory proposes that evolutionary change at the molecular level occurs primarily through the fixation of neutral mutations by genetic drift. The rate at which one neutral mutation replaces another depends only on the mutation rate, which should be fairly constant for any particular gene. If the rate at which a protein evolves is roughly constant over time, the amount of molecular change that a protein has undergone can be used as a **molecular clock** to date evolutionary events.

For example, the enzyme cytochrome *c* could be examined in two organisms known from fossil evidence to have had a common ancestor 400 million years ago. By determining the number of differences in the cytochrome *c* amino acid sequences in each organism, we could calculate the number of substitutions that have occurred per amino acid site. The occurrence of 20 amino acid substitutions since the two organisms diverged indicates an average rate of 5 substitutions per 100 million years. Knowing how fast the molecular clock ticks allows us to use molecular changes in cytochrome *c* to date other evolutionary events: if we found that cytochrome *c* in two organisms differed by 15 amino acid substitutions, our molecular clock would suggest that they diverged some 300 million years ago. If we assumed some error in our estimate of the rate of amino acid substitution, statistical analysis would show that the true divergence time might range from 160 million to 440 million years. The molecular clock is analogous to geological dating based on the radioactive decay of elements.

The molecular clock was proposed by Emile Zuckerandl and Linus Pauling in 1965 as a possible means of dating evolutionary events on the basis of molecules in present-day organisms. A number of studies have examined the rate of evolutionary change in proteins (FIGURE 23.22), and the molecular clock has been widely used to date evolutionary events when the fossil record is absent or ambiguous. However, the results of several studies have shown that the molecular clock does not always tick at a constant rate, particularly over shorter time periods, and this method remains controversial.

Concepts

Different genes and different parts of the same gene evolve at different rates. Those parts of genes that have the least effect on function tend to evolve at the highest rates. The idea of the molecular clock is that individual proteins and genes evolve at a constant rate and that the differences in the sequences of present-day organisms can be used to date past evolutionary events.

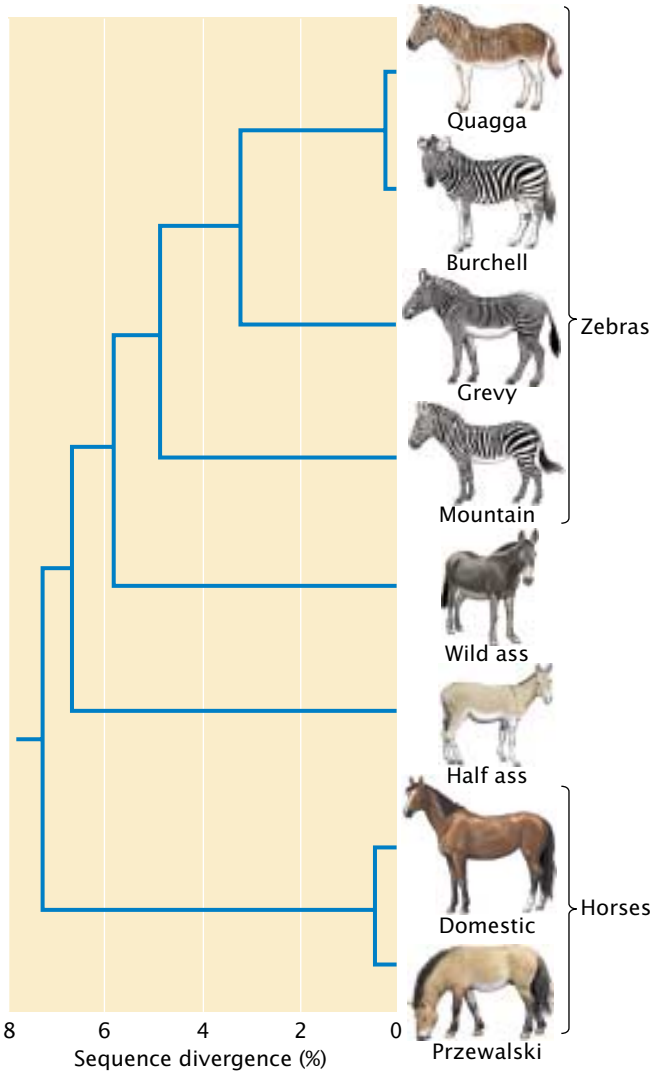


23.22 The molecular clock is based on the assumption of a constant rate of change in protein or DNA sequence. (a) Relation between the rate of amino acid substitution and time since divergence, based on amino acid sequences of α hemoglobin from the eight species shown in part *b*. The constant rate of evolution in protein and DNA sequences has been used as a molecular clock to date past evolutionary events. (b) Phylogeny of eight species and their approximate times of divergence, based on the fossil record.

