20

POPULATION GENETICS

Processes That Change Allelic Frequencies

STUDY OBJECTIVES

1. To develop ways to analyze population genetics problems 571
2. To analyze the effects of mutation, migration, and population size on the Hardy-Weinberg equilibrium 571
3. To study the ways in which natural selection results in organisms adapted to their environments 577

STUDY OUTLINE

Natural selection works on the variation found in nature, shown here by different banding patterns in tree snails (*Liguus fasciatus*), found mainly in southern Florida. (© J. H. Robinson/Photo Researchers, Inc.)

We continue our discussion of the genetics of the evolutionary process. This chapter is devoted to a discussion of some of the effects of violating, or relaxing, the assumptions of the Hardy-Weinberg equilibrium other than random mating, which we discussed in chapter 19. Here we consider the effects of mutation, migration, small population size, and natural selection on the Hardy-Weinberg equilibrium. These processes usually change allelic frequencies.

MODELS FOR POPULATION GENETICS

The steps we need to take to solve for equilibrium in population genetics models follow the same general pattern regardless of what model we are analyzing. We emphasize that these models were developed to help us understand the genetic changes taking place in a population. The models shed light on nonintuitive processes and help quantify intuitive processes. The steps in the models can be outlined as follows:

1. Set up an algebraic model.
2. Calculate allelic frequency in the next generation, q_{n+1}.
3. Calculate change in allelic frequency between generations, Δq.
4. Calculate the equilibrium condition, $\hat{q}$ (q-hat), at $\Delta q = 0$.
5. Determine, when feasible, if the equilibrium is stable.

MUTATION

Mutational Equilibrium

Mutation affects the Hardy-Weinberg equilibrium by changing one allele to another and thus changing allelic and genotypic frequencies. Consider a simple model in which two alleles, A and a, exist. A mutates to a at a rate of μ (mu), and a mutates back to A at a rate of ν (nu):

$$A \underset{\nu}{\overset{\mu}{\rightleftarrows}} a$$

If p_n is the frequency of A in generation n and q_n is the frequency of a in generation n, then the new frequency of a, q_{n+1}, is the old frequency of a plus the addition of a alleles from forward mutation and the loss of a alleles by back mutation. That is,

$$q_{n+1} = q_n + \mu p_n - \nu q_n \qquad (20.1)$$

in which μp_n is the increment of a alleles added by forward mutation, and νq_n is the loss of a alleles due to back mutation. Equation 20.1 takes into account not only the rate of forward mutation, μ, but also p_n, the frequency of A alleles available to mutate. Similarly, the loss of a to A alleles is the product of both the rate of back mutation, ν, and the frequency of the a allele, q_n. Equation 20.1 completes the second modeling step, derivation of an expression for q_{n+1}, allelic frequency after one generation of mutation pressure. The third step is to derive an expression for the change in allelic frequency between two generations. This change (Δq) is simply the difference between the allelic frequency at generation $n + 1$ and the allelic frequency at generation n. Thus, for the a allele

$$\Delta q = q_{n+1} - q_n = (q_n + \mu p_n - \nu q_n) - q_n \qquad (20.2)$$

which simplifies to

$$\Delta q = \mu p_n - \nu q_n \qquad (20.3)$$

The next step in the model is to calculate the equilibrium condition $\hat{q}$, or the allelic frequency when there is no change in allelic frequency from one generation to the next—that is, when Δq (equation 20.3) is equal to zero:

$$\Delta q = \mu p_n - \nu q_n = 0 \qquad (20.4)$$

Thus,

$$\mu p_n = \nu q_n \qquad (20.5)$$

Then, substituting $(1 - q_n)$ for p_n (since $p = 1 - q$), gives

$$\mu(1 - q_n) = \nu q_n$$

or, by rearranging:

$$\hat{q} = \frac{\mu}{\mu + \nu} \qquad (20.6)$$

And, since $p + q = 1$,

$$\hat{p} = \frac{\nu}{\mu + \nu} \qquad (20.7)$$

We can see from equations 20.6 and 20.7 that an equilibrium of allelic frequencies does exist. Also, the equilibrium value of allele a ($\hat{q}$) is directly proportional to the relative size of μ, the rate of forward mutation toward a. If $\mu = \nu$, the equilibrium frequency of the a allele ($\hat{q}$) will be 0.5. As μ gets larger, the equilibrium value shifts toward higher frequencies of the a allele.

Stability of Mutational Equilibrium

Having demonstrated that allelic frequencies can reach an equilibrium due to mutation, we can ask whether the mutational equilibrium is stable. A stable equilibrium is

one that returns to the original equilibrium point after being perturbed. An unstable equilibrium is one that will not return after being perturbed but, rather, continues to move away from the equilibrium point. As we mentioned in the last chapter, the Hardy-Weinberg equilibrium is a neutral equilibrium: It remains at the allelic frequency it moved to when perturbed.

Stable, unstable, and neutral equilibrium points can be visualized as marbles in the bottom of a concave surface (stable), on the top of a convex surface (unstable), or on a level plane (neutral; fig. 20.1). Although more sophisticated mathematical formulas exist for determining whether an equilibrium is stable, unstable, or neutral, we will use graphical analysis for this purpose.

Figure 20.2 introduces the process of graphical analysis, which provides an understanding of the dynamics of an event or process by representing the event in graphical form. In figure 20.2, we have graphed equation 20.3, the Δq equation of mutational dynamics. The ordinate, or y-axis, is Δq, the change in allelic frequency. The abscissa, or x-axis, is q, or allelic frequency. The diagonal line is the Δq equation, the relationship between Δq and q. Note that Δq can be positive (q is increasing) or negative (q is decreasing), whereas q is always positive (0–1.0). Graphical analysis can provide insights into the dynamics of many processes in population genetics.

The diagonal line in figure 20.2 crosses the $\Delta q = 0$ line at the equilibrium value ($\hat{q}$) of 0.167. This line also shows us the changes in allelic frequency that occur in a population not at the equilibrium point. We will look at two examples of populations under the influence of mutation pressure, but not at equilibrium: one at $q = 0.1$ (below equilibrium) and one at $q = 0.9$ (above equilibrium).

If we substitute $q = 0.1$ into equation 20.3, we get a Δq value of 4×10^{-6}. If we substitute $q = 0.9$ into the equation, we get a Δq value of -4.4×10^{-5}. In other words, when the population is below equilibrium, q increases ($\Delta q = +4 \times 10^{-6}$); if the population is above equilibrium, q decreases ($\Delta q = -4.4 \times 10^{-5}$). We can read these same conclusions directly from the graph in figure 20.2.

We can see that the mutational equilibrium is a stable one. Any population whose allelic frequency is not at the equilibrium value tends to return to that equilibrium value. A shortcoming of this model is that it provides no obvious information revealing the time frame for reaching equilibrium. To derive the equations needed to determine this parameter is beyond our scope. (We could use computer simulation or integrate equation 20.3 with respect to time.) In a large population, any great change in allelic frequency caused by mutation pressure alone takes an extremely long time. Most mutation rates are on

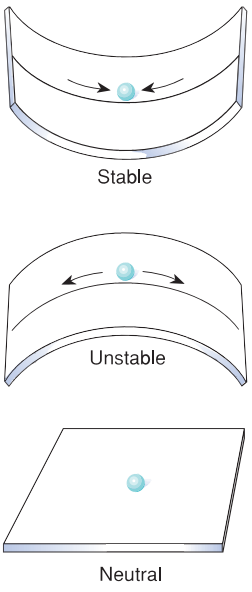

Figure 20.1 Types of equilibria: stable, unstable, and neutral.

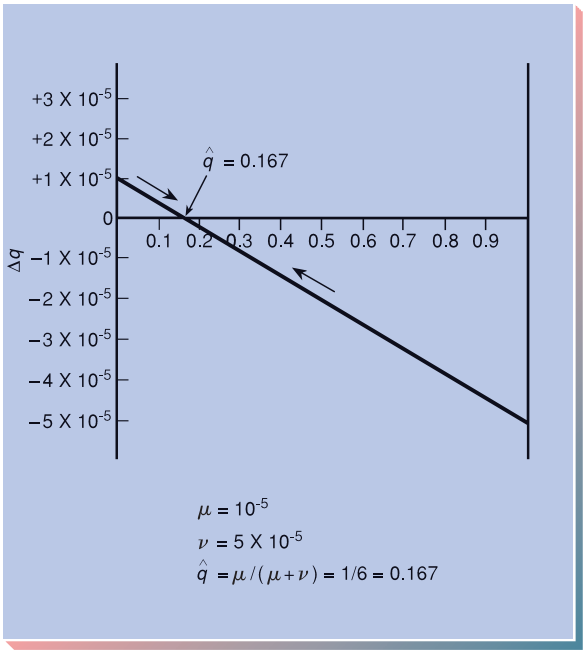

Figure 20.2 Graphical analysis of mutational equilibrium. The graph of the mutational Δq equation shows that when the population is perturbed from the equilibrium point ($q = 0.167$), it returns to that equilibrium point. At q values above equilibrium, change is negative, tending to return the population to equilibrium. At q values below equilibrium, change is positive, also tending to return the population to equilibrium.

the order of 10^{-5}, and equation 20.3 shows that change will be very slow with values of this magnitude. For example, if $\mu = 10^{-5}$, $\nu = 10^{-6}$, and $p = q = 0.5$, $\Delta q = (0.5 \times 10^{-5}) - (0.5 \times 10^{-6}) = 4.5 \times 10^{-6}$, or 0.0000045. It usually takes thousands of generations to get near equilibrium, which is approached asymptotically.