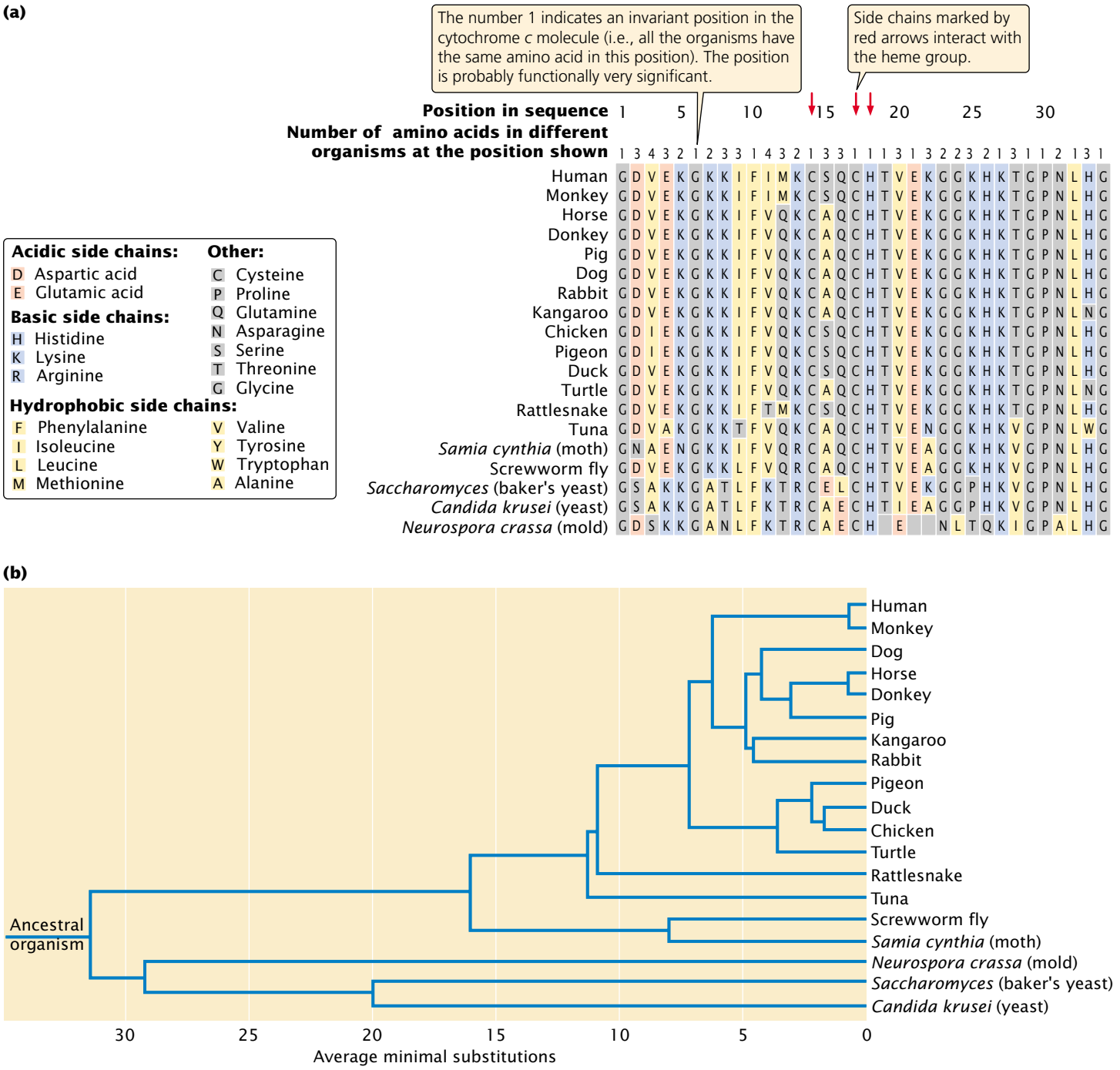
Molecular Phylogenies

As already mentioned, a phylogeny is an evolutionary history of a group of organisms, usually represented as a tree (FIGURE 23.23). The branches of the phylogenetic tree represent the ancestral relationships between the organisms, and the length of each branch is proportional to the amount of evolutionary change that separates the members of the phylogeny.

Before the rise of molecular biology, phylogenies were based largely on anatomical, morphological, or behavioral traits. Evolutionary biologists attempted to gauge the relationships among organisms by assessing the overall degree of similarity or by tracing the appearance of key characteristics of these traits. The first phylogenies constructed from molecular