As you can see from the low values of mutation rates, it would usually be nearly impossible to detect perturbations to the Hardy-Weinberg equilibrium by mutation in any one generation. The mutation rate can, however, determine the eventual allelic frequencies at equilibrium if no other factors act to perturb the gradual changes that mutation rates cause. Mutation can also affect final allelic frequencies when it restores alleles that natural selection is removing, a situation we will discuss at the end of the chapter. More important, mutation provides the alternative alleles that natural selection acts upon.

MIGRATION

Migration is similar to mutation in the sense that it adds or removes alleles and thereby changes allelic frequencies. Human populations are frequently affected by migration.

Assume two populations, natives and migrants, both containing alleles *A* and *a* at the *A* locus, but at different frequencies (p_N and q_N versus p_M and q_M), as shown in figure 20.3. Assume that a group of migrants joins the native population and that this group of migrants makes up a fraction m (e.g., 0.2) of the new conglomerate population. Thus, the old residents, or natives, will make up a proportionate fraction ($1 - m$; e.g., 0.8) of the combined population. The conglomerate *a*-allele frequency, q_c, will be the weighted average of the allelic frequencies of the natives and migrants (the allelic frequencies weighted—multiplied—by their proportions):

$$q_c = mq_M + (1 - m)q_N \qquad \textbf{(20.8)}$$

$$q_c = q_N + m(q_M - q_N) \qquad \textbf{(20.9)}$$

The change in allelic frequency, a, from before to after the migration event is

$$\Delta q = q_c - q_N = [q_N + m(q_M - q_N)] - q_N \qquad \textbf{(20.10)}$$

$$\Delta q = m(q_M - q_N) \qquad \textbf{(20.11)}$$

We then find the equilibrium value, $\hat{q}$ (at $\Delta q = 0$). Remembering that, in a product series, any multiplier with the value of zero makes the whole expression zero, Δq will be zero when either

$$m = 0 \text{ or } q_M - q_N = 0;\ q_M = q_N$$

The conclusions we can draw from this model are intuitive. Migration can upset the Hardy-Weinberg equilibrium. Allelic frequencies in a population under the influence of migration will not change if either the size of the migrant group drops to zero (m, the proportion of the conglomerate made up of migrants, drops to zero) or the allelic frequencies in the migrant and resident groups become identical.

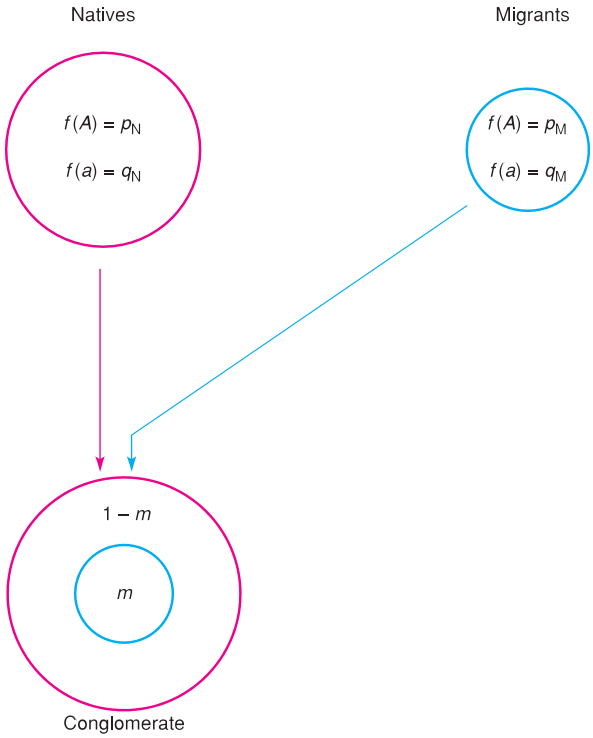

Figure 20.3 Diagrammatic view of migration. A group of migrants enters a native population, making up a proportion, m, of the final conglomerate population.

This migration model can be used to determine the degree to which alleles from one population have entered another population. It can analyze the allele interactions in any two populations. We can, for example, analyze the amount of admixture of alleles from Mongol populations with eastern European populations to explain the relatively high levels of blood type B in eastern European populations (if we make the relatively unrealistic assumption that each of these groups is homogeneous). The calculations are also based on a change happening all in one generation, which did not happen. Blood type and other loci can be used to determine allelic frequencies in western European, eastern European, and Mongol populations. We can rearrange equation 20.9 to solve for m, the proportion of migrants:

$$m = \frac{q_c - q_N}{q_M - q_N} \qquad \textbf{(20.12)}$$

From one sample, we find that the *B* allele is 0.10 in western Europe, taken as the resident or native population (q_N); 0.12 in eastern Europe, the conglomerate population (q_C); and 0.21 in Mongols, the migrants (q_M). Substituting these values into equation 20.12 gives a value for m of 0.18. That is, given the stated assumptions, 18% of the alleles in the eastern European population were brought in by genetic mixture with Mongols.

When a migrant group first joins a native group, before genetic mixing (mating) takes place, the Hardy-Weinberg equilibrium of the conglomerate population is perturbed, even though both subgroups are themselves in Hardy-Weinberg proportions. A decrease will occur in heterozygotes in the conglomerate population as compared to what we would predict from the allelic frequencies of that population (the average allelic frequencies of the two groups). This is a phenomenon of subdivision referred to as the **Wahlund effect.** The reason this happens is because the relative proportions of heterozygotes increase at intermediate allelic frequencies. As allelic frequencies rise above or fall below 0.5, the relative proportion of heterozygotes decreases.

In a conglomerate population, the allelic frequencies will be intermediate between the values of the two subgroups because of averaging. This generally means the predicted proportion of heterozygotes will be higher than the actual average proportion of heterozygotes in the two subgroups. An example is worked out in table 20.1. Assume that the two subgroups each make up 50% of the conglomerate population. In subgroup 1, $p = 0.1$ and $q = 0.9$; in subgroup 2, $p = 0.9$ and $q = 0.1$. Each subgroup will have 18% heterozygotes. The average, $(0.18 + 0.18)/2 = 0.18$, is the proportion of heterozygotes actually in the population. However, the conglomerate allelic frequencies are $p = 0.5$ and $q = 0.5$, leading to the expectation that 50% of the population will be heterozygotes. Hence, the observed frequency of heterozygotes is lower than the expected frequency (i.e., the Wahlund effect).

We should note that the same logic holds even if both populations have allelic frequencies above or below 0.5. Also, this effect happens when an observer samples what he or she thinks is a single population but is actually a population subdivided into several demes. When most population geneticists sample a population and find a deficiency of heterozygotes, they first think of inbreeding and then of subdivision, the Wahlund effect. (A further complication is that inbreeding leads to subdivision, and subdivision leads to inbreeding. Statistics have been developed to try to separate the effects of these two phenomena.) As soon as random mating occurs in a subdivided population, Hardy-Weinberg equilibrium is established in one generation. We refer to a population in which the individuals are mating at random as unstructured or **panmictic.**

Table 20.1 **The Wahlund Effect: Heterozygote Frequencies Are Below Expected in a Conglomerate Population**

	Subgroup I	Subgroup 2	Conglomerate	
p	0.1	0.9	0.5	
q	0.9	0.1	0.5	
			Expected	Observed
p^2	0.01	0.81	0.25	0.41
$2pq$	0.18	0.18	0.50	0.18
q^2	0.81	0.01	0.25	0.41

Note: In this example, the subgroups are of equal sizes.

SMALL POPULATION SIZE

Another variable that can upset the Hardy-Weinberg equilibrium is small population size. The Hardy-Weinberg equilibrium assumes an infinitely large population because, as defined, it is **deterministic,** not stochastic. That is, the Hardy-Weinberg equilibrium predicts exactly what the allelic and genotypic frequencies should be after one generation; it ignores variation due to sampling error. Obviously, every population of organisms on earth violates the Hardy-Weinberg assumption of infinite population size.