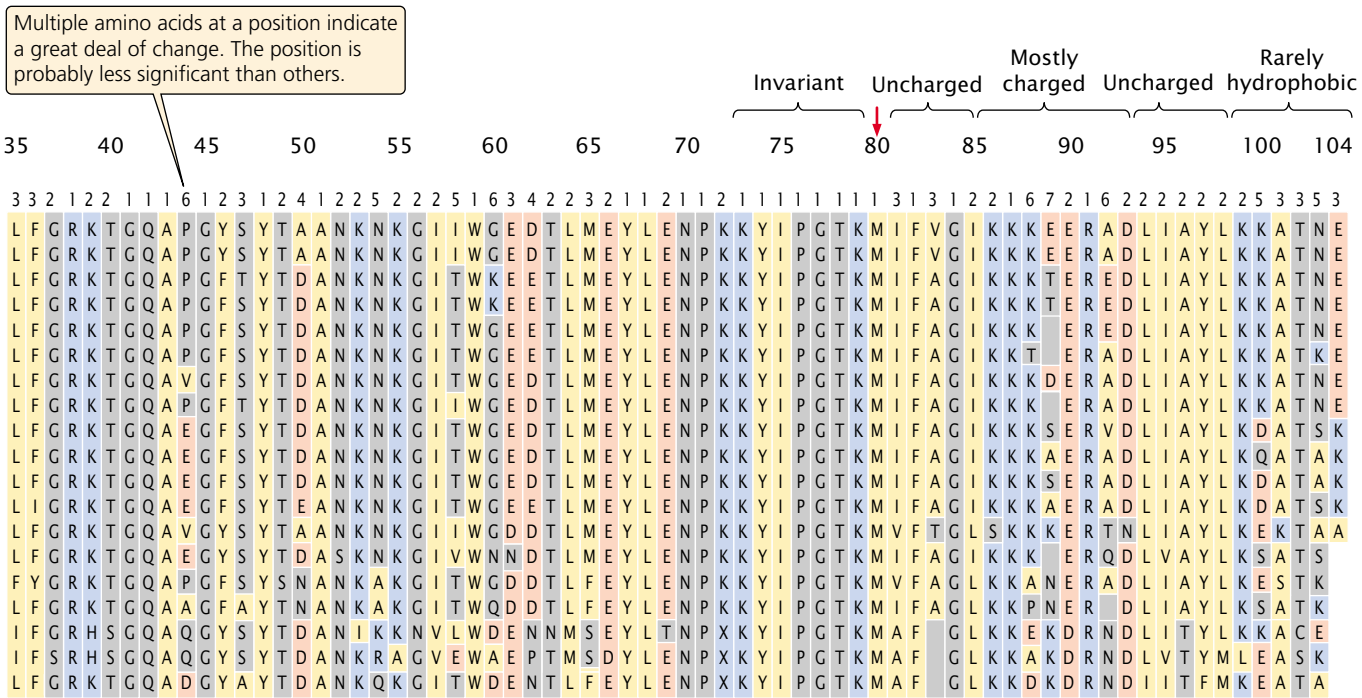


23.23 A phylogeny is the evolutionary history—the ancestral relationships—of a group of organisms. This branching diagram shows the phylogeny of horses based on mitochondrial DNA sequences. DNA of the extinct quagga was extracted from skins from preserved museum specimens.



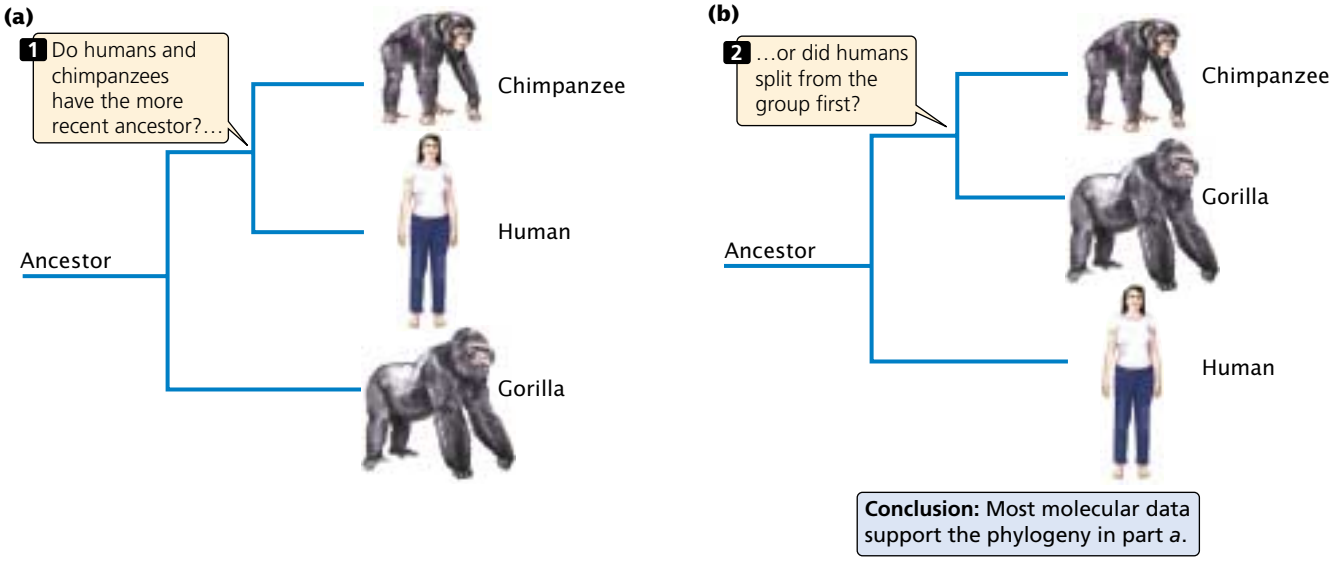
23.24 A phylogeny based on amino acid sequences of the cytochrome *c* molecule.

data were based on amino acid sequences of proteins such as cytochrome *c* (Figure 23.24), but, more recently, phylogenies have been based on DNA sequences. One example is the use of DNA sequences to study the relationship of humans to the other apes. Charles Darwin originally proposed that chimpanzees and gorillas were closely related to humans. However, subsequent study has placed humans in the family Hominidae and the great apes (chimpanzees, gorilla, orangutan, and gibbon) in the family Pongidae. Some researchers suggested that gibbons belong to a third family; others proposed that humans are most closely related to orangutans. Molecular data support the hypothesis that humans, chimpanzees, and

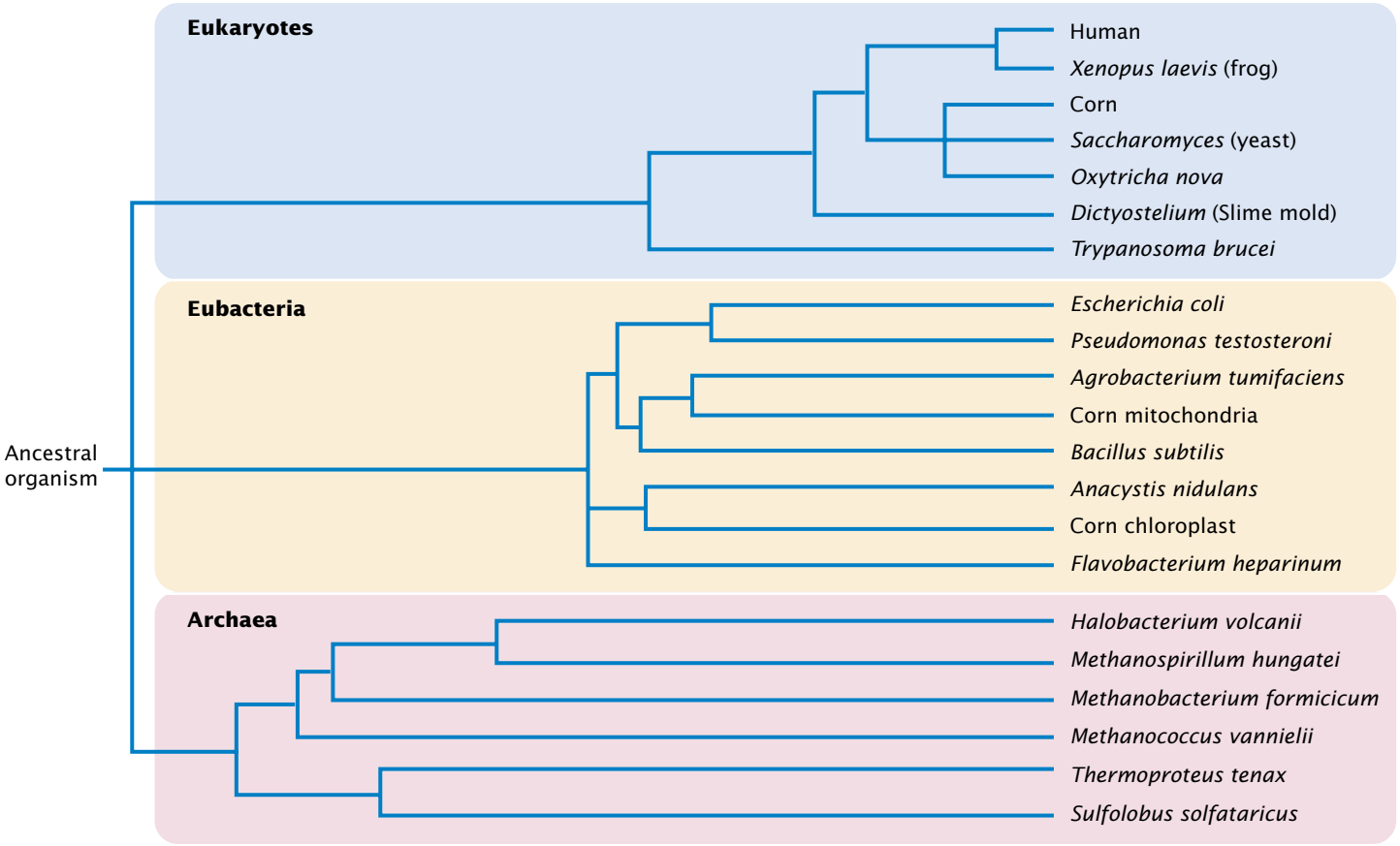


gorillas are most closely related and that orangutans and gibbons diverged from the other apes at a much earlier date. Growing evidence favors a close relationship between humans and chimpanzees (◖ FIGURE 23.25).

Because molecular data can be collected from virtually any organism, comparisons can be made between evolutionary distant organisms. For example, DNA sequences have been used to examine the primary divisions of life and to construct universal phylogenies. On the basis of 16S rRNA, Norman Pace and his colleagues constructed a universal tree of life that included all the major groups of organisms (◖ FIGURE 23.26). The results of their studies



◖ 23.25 **Two possible phylogenies of the human, chimpanzee, and gorilla relationships.** One phylogeny suggests (a) that humans and chimpanzees have the more recent ancestor. The other phylogeny suggests (b) that humans split from the group first and that chimpanzees and gorillas have the more recent ancestor.



23.26 A universal tree of life can be constructed from 16S rRNA sequences. Note that sequences from corn mitochondria and chloroplasts are most similar to sequences from eubacteria, confirming the endosymbiotic hypothesis that these eukaryotic organelles evolved from bacteria (see Chapter 20).

revealed that there are three divisions of life: the eubacteria (the common bacteria), the archaea (a distinct group of lesser-known prokaryotes), and the eukaryotes.

Concepts

Molecular data can be used to infer phylogenies (evolutionary histories) of groups of living organisms.

www.whfreeman.com/pierce Current research in molecular evolution

Connecting Concepts Across Chapters

The central theme of this chapter has been genetic evolution—how the genetic composition of a population changes with time. Unlike transmission and molecular genetics, which focus on individuals and particular genes, this chapter has focused on the genetic makeup of *groups* of individuals. To describe the genes in these groups, we must rely on mathematics and statistical tools; population genetics is therefore fundamentally quantitative in nature. Mathematical models are commonly used in population genetics to describe processes that bring about change in genotypic and allelic frequencies. These models are, by necessity, simplified representations of the real world, but they nevertheless can be sources of insight into how various factors influence the processes of genetic change.

Our study of population genetics depends on and synthesizes much of the information that we have covered in other parts of this book. Describing the genetic composition of a population requires an understanding of the principles of heredity (Chapters 3 through 5) and how genes are changed by mutation (Chapter 17). Our examination of molecular evolution in the second half of the chapter presupposes an understanding of how genes are encoded in DNA, replicated, and expressed (Chapters 10 through 15). It includes the use of molecular tools, such as restriction enzymes, DNA sequencing, and PCR, which are covered in Chapter 18.