Sampling Error

The zygotes of every generation are a sample of gametes from the parent generation. Sampling errors are the changes in allelic frequencies from one generation to the next that are due to inexact sampling of the alleles of the parent generation. Toss a coin one hundred times, and chances are, it will not land heads exactly fifty times. However, as the number of coin tosses increases, the percentage of heads will approach 50%, a percentage reached with certainty only after an infinite number of tosses. The same applies to any sampling problem, from drawing cards from a deck to drawing gametes from a gene pool.

If small population size is the only factor causing deviation from Hardy-Weinberg equilibrium, it will cause the allelic frequencies of a population to fluctuate from generation to generation in the process known as random genetic drift. In other words, an *Aa* heterozygote will sometimes produce several offspring that have only the *A* allele, or sometimes random mortality will kill a disproportionate number of *aa* homozygotes. In either case, the next generation may not have the same allelic frequencies as the

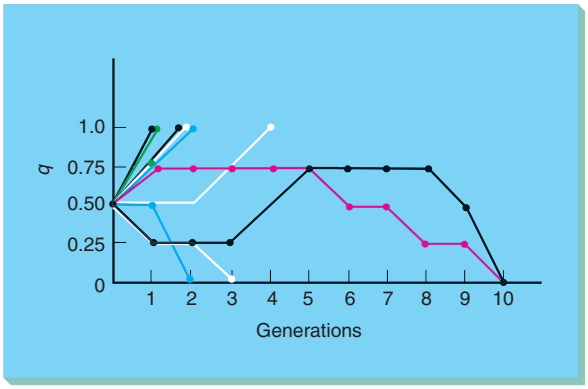

Figure 20.4 Random genetic drift. Ten populations, each consisting of two individuals with initial $q = 0.5$, all go to fixation or loss of the *a* allele (four or zero copies) within ten generations due to the sampling error of gametes. Once the *a* allele has been fixed or lost, no further change in allelic frequency will occur (barring mutation or migration). We show a population of only two individuals to exaggerate the effects of random genetic drift.

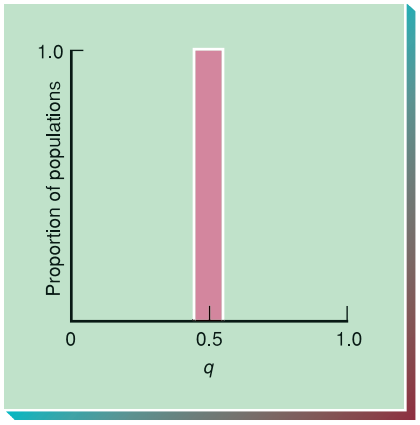

Figure 20.5 Initial conditions of random drift model. One thousand populations, each of size one hundred, and each with an allelic frequency (q) of 0.5.

present generation. The end result will be either fixation or loss of any given allele ($q = 1$ or $q = 0$; fig. 20.4), although which will be fixed or lost depends on the original allelic frequencies. The rate of approach to reach the fixation-loss endpoint depends on the size of the population.

Simulation of Random Genetic Drift

We can investigate the process of random genetic drift mathematically by starting with a large number of populations of the same finite size and observing how the distribution of allelic frequencies among the populations changes in time due only to random genetic drift. For example, we can start with one thousand hypothetical populations, each containing one hundred individuals, with the frequency of the *a* allele, q, 0.5 in each (fig. 20.5). We measure time in generations, t, as a function of the population size, N (one hundred in this example). For instance, $t = N$ is generation one hundred, $t = N/5$ is generation twenty, and $t = 3N$ is generation three hundred. Then, by using computer simulation (or the **Fokker-Planck equation,** which physicists use to describe diffusion processes such as Brownian motion), we generate the series of curves shown in figure 20.6. These curves show that as the number of generations increases, the populations begin to diverge from $q = 0.5$. Approximately the same number of populations go to q values above 0.5 as go to q values below 0.5. Therefore, the distribution spreads symmetrically. When the distribution of allelic frequencies reaches the sides of the graph, some populations become fixed for the *a* allele and some lose it. In a sense, the sides act as sinks:

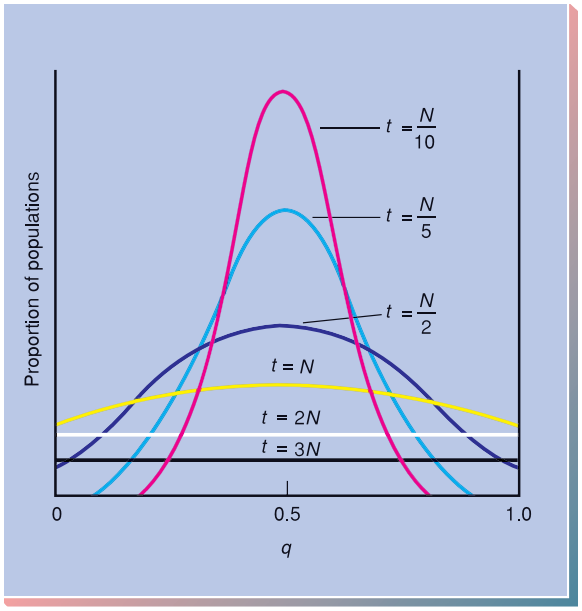

Figure 20.6 Genetic drift in small populations: $q = 0.5$. After time passes, the populations of figure 20.5 begin to diverge in their allelic frequencies. Time is measured in population size (N), showing that the effects of random genetic drift are qualitatively similar in populations of all sizes; the only difference is the timescale. (From M. Kimura, "Solution of a process of random genetic drift with a continuous model," *Proceedings of the National Academy of Sciences, USA,* 41:144-50, 1955. Reprinted by permission.)

Any population that has the *a* allele lost or fixed will be permanently removed from the process of random genetic drift. Without mutation to bring one or the other allele

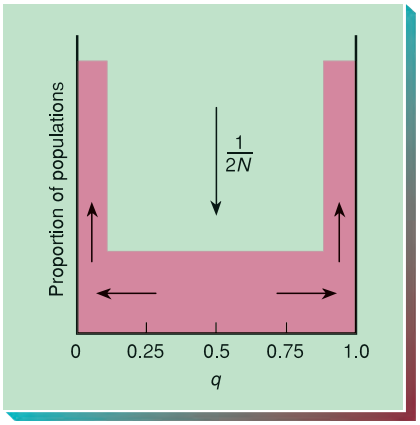

Figure 20.7 Continued genetic drift in the one thousand populations, each numbering one hundred in size, shown in figures 20.5 and 20.6. After approximately 2*N* generations, the distribution is flat, and populations are going to loss or fixation of the *a* allele at a rate of 1/2*N* populations per generation. (From S. Wright, "Evolution in Mendelian Populations," *Genetics,* 97:114. Copyright © 1931 Genetics Society of America.)

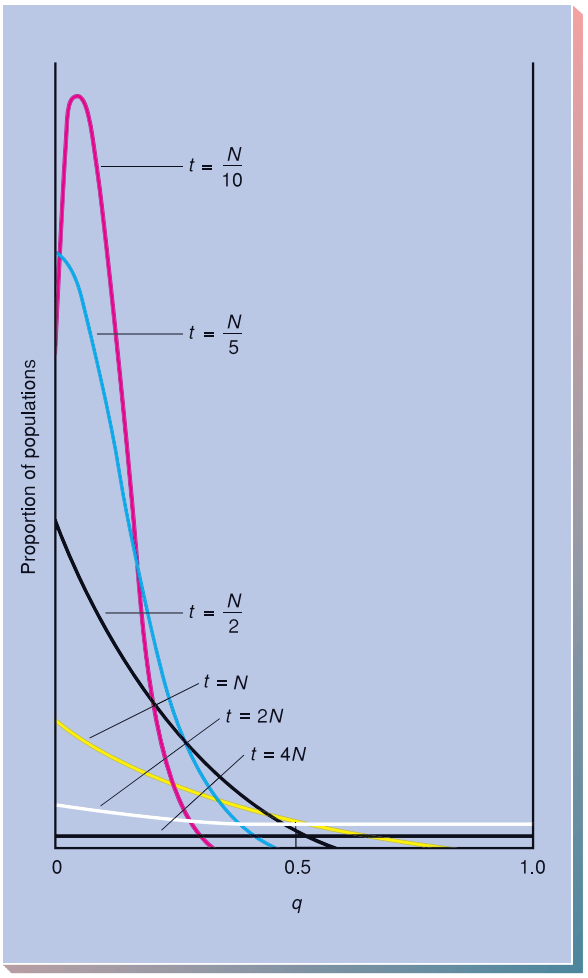

Figure 20.8 Random genetic drift in small populations with $q = 0.1$. Compare this figure with figure 20.6. In this case, the probability of fixation of the *a* allele is 0.1, and the probability of its loss is 0.9. (From M. Kimura, "Solution of a process of random genetic drift with a continuous model," *Proceedings of the National Academy of Sciences, USA,* 41:144–50, 1955. Reprinted by permission.)

back into the gene pool, these populations maintain a constant allelic frequency of zero or 1.0.

At a point between *N* (one hundred) and 2*N* (two hundred) generations, the distribution of allelic frequencies flattens out and begins to lose populations to the edges (fixation or loss) at a constant rate, as shown in figure 20.7. The rate of loss is about 1/2*N* (1/200), or 0.5% of the populations per generation. If the initial allelic frequency was not 0.5, everything is shifted in the distribution (fig. 20.8), but the basic process is the same—in all populations, sampling error causes allelic frequencies to drift toward fixation or elimination. If no other factor counteracts this drift, every population is destined to eventually be either fixed for or deficient in any given allele.

The amount of time the process takes depends on population size. The example used here was based on small populations of one hundred. If we substitute one million for one hundred in figure 20.6, a flat distribution of populations would not be reached for two million generations, rather than two hundred generations. Thus, a population experiences the effect of random genetic drift in inverse proportion to its size: Small populations rapidly fix or lose a given allele, whereas large populations take longer to show the same effects. Genetic drift also shows itself in several other ways.

Founder Effects and Bottlenecks

Several well-known genetic phenomena are caused by populations starting at or proceeding through small numbers. When a population is initiated by a small, and therefore genetically unrepresentative, sample of the parent population, the genetic drift observed in the subpopulation is referred to as a **founder effect.** A classic human example is the population founded on Pitcairn Island by several of the *Bounty* mutineers and some Polynesians. The unique combination of Caucasian and Polynesian traits that characterizes today's Pitcairn Island population resulted from the small number of founders for the population.

Sometimes populations go through **bottlenecks,** periods of very small population size, with predictable ge-

netic results. After the bottleneck, the parents of the next generation have been reduced to a small number and may not be genetically representative of the original population. The field mice on Muskeget Island, Massachusetts, have a white forehead blaze of hair not commonly found in nearby mainland populations. Presumably, the island population went through a bottleneck at the turn of the century, when cats on the island reduced the number of mice to near zero. The population was reestablished by a small group of mice that happened by chance to contain several animals with this forehead blaze.

NATURAL SELECTION

Although mutation, migration, and random genetic drift all influence allelic frequencies, they do not necessarily produce populations of individuals that are better adapted to their environments. Natural selection, however, tends to that end. The consequence of natural selection, Darwinian evolution, is considered in detail in the next chapter. We discuss here the algebra behind the process of natural selection. Artificial selection, as practiced by animal and plant breeders, follows the same rules.

How Natural Selection Acts

Selection, or **natural selection,** is a process whereby one phenotype and, therefore, one genotype leaves relatively more offspring than another genotype, measured by both reproduction and survival. Selection is thus a matter of **reproductive success,** the relative contribution of that genotype to the next generation. It is important to remember that selection acts on whole organisms and thus on phenotypes. However, we analyze the process by looking directly at the genotype, usually only at one locus.

Fitness

A measure of reproductive success is the **fitness,** or **adaptive value,** of a genotype. A genotype that, compared with other genotypes, leaves relatively more offspring that survive to reproduce has the higher fitness. (Note that this use of the word *fitness* differs from our common notion of physical fitness.)

Fitness is usually computed to vary from zero to one (0–1) and is always related to a given population at a given time. For example, in a normal environment, fruit flies with long wings may be more fit than fruit flies with short wings. But in a very windy environment, a fruit fly with limited flying ability may survive better than one with the long-winged genotype, which will be blown around by the wind. Thus, fitness (usually assigned the letter W) is relative to a given circumstance. In a given environment, the genotype that leaves the most offspring is usually assigned a fitness of $W = 1$, and a lethal genotype has a fitness of $W = 0$. Any other genotype has a fitness value between zero and one. A number of factors can decrease this fitness value, W, below one. A **selection coefficient** measures the sum of forces acting to prevent reproductive success. It is usually represented by the letter s or t and is defined by the fitness equation

$$W = 1 - s \qquad (20.13)$$

and

$$s = 1 - W \qquad (20.14)$$

Thus, as the selection coefficient increases, fitness decreases, and vice versa.

Components of Fitness

Natural selection can act at any stage of the life cycle of an organism. It usually acts in one of four ways. (1) The reproductive success of a genotype can be affected by prenatal, juvenile, or adult survival. Differential survival of genotypes is referred to as viability selection or **zygotic selection.** (2) A heterozygote can produce gametes with differential success when one of its alleles fertilizes more often than the other. This is termed **gametic selection.** A well-studied case is the *t*-allele (tailless) locus in house mice; the gametes of as many as 95% of the heterozygous males of the *Tt* genotype carry the *t* allele. (This phenomenon is also referred to as **segregation distortion** or **meiotic drive.**) Selection can also take place in two areas of the reproductive segment of an organism's life cycle. (3) Some genotypes may mate more often than others (have greater mating success), resulting in **sexual selection.** Sexual selection usually occurs when members of the same sex compete for mates or when females have some form of choice. Adaptations for fighting, such as antlers in male elk, or displaying, such as the peacock's tail, are the results of sexual selection. (4) Finally, some genotypes may be more fertile than other genotypes, resulting in **fecundity selection.** The particular variable of the life cycle that selection acts upon is termed a **component of fitness.**

Effects of Selection

Figure 20.9 shows the three main ways that the sum total of selection can act. **Directional selection** works by continuously removing individuals from one end of the phenotypic (and therefore, presumably, genotypic) distribution (e.g., short-necked giraffes are removed). Removal means disappearance through death or failure to reproduce (genetic death). Thus, the mean is constantly shifted toward

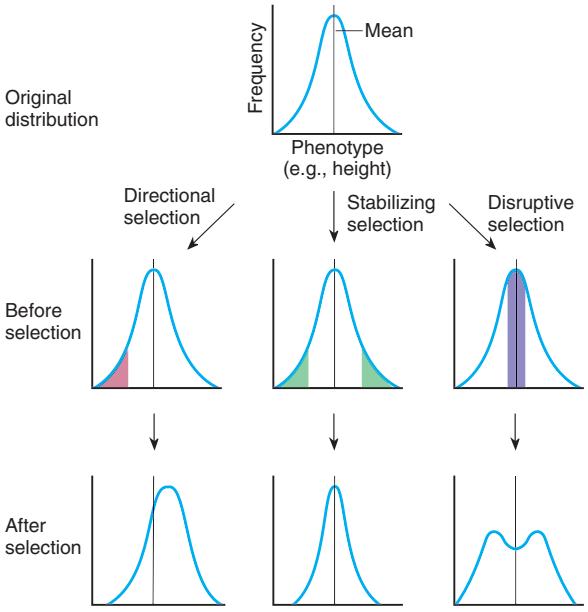

Figure 20.9 Directional, stabilizing, and disruptive selection. *Colored areas* show the groups being selected against. At the *top* is the original distribution of individuals. The final distributions after selection appear in the *bottom row.*

the other end of the phenotypic distribution; in our example, the mean shifts toward long-necked giraffes. The evolution of neck length in giraffes, presumably by directional selection, has been documented from the geologic record.

Stabilizing selection (fig. 20.9) works by constantly removing individuals from both ends of a phenotypic distribution, thus maintaining the same mean over time. Stabilizing selection now works on giraffe neck length—it is neither increasing nor decreasing. **Disruptive selection** works by favoring individuals at both ends of a phenotypic distribution at the expense of individuals in the middle. It, like stabilizing selection, should maintain the same mean value for the phenotypic distribution. Disruptive selection has been carried out successfully in the laboratory for bristle number in *Drosophila.* Starting with a population with a mean number of sternopleural chaeta (bristles on one of the body plates) of about eighteen, investigators succeeded after twelve generations of getting a fly population with one peak of bristle numbers at about sixteen and another at about twenty-three (fig. 20.10).

Selection Against the Recessive Homozygote

We can analyze selection by using our standard model-building protocol of population genetics—namely, define the initial conditions; allow selection to act; calculate the allelic frequency after selection (q_{n+1}); calculate Δq (change in allelic frequency from one generation to the next); then calculate equilibrium frequency, $\hat{q}$, when Δq becomes zero; and examine the stability of the equilibrium. In the analysis that follows, we consider a single autosomal locus in a diploid, sexually reproducing species with two alleles and assume that selection acts directly on the phenotypes in a simple fashion (i.e., it occurs at a single stage in the life of the organism, such as larval mortality in *Drosophila*). After selection, the individuals remaining within the population mate at random to form a new generation in Hardy-Weinberg proportions.