CONCEPTS SUMMARY

- Population genetics examines the genetic composition of groups of individuals and how this composition changes with time.
- A Mendelian population is a group of interbreeding, sexually reproducing individuals, whose set of genes constitutes the population's gene pool. Evolution occurs through changes in this gene pool.
- Genetic variation and the forces that shape it are important in population genetics. A population's genetic composition can be described by its genotypic and allelic frequencies.
- The Hardy-Weinberg law describes the effect of reproduction and Mendel's laws on the allelic and genotypic frequencies of a population. It assumes that a population is large, randomly mating, and free from the effects of mutation, migration, and natural selection. When these conditions are met, the allelic frequencies do not change and the genotypic frequencies stabilize after one generation in the Hardy-Weinberg equilibrium proportions p^2, $2pq$, and q^3, where p and q equal the frequencies of the alleles.
- Nonrandom mating affects the frequencies of genotypes but not alleles. Positive assortative mating is preferential mating between like individuals; negative assortative mating is preferential mating between unlike individuals.
- Inbreeding, a type of positive assortative mating, increases the frequency of homozygotes while decreasing the frequency of heterozygotes. Inbreeding is frequently detrimental because it increases the appearance of lethal and deleterious recessive traits.
- Mutation, migration, genetic drift, and natural selection can change allelic frequencies.
- Recurrent mutation eventually leads to an equilibrium, with the allelic frequencies being determined by the relative rates of forward and reverse mutation. Change due to mutation in a single generation is usually very small because mutation rates are low.
- Migration, the movement of genes between populations, increases the amount of genetic variation within populations and decreases differences between populations. The magnitude of change depends both on the differences in allelic frequencies between the populations and on the magnitude of migration.
- Genetic drift, the change in allelic frequencies due to chance factors, is important when the effective population size is small. Genetic drift occurs when a population consists of a small number of individuals, is established by a small number of founders, or undergoes a major reduction in size. Genetic drift changes allelic frequencies, reduces genetic variation within populations, and causes genetic divergence among populations.
- Natural selection is the differential reproduction of genotypes; it is measured by the relative reproductive successes of genotypes (fitnesses). The effects of natural selection on allelic frequency can be determined by applying the general selection model. Directional selection leads to the fixation of one allele. The rate of change in allelic frequency due to selection depends on the intensity of selection, the dominance relations, and the initial frequencies of the alleles.
- Mutation and natural selection can produce an equilibrium, in which the number of new alleles introduced by mutation is balanced by the elimination of alleles through natural selection.
- Molecular methods offer a number of advantages for the study of evolution. The use of protein electrophoresis to study genetic variation in natural populations showed that most natural populations have large amounts of genetic variation in their proteins. Two hypotheses arose to explain this variation. The neutral-mutation hypothesis proposed that molecular variation is selectively neutral and is shaped largely by mutation and genetic drift. The balance model proposed that molecular variation is maintained largely by balancing selection.
- Different parts of the genome show different amounts of genetic variation. In general, those that have the least effect on function evolve at the highest rates.
- The molecular-clock hypothesis proposes a constant rate of nucleotide substitution, providing a means of dating evolutionary events by looking at nucleotide differences between organisms.
- Molecular data are often used for constructing phylogenies.

IMPORTANT TERMS

Mendelian population (p. 000)
gene pool (p. 000)
genotypic frequency (p. 000)
allelic frequency (p. 000)
Hardy-Weinberg law (p. 000)
Hardy-Weinberg equilibrium (p. 000)
positive assortative mating (p. 000)
negative assortative mating (p. 000)
inbreeding (p. 000)
outcrossing (p. 000)
inbreeding coefficient (p. 000)
inbreeding depression (p. 000)
equilibrium (p. 000)
migration (gene flow) (p. 000)
sampling error (p. 000)
genetic drift (p. 000)
effective population size (p. 000)
founder effect (p. 000)
genetic bottleneck (p. 000)
fixation (p. 000)
fitness (p. 000)
selection coefficient (p. 000)
directional selection (p. 000)
overdominance (p. 000)
underdominance (p. 000)
phylogeny (p. 000)
proportion of polymorphic loci (p. 000)
expected heterozygosity (p. 000)
neutral-mutation hypothesis (p. 000)
balance hypothesis (p. 000)
molecular clock (p. 000)

Worked Problems

1. The following genotypes were observed in a population:

Genotype	Number
HH	40
Hh	45
hh	50

(a) Calculate the observed genotypic and allelic frequencies for this population.

(b) Calculate the numbers of genotypes expected if this population were in Hardy-Weinberg equilibrium.

(c) Using a chi-square test, determine whether the population is in Hardy-Weinberg equilibrium.

• Solution

(a) The observed genotypic and allelic frequencies are calculated by using Equations 23.1 and 23.3:

$$f(HH) = \frac{\text{number of } HH \text{ individuals}}{N} = \frac{40}{135} = .30$$

$$f(Hh) = \frac{\text{number of } Hh \text{ individuals}}{N} = \frac{45}{135} = .33$$

$$f(hh) = \frac{\text{number of } hh \text{ individuals}}{N} = \frac{50}{135} = .37$$

$$p = f(H) = \frac{2n_{HH} + n_{Hh}}{2N} = \frac{2(40) + (45)}{2(135)} = .46$$

$$q = f(h) = (1 - p) = (1 - .46) = .54$$

(b) If the population is in Hardy-Weinberg equilibrium, the expected numbers of genotypes are:

$$HH = p^2 \times N = (.46)^2 \times 135 = 28.57$$

$$Hh = 2pq \times N = 2(.46)(.54) \times 135 = 67.07$$

$$hh = q^2 \times N = (.54)^2 \times 135 = 39.37$$

(c) The observed and expected numbers of the genotypes are:

Genotype	Number observed	Number expected
HH	40	28.57
Hh	45	67.07
hh	50	39.37

These numbers can be compared by using a chi-square test:

$$\chi^2 = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$$
$$= \sum \frac{(40 - 28.57)^2}{28.57} + \frac{(45 - 67.07)^2}{67.07} + \frac{(50 - 39.37)^2}{39.37}$$
$$= 4.57 + 7.26 + 2.87 = 14.70$$

The degrees of freedom associated with this chi-square value are $n - 2$, where n equals the number of expected genotypes, or $3 - 2 = 1$. By examining Table 3.4, we see that the probability associated with this chi-square and the degrees of freedom is $P < .001$, which means that the difference between the observed and expected values is unlikely to be due to chance. Thus, there is a significant difference between the observed numbers of genotypes and the numbers that we would expect if the population were in Hardy-Weinberg equilibrium. We conclude that the population is not in equilibrium.

2. A recessive allele for red hair (r) has a frequency of .2 in population I and a frequency of .01 in population II. A famine in population I causes a number of people in population I to migrate to population II, where they reproduce randomly with the members of population II. Geneticists estimate that, after migration, 15% of the people in population II consist of people who migrated from population I. What will be the frequency of red hair in population II after the migration?