Selection Model

In table 20.2, we outline the model for selection against the homozygous recessive genotype. The initial population is in Hardy-Weinberg equilibrium. Even with selection acting during the life cycle of the organism, Hardy-Weinberg proportions will be reestablished anew after each round of random mating, although presumably at new allelic frequencies. All selection models start out the same way. They diverge at the point of assigning fitness, which depends on the way natural selection is acting. In the model in table 20.2, the dominant homozygote and the heterozygote have

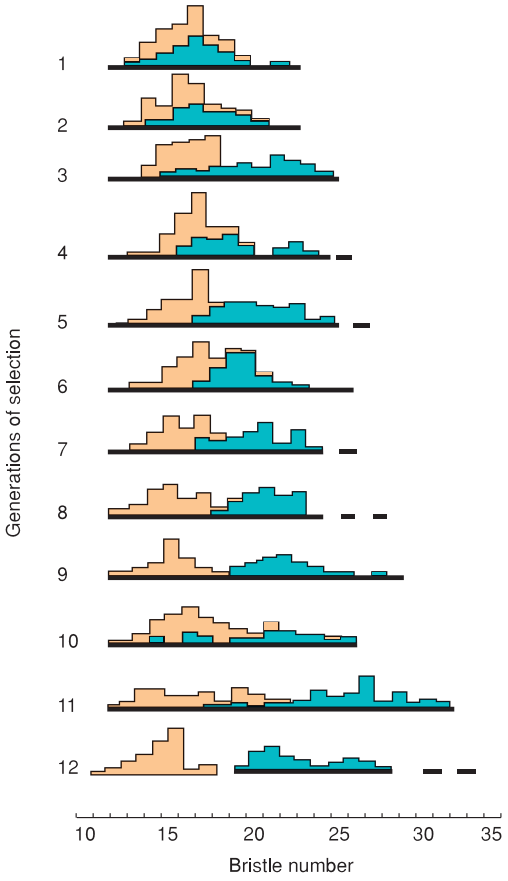

Figure 20.10 Disruptive selection in *Drosophila melanogaster.* After twelve generations of selection for flies with either many or few bristles (chaetae) on the sternopleural plate, the population was bimodal. In other words, many flies in the population had either few or many bristles, but few flies had an intermediate bristle number. (Reprinted with permission from *Nature,* Vol. 193, J. M. Thoday and J. B. Gibson, "Isolation by Disruptive Selection." Copyright © 1962 Macmillan Magazines Limited.)

the same fitness ($W = 1$). Natural selection cannot differentiate between the two genotypes because they both have the same phenotype. The recessive homozygote, however, is being selected against, which means that it has a lower fitness than the two other genotypes ($W = 1 - s$).

After selection, the ratio of the different genotypes is determined by multiplying their frequencies (Hardy-Weinberg proportions) by their fitnesses. The procedure follows from the definition of fitness, which in this case is a relative survival value. Thus, only $1 - s$ of the *aa* genotype survives for every one of the other two genotypes. For example, if s were 0.4, then the fitness of the *aa* type would be $1 - s$, or 0.6. For every ten *AA* and *Aa* individuals that survive to reproduce, only six *aa* individuals would survive to reproduce. The total of the three genotypes after selection is $1 - sq^2$. That is,

$$p^2 + 2pq + q^2(1 - s) = p^2 + 2pq + q^2 - sq^2$$
$$= 1 - sq^2$$

Table 20.2 **Selection Against the Recessive Homozygote: One Locus with Two Alleles, *A* and *a***

	Genotype			
	AA	***Aa***	***aa***	**Total**
Initial genotypic frequencies	p^2	$2pq$	q^2	1
Fitness (W)	1	1	$1-s$	
Ratio after selection	p^2	$2pq$	$q^2(1-s)$	$1-sq^2 = \overline{W}$
Genotypic frequencies after selection	$\frac{p^2}{\overline{W}}$	$\frac{2pq}{\overline{W}}$	$\frac{q^2(1-s)}{\overline{W}}$	1

Mean Fitness of a Population

The value $(1 - sq^2)$ is referred to as the **mean fitness of the population, $\overline{W}$,** because it is the sum of the fitnesses of the genotypes multiplied (weighted) by the frequencies at which they occur. Thus, it is a weighted mean of the fitnesses, weighted by their frequencies. The new ratios of the three genotypes can be returned to genotypic frequencies by simply dividing by the mean fitness of the population, $\overline{W}$, as in the last line of table 20.2. (Remember that a set of numbers can be converted to proportions of unity by dividing them by their sum.) The new genotypic frequencies are thus the products of their original frequencies times their fitnesses, divided by the mean fitness of the population.

After selection, the new allelic frequency (q_{n+1}) is the proportion of *aa* homozygotes plus half the proportion of heterozygotes, or

$$q_{n+1} = \frac{q^2(1 - s)}{1 - sq^2} + \frac{pq}{1 - sq^2}$$
$$= \frac{q(q - sq + p)}{1 - sq^2}$$
$$= \frac{q(1 - sq)}{1 - sq^2} \qquad \textbf{(20.15)}$$

This model can be simplified somewhat if we assume that the *aa* genotype is lethal. Its fitness would be zero,

and s, the selection coefficient, would be one. Equation 20.15 would then change to

$$q_{n+1} = \frac{q(1-q)}{1-q^2} \qquad \textbf{(20.16)}$$

Since $(1 - q^2)$ is factorable into $(1 - q)(1 + q)$, equation 20.16 becomes

$$q_{n+1} = \frac{q(1-q)}{(1-q)(1+q)}$$

$$= \frac{q}{(1+q)} \qquad \textbf{(20.17)}$$

The change in allelic frequency is then calculated as

$$\Delta q = q_{n+1} - q = \frac{q}{1+q} - q$$

To solve this equation, q is multiplied by $(1 + q)/(1 + q)$ so that both parts of the expression are over a common denominator:

$$\Delta q = \frac{q - q(1+q)}{1+q}$$

$$= \frac{-q^2}{1+q} \qquad \textbf{(20.18)}$$

This is the expression for the change in allelic frequency caused by selection. Since selection will not act again until the same stage in the life cycle during the next generation, equation 20.18 is also an expression for the change in allelic frequency between generations.

Two facts should be apparent from equation 20.18. First, the frequency of the recessive allele (q) is declining, as indicated by the negative sign of the fraction. This fact should be intuitive because of the way selection was defined in the model (eliminating *aa* homozygotes). Second, the change in allelic frequency is proportional to q^2, which appears in the numerator of the expression. In other words, allelic frequency is declining as a relative function of the number of homozygous recessive individuals in the population. This fact is consistent with the premise of the selection model (with selection against the homozygous recessive genotype). This final formula supports the methodology of the model.

Equilibrium Conditions

Next we calculate the equilibrium q by setting the Δq equation equal to zero, since a population in equilibrium will show no change in allelic frequencies from one generation to the next:

$$\frac{-q^2}{1+q} = 0 \qquad \textbf{(20.19)}$$

For a fraction to be zero, the numerator must equal zero. Thus, $q^2 = 0$, and $\hat{q} = 0$. At equilibrium, the *a* allele should be entirely removed from the population. If the *aa* homozygotes are being removed, and if there is no mutation to return *a* alleles to the population, then eventually the *a* allele disappears from the population.

Time Frame for Equilibrium

One shortcoming of this selection model is that it is not immediately apparent how many generations will be required to remove the *a* allele. The deficiency can be compensated for by using a computer simulation or by introducing a calculus differential into the model. Either method would produce the frequency-time graph of figure 20.11. This figure clearly shows that the *a* allele is removed more quickly when selection is stronger (when s is larger) and that the curves appear to be asymptotic—the *a* allele is not immediately eliminated and would not be entirely removed until an infinitely large number of generations had passed. There is a reason for the asymptotic behavior of the graph: As the *a* allele becomes rarer and rarer, it tends to be found in heterozygotes (table 20.3). Since selection can remove only *aa* homozygotes, an *a* allele hidden in an *Aa* heterozygote will not be selected against. When $q = 0.5$, there are two heterozygotes for every *aa* homozygote. When

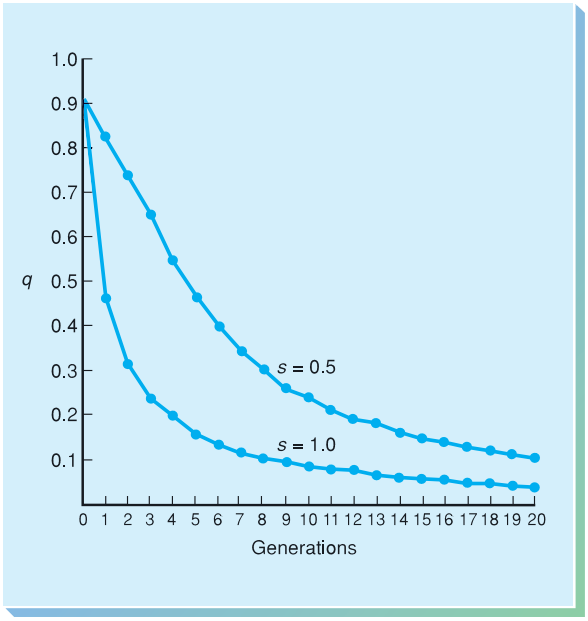

Figure 20.11 Decline in q (the frequency of the *a* allele) under different intensities of selection against the *aa* homozygote. Note that the loss of the *a* allele is asymptotic in both cases, but the drop in allelic frequency is more rapid with the larger selection coefficient.

Table 20.3 **Relative Occurrence of Heterozygotes and Homozygotes as Allelic Frequency Declines: $q = f(a)$; $p = f(A)$**

q	$f(Aa)$ $(2pq)$	$f(aa)$ (q^2)	$f(Aa)/f(aa)$
0.5	0.50	0.25	2
0.2	0.32	0.04	8
0.1	0.18	0.01	18
0.01	0.0198	0.0001	198
0.001	0.001998	0.000001	1,998

$q = 0.001$, there are almost two thousand heterozygotes per *aa* homozygote. Remember, only the recessive homozygote is selected against. Natural selection cannot distinguish the dominant homozygote from the heterozygote.

Selection-Mutation Equilibrium

Although a deleterious allele is eliminated slowly from a population, the time frame is so great that there is opportunity for mutation to bring the allele back. Given a population in which alleles are removed by selection and added by mutation, the point at which no change in allelic frequency occurs, the **selection-mutation equilibrium,** may be determined as follows. The new frequency (q_{n+1}) of the recessive *a* allele after nonlethal selection ($s < 1$) against the recessive homozygote is obtained by equation 20.15:

$$q_{n+1} = \frac{q(1 - sq)}{1 - sq^2}$$

Change in allelic frequency under this circumstance will thus be

$$\Delta q = q_{n+1} - q = \frac{q(1 - sq)}{(1 - sq^2)} - \frac{q(1 - sq^2)}{(1 - sq^2)}$$

$$= \frac{q - sq^2 - q + sq^3}{(1 - sq^2)}$$

$$= \frac{-sq^2(1 - q)}{1 - sq^2} \qquad \textbf{(20.20)}$$

Equation 20.20 is the general form of equation 20.18 for any value of *s*. The change in allelic frequency due to mutation can be found by using equation 20.4:

$$\Delta q = \mu p - \nu q$$

where μ and ν are the rate of forward and back mutation, respectively. When equilibrium exists, the change from selection will just balance the change from mutation. Thus,

$$\mu p - \nu q + \frac{-sq^2(1 - q)}{1 - sq^2} = 0$$

and

$$\mu p - \nu q = \frac{sq^2(1 - q)}{1 - sq^2} \qquad \textbf{(20.21)}$$

Now, some judicious simplifying is justified, because in a real situation, q will be very small because the *a* allele is being selected against. Thus, νq will be close to zero, and $1 - sq^2$ will be close to unity. Equation 20.21, therefore, becomes:

$$\mu p \cong sq^2(1 - q)$$

$$\mu(1 - q) \cong sq^2(1 - q)$$

$$q^2 \cong \mu/s$$

$$\hat{q} \cong \sqrt{\mu/s} \qquad \textbf{(20.22)}$$

In the case of a recessive lethal, *s* would be unity, so

$$q^2 \cong \mu \text{ and } \hat{q} \cong \sqrt{\mu}$$

If a recessive homozygote has a fitness of 0.5 ($s = 0.5$) and a mutation rate, μ, of 1×10^{-5}, the allelic frequency at selection-mutation equilibrium will be

$$\hat{q} \cong \sqrt{\mu/s} \cong \sqrt{1 \times 10^{-5}/0.5} \cong \sqrt{2 \times 10^{-5}}$$

$$\cong 0.004$$

If the recessive phenotype were lethal, then

$$\hat{q} \cong \sqrt{\mu/s} \cong \sqrt{1 \times 10^{-5}/1}$$

$$\cong 0.003$$

These are very low equilibrium values for the *a* allele.

Types of Selection Models

In view of the limited ways that fitnesses can be assigned, only a limited number of selection models are possible. Table 20.4 lists all possible selection models if we assume that fitnesses are constants and the highest fitness is one. (You might now go through the list of models and determine the equilibrium conditions for each.) Note that two possible fitness distributions are missing. There is no model in which fitnesses are $1-s$, 1, and 1 for the A_1A_1, A_1A_2, and A_2A_2 genotypes, respectively (remembering that $p = f[A_1]$ and $q = f[A_2]$). That model is for selection against the A_1A_1 homozygote. Some reflection should show that this is the same model as model 1 of table 20.4, except that the A_1 allele is acting like a recessive allele. In other words, natural selection acts against A_1A_1 homozygotes, but not against the A_1A_2 and A_2A_2 genotypes. Thus, the model reduces to model 1 if we treat A_1 as the recessive allele and A_2 as the

Table 20.4 All Possible One-Locus, Two-Allele Selection Models (Assuming All Selection Coefficients Are Constants)

	Genotypic Fitness		
Type of Selection	A_1A_1	A_1A_2	A_2A_2
1. Against recessive homozygotes	1	1	$1 - s$
2. Against heterozygotes	1	$1 - s$	1
3. Against one allele	1	$1 - s_1$	$1 - s_2$
4. Against homozygotes	$1 - s_1$	1	$1 - s_2$

dominant allele. Similarly, the $(1 - s_1, 1 - s_2, 1)$ model is eliminated for the same reason (allele A_2 is acting like the dominant allele and A_1 like the recessive allele). We now describe the outcome of each of the models in the table.

In both models 1 and 3 (table 20.4), selection is against genotypes containing the A_2 allele. Model 1, which we just derived in detail, is the model for a deleterious recessive allele. Almost any enzyme defect in a metabolic pathway fits this model, such as PKU, alkaptonuria, Tay-Sachs disease, and so on. In model 3, however, natural selection can detect the heterozygote, as is the case with deleterious alleles that are not completely recessive. An example would be the hemoglobin anomaly called thalassemia, a disorder common in some European and Asian populations, that produces a severe anemia in homozygotes and a milder anemia in heterozygotes. It should be clear that selection can more quickly eliminate a partially recessive allele than a completely recessive allele because the allele can no longer "hide" in the heterozygote.

Dominant or semidominant alleles (model 3) are usually more quickly removed from a population because they are completely open to selection. It takes an infinite number of generations to remove a recessive lethal allele, but only one generation for natural selection to remove a completely dominant lethal allele (see model 3, where $s_1 = s_2 = 1$). Examples of dominant deleterious traits in people are Huntington disease, facioscapular muscular dystrophy, and chondrodystrophy.

Model 2 is interesting because selection against the heterozygote leads to an unstable equilibrium at $q = 0.5$. If one heterozygote is removed by selection, one each of the two alleles is eliminated. However, if p and q are not equal (and thus not equal to 0.5), then one A_1 allele is not the same proportion of the A_1 alleles as one A_2 allele is of all the A_2 alleles. In other words, in a population of fifty individuals with $q = 0.1$ and $p = 0.9$, one A_2 allele is 10% (1/10) of the A_2 alleles, whereas one A_1 allele is only 1.1% (1/90) of the A_1 alleles. Removing one each of the two alleles causes a decrease in q. Therefore, a population following model 2 is at equilibrium at $p = q = 0.5$. However, this is an unstable equilibrium. Any perturbation that changes the allelic frequencies causes the rarer allele to be selected against and eventually removed from the population. An example is the maternal-fetal incompatibility at the Rh locus in human beings. The disease erythroblastosis occurs only in heterozygous fetuses (Rh^+Rh^-) in Rh-negative (Rh^-Rh^-) mothers. Heterozygotes are, therefore, selected against.