• Solution

From Equation 23.16, the allelic frequency in a population after migration (q'_{II}) is

$$q'_{II} = q_I(m) + q_{II}(1 - m)$$

where q_I and q_{II} are the allelic frequencies in population I (migrants) and population II (residents), respectively, and m is the proportion of population II that consist of migrants. In this problem, the frequency of red hair is .2 in population I and .01 in population II. Because 15% of population II consists of migrants, $m = .15$. Substituting these values into Equation 23.16, we obtain:

$$q'_{II} = .2(.15) + (.01)(1 - .15) = .03 + .0085 = .0385$$

This is the expected frequency of the allele for red hair in population II after migration. Red hair is a recessive trait; if mating is random for hair color, the frequency of red hair in population II after migration will be:

$$f(rr) = q^2 = (.0385)^2 = .0015$$

3. Two populations have the following numbers of breeding adults:

Population A: 60 males, 40 females

Population B: 5 males, 95 females

(**a**) Calculate the effective population sizes for populations A and B.

(**b**) What predications can you make about the effects of the different sex ratios of these populations on their gene pools?

• Solution

(**a**) The effective population size can be calculated by using Equation 23.19:

$$N_e = \frac{4 \times n_{\text{males}} \times n_{\text{females}}}{n_{\text{males}} + n_{\text{females}}}$$

For population A:

$$N_e = \frac{4 \times 60 \times 40}{60 + 40} = 96$$

For population B:

$$N_e = \frac{4 \times 5 \times 95}{5 + 95} = 19$$

Although each population has a total of 100 breeding adults, the effective population size of population B is much smaller because it has a greater disparity between the numbers of males and females.

(**b**) The effective population size determines the amount of genetic drift that will occur. Because the effective population size of B is much smaller than that of population A, we can predict that population B will undergo more genetic drift, leading to greater changes in allelic frequency, greater loss of genetic variation, and greater genetic divergence from other populations.

4. Alcohol is a common substance in rotting fruit, where fruit fly larvae grow and develop; larvae use the enzyme alcohol dehydrogenase (ADH) to detoxify the effects of this alcohol. In some fruit-fly populations, two alleles are present at the locus than encodes ADH: ADH^F, which encodes a form of the enzyme that migrates rapidly (fast) on an electrophoretic gel; and ADH^S, which encodes a form of the enzyme that migrates slowly on an electrophoretic gel. Female fruit flies with different *ADH* genotypes produce the following numbers of offspring when alcohol is present:

Genotype	Mean number of offspring
ADH^FADH^F	120
ADH^FADH^S	60
ADH^SADH^S	30

(**a**) Calculate the relative fitnesses of females having these genotypes.

(**b**) If a population of fruit flies has an initial frequency of ADH^F equal to .2, what will be the frequency in the next generation when alcohol is present?

• Solution

(**a**) Fitness is the relative reproductive output of a genotype and is calculated by dividing the average number of offspring produced by that genotype by the mean number of offspring produced by the most prolific genotype. The fitnesses of the three *ADH* genotypes therefore are:

Genotype	Mean number of offspring	Fitness
ADH^FADH^F	120	$W_{FF} = \frac{120}{120} = 1$
ADH^FADH^S	60	$W_{FS} = \frac{60}{120} = .5$
ADH^SADH^S	30	$W_{SS} = \frac{30}{120} = .25$

(**b**) To calculate the frequency of the ADH^F allele after selection, we can use the table method. The frequencies of the three genotypes before selection are the Hardy-Weinberg equilibrium frequencies of p^2, $2pq$, and q^2. We multiply each of these frequencies by the fitness of each genotype to obtain the frequencies after selection. These products are summed to obtain the mean fitness of the population ($\overline{W}$), and the products are then divided by the mean fitness to obtain the relative genotypic frequencies after selection as shown here:

	ADH^FADH^F	ADH^FADH^S	ADH^SADH^S
Initial genotypic frequencies:	$p^2 = (.2)^2 = .04$	$2pq = 2(.2)(.8) = .32$	$q^2 = (.8)^2 = 0.64$
Fitnesses:	$W_{FF} = 1$	$W_{FS} = .5$	$W_{22} = .25$
Proportionate contribution of genotypes to population:	$p^2W_{FF} = .04(1) = .04$	$2pqW_{FS} = (.32)(.5) = .16$	$q^2W_{SS} = (.64)(.25) = .16$
Relative genotypic frequency after selection:	$\frac{p^2W_{FF}}{\overline{W}} = \frac{.04}{.36} = .11$	$\frac{2pqW_{FS}}{\overline{W}} = \frac{.16}{.36} = .44$	$\frac{q^2W_{SS}}{\overline{W}} = \frac{.16}{.36} = .44$

$\overline{W} = .04 + .16 + .16 = .36$

To calculate the allelic frequency after selection, we use Equation 23.4:

$$p = f(ADH^F) = f(ADH^FADH^F) + \tfrac{1}{2} f(ADH^FADH^S)$$
$$= .11 + \tfrac{1}{2}(.44) = .33$$

We predict that the frequency of ADH^F will increase from .2 to .33.

The New Genetics
MINING GENOMES

Population Genetics: Analyses and Simulations

In this exercise, you will analyze real molecular data, primarily generated by high-school and college students, to learn how allele frequencies and genotype distributions can be used to study human populations. To do so, you will use the databases and statistical tools at the Dolan DNA Learning Center of Cold Spring Harbor Laboratory. In addition, you will use simulations to explore how factors such as population size, selection pressure, and genetic drift interact to cause allele frequencies to change.

COMPREHENSION QUESTIONS

1. What is a Mendelian population? How is the gene pool of a Mendelian population usually described? What are the predictions given by the Hardy-Weinberg law?
* 2. What assumptions must be met for a population to be in Hardy-Weinberg equilibrium?
3. What is random mating?
* 4. Give the Hardy-Weinberg expected genotypic frequencies for (**a**) an autosomal locus with three alleles, and (**b**) an X-linked locus with two alleles.
5. Define inbreeding and briefly describe its effects on a population.
6. What determines the allelic frequencies at mutational equilibrium?
* 7. What factors affect the magnitude of change in allelic frequencies due to migration?
8. Define genetic drift and give three ways that it can arise. What effect does genetic drift have on a population?
* 9. What is effective population size? How does it affect the amount of genetic drift?
10. Define natural selection and fitness.
11. Briefly discuss the differences between directional selection, overdominance, and underdominance. Describe the effect of each type of selection on the allelic frequencies of a population.
12. What factors affect the rate of change in allelic frequency due to natural selection?
*13. Compare and contrast the effects of mutation, migration, genetic drift, and natural selection on genetic variation within populations and on genetic divergence between populations.
14. Give some of the advantages of using molecular data in evolutionary studies.
*15. What is the key difference between the neutral-mutation hypothesis and the balance hypothesis?
16. Outline the different rates of evolution that are typically seen in different parts of a protein-encoding gene. What might account for these differences?
*17. What is the molecular clock?

APPLICATION QUESTIONS AND PROBLEMS

18. How would you respond to someone who said that models are useless in studying population genetics because they represent oversimplifications of the real world?
*19. Voles *(Microtus ochrogaster)* were trapped in old fields in southern Indiana and were genotyped for a transferrin locus. The following numbers of genotypes were recorded.

T^ET^E	T^ET^F	T^FT^F
407	170	17

Calculate the genotypic and allelic frequencies of the transferrin locus for this population.

20. Orange coat color in cats is due to an X-linked allele (X^O) that is codominant to the allele for black (X^+). Genotypes of the orange locus of cats in Minneapolis and St. Paul, Minnesota, were determined and the following data were obtained.

X^OX^O females	11
X^OX^+ females	70
X^+X^+ females	94
X^OY males	36
X^+Y males	112

Calculate the frequencies of the X^O and X^+ alleles for this population.

21. A total of 6129 North American Caucasians were blood typed for the MN locus, which is determined by two codominant alleles, L^M and L^N. The following data were obtained:

Blood type	Number
M	1787
MN	3039
N	1303

Carry out a chi-square test to determine whether this population is in Hardy-Weinberg equilibrium at the MN locus.

22. Genotypes of leopard frogs from a population in central Kansas were determined for a locus that encodes the enzyme malate dehydrogenase. The following numbers of genotypes were observed:

Genotype	Number
M^1M^1	20
M^1M^2	45
M^2M^2	42
M^1M^3	4
M^2M^3	8
M^3M^3	6
Total	125

(**a**) Calculate the genotypic and allelic frequencies for this population.

(**b**) What would be the expected numbers of genotypes if the population were in Hardy-Weinberg equilibrium?

23. Full color (*D*) in domestic cats is dominant over dilute color (*d*). Of 325 cats observed, 194 have full color and 131 have dilute color.

(**a**) If these cats are in Hardy-Weinberg equilibrium for the dilution locus, what is the frequency of the dilute allele?

(**b**) How many of the 194 cats with full color are likely to be heterozygous?

24. Tay-Sachs disease is an autosomal recessive disorder. Among Ashkenazi Jews, the frequency of Tay-Sachs disease is 1 in 3600. If the Ashkenazi population is mating randomly for the Tay-Sachs gene, what proportion of the population consists of heterozygous carriers of the Tay-Sachs allele?

25. In the plant *Lotus corniculatus,* cyanogenic glycoside protects the plants against insect pests and even grazing by cattle. This glycoside is due to a simple dominant allele. A population of *L. corniculatus* consists of 77 plants that possess cyanogenic glycoside and 56 that lack the compound. What is the frequency of the dominant allele that results in the presence of cyanogenic glycoside in this population?

*26. Color blindness in humans is an X-linked recessive trait. Approximately 10% of the men in a particular population are color blind.

(**a**) If mating is random for the color-blind locus, what is the frequency of the color-blind allele in this population?

(**b**) What proportion of the women in this population are expected to be color-blind?

(**c**) What proportion of the women in the population are expected to be heterozygous carriers of the color-blind allele?

*27. The human MN blood type is determined by two codominant alleles, L^M and L^N. The frequency of L^M in Eskimos on a small Arctic island is .80. If the inbreeding coefficient for this population is .05, what are the expected frequencies of the M, MN, and N blood types on the island?

28. Demonstrate mathematically that full sib mating ($F = 1/4$) reduces the heterozygosity by $1/4$ with each generation.

29. The forward mutation rate for piebald spotting in guinea pigs is 8×10^{-5}; the reverse mutation rate is 2×10^{-6}. Assuming that no other evolutionary forces are present, what is the expected frequency of the allele for piebald spotting in a population that is in mutational equilibrium?

*30. In German cockroaches, curved wing (*cv*) is recessive to normal wing (cv^+). Bill, who is raising cockroaches in his dorm room, finds that the frequency of the gene for curved wings in his cockroach population is .6. In the apartment of his friend Joe, the frequency of the gene for curved wings is .2. One day Joe visits Bill in his dorm room, and several cockroaches jump out of Joe's hair and join the population in Bill's room. Bill estimates that 10% of the cockroaches in his dorm room now consists of individual roaches that jumped out of Joe's hair. What will be the new frequency of curved wings among cockroaches in Bill's room?

31. A population of water snakes is found on an island in Lake Erie. Some of the snakes are banded and some are unbanded; banding is caused by an autosomal allele that is recessive to an allele for no bands. The frequency of banded snakes on the island is .4, whereas the frequency of banded snakes on the mainland is .81. One summer, a large number of snakes migrate from the mainland to the island. After this migration, 20% of the island population consists of snakes that came from the mainland.