In model 4, selection is against homozygotes. This model is called the **heterozygote advantage,** and we will derive its equilibrium condition because the results are important to evolutionary theory (table 20.5). At equilibrium

$$\Delta q = \frac{pq(s_1 p - s_2 q)}{\overline{W}} \qquad \textbf{(20.23)}$$

For this expression to be zero, either

$$p = 0,\ q = 0,\ \text{or}\ (s_1 p - s_2 q) = 0$$

If $p = 0$ or $q = 0$, the result is trivial; the equilibrium exists only because of the absence of one of the alleles. The more meaningful equilibrium occurs when $s_1 p - s_2 q = 0$. In that case

$$s_1 p = s_2 q \text{ or } s_1(1 - q) = s_2 q$$

and

$$\hat{q} = \frac{s_1}{s_1 + s_2} \qquad \textbf{(20.24)}$$

Table 20.5 Selection Model of Heterozygote Advantage: The A Locus with A_1 and A_2 Alleles

	Genotype			
	A_1A_1	A_1A_2	A_2A_2	**Total**
Initial genotypic frequencies	p^2	$2pq$	q^2	1
Fitness (W)	$1 - s_1$	1	$1 - s_2$	
Ratio after selection	$p^2(1 - s_1)$	$2pq$	$q^2(1 - s_2)$	$1 - s_1p^2 - s_2q^2 = \overline{W}$
Genotypic frequencies after selection	$\frac{p^2(1 - s_1)}{\overline{W}}$	$\frac{2pq}{\overline{W}}$	$\frac{q^2(1 - s_2)}{\overline{W}}$ [1]	

BOX 20.1

Experimental Methods

A General Computer Program to Simulate the Approach to Allelic Equilibrium Under Heterozygote Advantage

It is surprising how much insight we can gain into the processes of population genetics by modeling them on a computer. The simple computer program presented here calculates changing allelic frequencies due to random mating when alleles at a locus are under a heterozygote-advantage selection regime. The program is written in the Microsoft® Visual Basic language. You can simulate any of the selection models described in chapter 20 by simply changing the variables. Also, this program can model many of the other processes discussed in this and the last chapter; usually, only a few lines need to be changed to look at an entirely different process. Other computer programs can substitute. Output should be graphed. The program should be rerun several times with various sets of values for the allelic frequencies and fitnesses. If the outcome isn't clear by twenty-five generations, the number of generations can be increased with a few small changes in the program.

In the computer program (fig. 1), p is set to 0.9, q is $1 - p$ (0.1), and the three fitnesses are named w11, w12, and w22 for the *AA*, *Aa*, and *aa* genotypes, respectively. In this case, w11 is set to 0.4, w12 to 1, and w22 to 0.6, a model of heterozygote advantage;

continued

```
Sub Command1_Click ()
    Static q(25)
    Static p(25)
    Picture1.Cls

'Set variables

    p(1) = .9
    w11 = .4
    w12 = 1
    w22 = .6
    q(1) = 1 - p(1)

'Calculate p and q values

    For i = 2 To 25
      wbar = p(i - 1) ^ 2 * w11 + 2 * p(i - 1) * q(i - 1) * w12 + q(i - 1) ^ 2 * w22
      q(i) = (q(i - 1) ^ 2 * w22 + p(i - 1) * q(i - 1) * w12) / wbar
      p(i) = 1 - q(i)
    Next i

'Draw axes and grid

    Picture1.Scale (-1, 1.1)-(26, -.1)
    Picture1.Line (0, 0)-(0, 1)
      For i = 0 To 10
        Picture1.Line (0, .1 * i)-(25, .1 * i)
      Next i
      For i = 5 To 25 Step 5
        Picture1.Line (i, 0)-(i, 1)
      Next i

'Draw q values

    Picture1.DrawWidth = 5
      For i = 1 To 25
        Picture1.PSet (i, q(i))
      Next i
    End Sub
```

Figure 1 A Microsoft® Visual Basic computer program for the simulation of heterozygote advantage. The first statement indicates that the program is run by clicking a command button. Twenty-five values of q and p are calculated and stored for printing. The program also prints a grid of lines at increments of $q = 0.1$ and generations $= 5$.

BOX 20.1 CONTINUED

the number of generations is twenty-five. The program calculates the mean fitness of the population, wbar, as p^2(w11) + 2pq(w12) + q^2(w22); it then calculates the new allelic frequencies after one generation of selection (the new proportion of *aa* homozygotes plus half the proportion of heterozygotes). The program then repeats this process twenty-five times, storing each new q in the array q(i). The graphic output shown in figure 2 results. As you can see, q is approaching 0.6. If you would like to see the values generated by the program, appropriate print statements can be added.

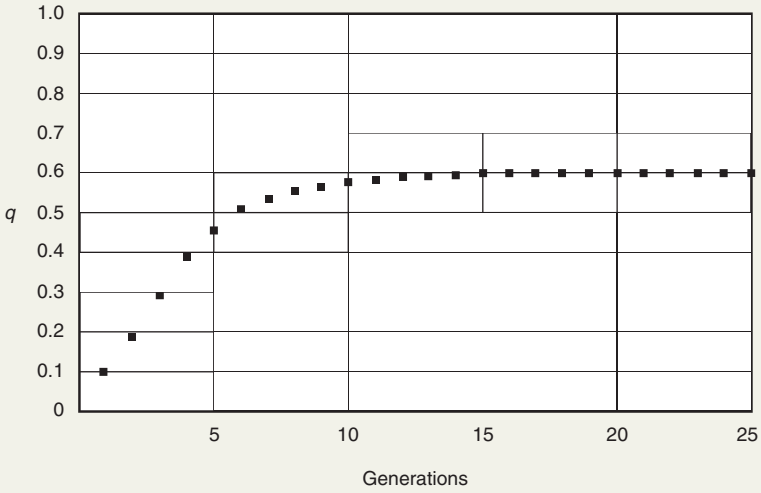

Figure 2 Graphical output of the computer program from figure 1, with axis labels drawn in. The frequency of the *a* allele, q, begins at 0.1 and asymptotes toward 0.6.

Since $p + q = 1$,

$$\hat{p} = \frac{s_2}{s_1 + s_2} \qquad \textbf{(20.25)}$$

Several interesting conclusions follow. First, unlike the other models of selection, this model allows a population to maintain both alleles. We can demonstrate that this equilibrium is stable by graphing the Δq value against q. Such a graph appears in figure 20.12, in which q is the frequency of allele A_2 and the fitnesses of genotypes A_1A_1, A_1A_2, and A_2A_2 are assumed to be 0.8, 1, and 0.7, respectively. Note that if the equilibrium is perturbed by an increase or decrease in q, the population returns to the point of equilibrium. Second, the equilibrium is independent of the original allelic frequencies since it involves only the selection coefficients, s_1 and s_2. Last, the equilibrium for each allele (equations 20.24 and 20.25) is directly proportional to the selection coefficient against the other allele. As the selection against A_1 increases (s_1 increases), the equilibrium shifts toward a higher value of q (more A_2 alleles; box 20.1).

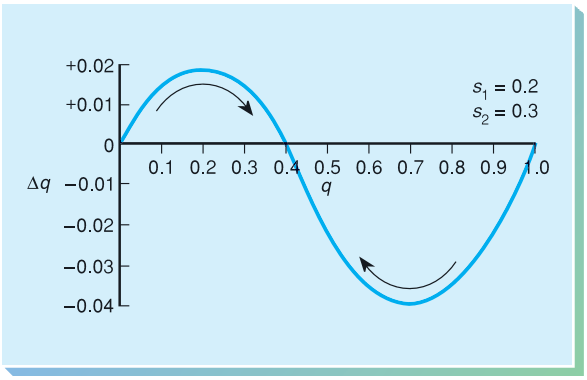

Figure 20.12 Plot of allelic frequency (q) versus change in allelic frequency (Δq) for a polymorphism maintained by heterozygote advantage. In this case, s_1 = 0.2 and s_2 = 0.3; the equilibrium value, $\hat{q}$, is 0.4. When perturbed, the population tends to return to this value unless the perturbation brings q to either 1.0 or 0.0, in which case the population is either fixed for the *a* allele or has lost it. In both cases, no further change in allelic frequency will take place, barring mutation or migration.

SUMMARY

STUDY OBJECTIVE 1: To develop ways to analyze population genetics problems 571

A five-step protocol is presented to determine equilibrium allelic frequencies.

STUDY OBJECTIVE 2: To analyze the effects of mutation, migration, and population size on the Hardy-Weinberg equilibrium 571–577

The effects of relaxing some of the assumptions of the Hardy-Weinberg equilibrium are analyzed. Both mutation and migration transport alleles in and out of a population. Mutation provides the variability on which natural selection acts, but it usually does not directly affect the equilibrium because mutation rates are usually very low. If two randomly mating populations merge, or if two randomly mating demes are mistakenly treated as a single deme, the conglomerate will be deficient in heterozygotes. This deviation is called the Wahlund effect.

Finite population size is a source of sampling error. It results in changes in allelic frequencies known as random genetic drift. The smaller the population, the more rapidly allelic frequencies change. The dynamics of random genetic drift were studied graphically.

STUDY OBJECTIVE 3: To study the ways in which natural selection results in organisms adapted to their environments 577–584

Natural selection is defined by differential reproductive success. Depending upon which phenotypes are most fit, natural selection can act in several ways to change allelic and genotypic frequencies. Selection against the recessive homozygote acts to remove the allele from the population. Mutation brings the allele back into the population. Thus, a selection-mutation equilibrium maintains the unfavorable allele at a relatively low frequency. Heterozygote advantage maintains both alleles in a population.

SOLVED PROBLEMS

PROBLEM 1: At a particular locus, there are two alleles, *B* and *b*. The mutation rate of *B* to *b* is 3.5×10^{-4}, whereas the mutation rate of *b* to *B* is 6×10^{-8}. What is the equilibrium frequency of the *b* allele, assuming no other factor is operating in this population to disturb the Hardy-Weinberg equilibrium?

Answer: We let $q = f(b)$, $\mu = 3.5 \times 10^{-4}$, and $\nu = 6 \times 10^{-8}$. We then simply substitute μ and ν into equation 20.6:

$$\hat{q} = \mu/(\mu + \nu) = 3.5 \times 10^{-4}/(3.5 \times 10^{-4} + 6 \times 10^{-8}) = 0.9998$$

PROBLEM 2: Given a population of about one million cicadas with a frequency of the *a* allele at the *A* locus of 0.75, what is the probability that the *a* allele will be lost due to random genetic drift? How much longer will the possible loss of the allele take than the loss of the allele would take in a population of one thousand?

Answer: Regardless of the size of a finite population, random genetic drift takes place. The probability of the loss of an allele with a frequency of 0.75 is 0.25; the probability of its fixation is 0.75 (see fig. 20.8). Since it is convenient to measure time (number of generations) within populations of finite size in units of population size, we can see that an event that takes *N* generations will be one thousand generations in the small population, but one million generations in the large population. Thus, random genetic drift occurs in the larger population at about one-thousandth the rate of the small population.

PROBLEM 3: In a laboratory colony of fruit flies, the fitnesses of the genotypes of an electrophoretic locus (malate dehydrogenase) are determined. Three genotypes, *FF*, *FS*, and *SS*, have fitnesses of 0.85, 1.0, and 0.6, respectively. What is the equilibrium frequency of the slow allele (*S*)?

Answer: If the fitnesses of the three genotypes *FF, FS,* and *SS* are as given, then the locus is exhibiting heterozygote advantage with selection coefficients of the two homozygotes of $s_1 = 0.15\ (1 - 0.85)$ and $s_2 = 0.4\ (1 - 0.6)$. If q is the frequency of the slow allele, then, using equation 20.24,

$$\hat{q} = s_1/(s_1 + s_2) = 0.15/(0.15 + 0.4) = 0.27$$

EXERCISES AND PROBLEMS*

MUTATION

1. Consider a locus with alleles A and a in a large, randomly mating population under the influence of mutation.
   a. If the mutation rate of A to a is 6×10^{-5}, and the back-mutation rate to A is 7×10^{-7}, what is the equilibrium frequency of a?
   b. If $q = 0.9$ in generation n, what would it be one generation later, under only the influence of mutation?

2. Derive an expression for mutation equilibrium when no back mutation is occurring.

3. Consider a population in which $p = 0.9$ and $q = 0.1$. If the forward mutation rate, $A \rightarrow a$, is 5×10^{-5} and the reverse mutation rate, $a \rightarrow A$, is 2×10^{-5}, calculate the equilibrium frequency, $\hat{q}$, of the a allele.

4. If the forward mutation rate, $A \rightarrow a$, is five times the reverse mutation rate, what is the equilibrium frequency of the a allele?

MIGRATION

5. The following data refer to the R° allele in the Rh blood system:

frequency in western Europeans = 0.62

frequency in eastern Europeans = 0.45

frequency in Mongols = 0.03

What is the total proportion of alleles that have entered the eastern European population?

6. Given the data from problem 1 of chapter 19, what factors could have caused the population to leave Hardy-Weinberg equilibrium? (*See also* SMALL POPULATION SIZE and NATURAL SELECTION)

7. In a population of nine hundred butterflies, the frequency (p) of the fast allele of the enzyme phosphoenol pyruvate is 0.6, and the frequency of the slow form (q) is 0.4. Ninety butterflies migrate to this population, and the migrants have a slow-allele frequency of 0.8. Calculate the allelic frequencies of the new population.

8. If the frequency of the N allele is 0.25 in a native population, 0.32 in a conglomerate population, and 0.4 in a migrant population, what percentage of the N alleles in the conglomerate population were derived from the migrant population?

9. In a particular population, the frequency of allele t was 0.25 in a migrant population and 0.45 in the conglomerate population. If the migration rate was 0.1, calculate the frequency of t in the original, native population.

SMALL POPULATION SIZE

10. In a population of five hundred individuals with a frequency of allele A of 0.7, what is the ultimate fate of the A allele? What is the probability that the population will eventually lose the A allele? How many are $N/5$ generations? $4N$ generations?

NATURAL SELECTION

11. Differentiate among stabilizing, directional, and disruptive selection.

12. Derive a model of selection in which the fitness of the heterozygote is half the fitness of one of the homozygotes and twice the fitness of the other. Give expressions for the following:
   a. Mean population fitness
   b. Equilibrium allelic frequency (stable?)

13. Derive an expression for the equilibrium allelic frequencies under a model in which selection acts against heterozygotes. Is the equilibrium stable?

14. Table 20.6 describes selection at the A locus in a given diploid species in which $p = f(A)$ and $q = f(a)$.
   a. Describe the type of selection occurring here. Why does the total equal one before selection but $\overline{W}$, after?
   b. Derive an equation for q after one generation of selection (q_{n+1}).
   c. This system will reach equilibrium, with $\hat{p} = s_2/(s_1 + s_2)$. If selection is twice as strong against aa as against AA, what are the equilibrium allelic frequencies? If $s_1 = 0.1$ and $s_2 = 0.3$, what percentage of heterozygotes is at equilibrium?

15. Given a locus with alleles A and a in a sexually reproducing, diploid population in Hardy-Weinberg equilibrium, set up a model and the initial formula for the frequency of the dominant allele after one generation (p_n+1) if selection acts against the dominant phenotype. What are the equilibrium conditions?

16. There is a locus with alleles A and a in a large, randomly mating, diploid, sexually reproducing population. Allele A mutates to a at a rate of μ, and no back mutation takes place. However, the aa homozygote is selected against with a fitness of $1 - s$. Give a formula for the equilibrium condition. If $\mu = 5 \times 10^{-5}$

* Answers to selected Exercises and Problems are on page A-22.

Table 20.6

	Genotypes			
	AA	*Aa*	*aa*	Total
Before selection	p^2	$2pq$	q^2	1
Fitness (W)	$1 - s_1$	1	$1 - s_2$	
After selection	$p^2(1 - s_1)$	$2pq$	$q^2(1 - s_2)$	$\overline{W} = 1 - s_1p^2 - s_2q^2$

and $s = 0.15$, what are the equilibrium allelic frequencies?

17. If a locus has alleles A_1 and A_2, what is the equilibrium frequency of A_1 if both homozygotes are lethal?

18. The following data were collected from a population of *Drosophila* segregating sepia (s) and wild-type (s^+) eye colors. One sample was taken when the eggs were deposited, and another was taken later among adults. Reconstruct the mode of selection.

	s^+s^+	s^+s	ss
Egg	25	50	25
Adult	30	60	10

19. The data in table 20.7 come from T. Dobzhansky's work with chromosomal inversions in *Drosophila pseudoobscura.* They represent four samples from various altitudes in the Sierra Nevada Mountains in California. What would you say about, and what would you do in the lab to determine, the fitnesses of the inversions? What factors could cause the changes in fitness?

20. In a particular population with two alleles at a locus, the frequency of *AA* individuals = 0.25, *Aa* = 0.5, and *aa* = 0.25. If the *AA* genotype fitness = 1, *Aa* = 0.8, and *aa* = 0.6, what will the frequencies of *A* and *a* be in the next generation? Assume mutations do not occur.

Table 20.7 Data from Dobzhansky's Work

Elevation of Sample	Inversion			
	ST	*AR*	*CH*	Others
6,800 ft	26	44	16	14
4,600 ft	32	37	19	12
3,000 ft	41	35	14	10
800 ft	46	25	16	13

Note: *ST* = standard; *AR* = Arrowhead; *CH* = Chiricahua

21. Calculate the frequency of the recessive *b* allele in a population one generation after selection if in the original population $q = f(b) = 0.7$ and the relative fitness of *bb* homozygotes is 0.4.

22. A type of dwarfism in dogs is caused by a recessive allele. The mutation rate from the normal to the mutant allele has been estimated at 5×10^{-5}, and the fitness of the dwarf is 0.2 when compared with normal individuals. Calculate the equilibrium frequency of the dwarf allele.

23. A recessive allele ($q = 0.5$) was initially neutral, but suddenly the environment changed and the recessive homozygote became lethal. What is q one generation after selection begins? What is the expected frequency of the recessive allele two generations after selection?

CRITICAL THINKING QUESTIONS

1. If the selection model of heterozygous disadvantage leads to the elimination of the rarer allele, why would such systems (e.g., the Rh blood system) still exist.

2. A scientist studied the distribution of electrophoretic genotypes in a sample of an insect species and found a deficiency of heterozygotes. How could this come about?

Suggested Readings for chapter 20 are on page B-20.