(**a**) Assuming that both the mainland population and the island population are in Hardy-Weinberg equilibrium for the alleles that affect banding, what is the frequency of the allele for bands on the island and on the mainland before migration?

(**b**) After migration has taken place, what will be the frequency of the banded allele on the island?

*32. Calculate the effective size of a population with the following numbers of reproductive adults:

(a) 20 males and 20 females

(b) 30 males and 10 females

(c) 10 males and 30 females

(d) 2 males and 38 females

33. Pikas are small mammals that live at high elevation in the talus slopes of mountains. Populations located on mountain tops in Colorado and Montana in North America are relatively isolated from one another, because the pikas don't occupy the low-elevation habitats that separate the mountain tops and don't venture far from the talus slopes. Thus, there is little gene flow between populations. Furthermore, each population is small in size and was founded by a small number of pikas.

A group of population geneticists propose to study the amount of genetic variation in a series of pika populations and to compare the allelic frequencies in different populations. On the basis of biology and the distribution of pikas, what do you predict the population geneticists will find concerning the within- and between-population genetic variation?

34. In a large, randomly mating population, the frequency of the allele (*s*) for sickle-cell hemoglobin is .028. The results of studies have shown that people with the following genotypes at the beta-chain locus produce the average numbers of offspring given:

Genotype	Average number of offspring produced
SS	5
Ss	6
ss	0

(a) What will be the frequency of the sickle-cell allele *(s)* in the next generation?

(b) What will be the frequency of the sickle cell allele at equilibrium?

35. Two chromosomal inversions are commonly found in populations of *Drosophila pseudoobscura:* Standard *(ST)* and Arrowhead *(AR)*. When treated with the insecticide DDT, the genotypes for these inversions exhibit overdominance, with the following fitnesses:

Genotype	Fitness
ST/*ST*	.47
ST/*AR*	1
AR/*AR*	.62

What will be the frequency of *ST* and *AR* after equilibrium has been reached?

*36. In a large, randomly mating population, the frequency of an autosomal recessive lethal allele is .20. What will be the frequency of this allele in the next generation?

37. A certain form of congenital glaucoma results from an autosomal recessive allele. Assume that the mutation rate is 10^{-5} and that persons having this condition produce, on the average, only about 80% of the offspring produced by persons who do not have glaucoma.

(a) At equilibrium between mutation and selection, what will be the frequency of the gene for congenital glaucoma?

(b) What will be the frequency of the disease in a randomly mating population that is at equilibrium?

CHALLENGE QUESTIONS

38. The Barton Springs salamander is an endangered species found only in a single spring in the city of Austin, Texas. There is growing concern that a chemical spill on a nearby freeway could pollute the spring and wipe out the species. To provide a source of salamanders to repopulate the spring in the event of such a catastrophe, a proposal has been made to establish a captive breeding population of the salamander in a local zoo. You are asked to provide a plan for the establishment of this captive breeding population, with the goal of maintaining as much of the genetic variation of the species as possible in the captive population. What factors might cause loss of genetic variation in the establishment of the captive population? How could loss of such variation be prevented? Assuming that it is feasible to maintain only a limited number of salamanders in captivity, what procedures should be instituted to ensure the long-term maintenance of as much of the variation as possible?

SUGGESTED READINGS

Avise, J. C. 1994. *Molecular Markers, Natural History, and Evolution.* New York: Chapman and Hall.
An excellent review of how molecular techniques are being used to examine evolutionary questions.

Buri, P. 1956. Gene frequency in small populations of mutant *Drosophila. Evolution* 10:367–402.
Buri's famous experiment demonstrating the effects of genetic drift on allelic frequencies.

Hardy, G. H. 1908. Mendelian proportions in a mixed population. *Science* 28:49–50.
Original paper by G. H. Hardy outlining the Hardy-Weinberg law.

Hartl, D. L., and A. G. Clark. 1997. *Principles of Population Genetics,* 3d ed. Sunderland, MA: Sinauer.
An advanced textbook in population genetics.

MacIntyre, R. J., Ed. 1985. *Molecular Evolutionary Genetics.* New York: Plenum.
Contributors treat various aspects of molecular evolution.

Mettler, L. E., T. G. Gregg, and H. S. Schaffer. 1998. *Population Genetics and Evolution,* 2d ed. Englewood Cliffs. NJ: Prentice Hall.
A short, readable textbook on population genetics.

Nei, M., and S. Kumar. 2000. *Molecular Evolution and Phylogenetics.* Oxford: Oxford University Press.
An advanced textbook on the methods used in the study of molecular evolution.

Ou, C., C. A. Ciesielski, G. Myers, C. I. Bandea, et al. 1992. Molecular epidemiology of HIV transmission in a dental practice. *Science* 256:1165–1171.
Study of molecular evolution of HIV in a Florida dental practice.

Provine, W. B. 2002. *The Origins of Theoretical Population Genetics,* 2nd ed. Chicago: Chicago University Press.
A complete history of the origins of population genetics as a field of study.

Saccheri, I., M. Kuussaari, M. Kankare, P. Vikman, W. Fortelius, and I. Hanski. 1998. Inbreeding and extinction in a butterfly metapopulation. *Nature* 392:491–494.
Discusses the role of inbreeding in population extinction of butterflies.

Vial, C., P. Savolainen, J. E. Maldonado, I. R. Amorim, J. E. Rice, R. L. Honeyutt, K. A. Cranall, J. Lundeberg, and R. K. Wayne. 1997. Multiple and ancient origins of the domestic dog. *Science* 276:1687–1689.
Using the molecular clock and mitochondrial DNA sequences, these geneticists estimate that the dog was domesticated more than 100,000 years ago.