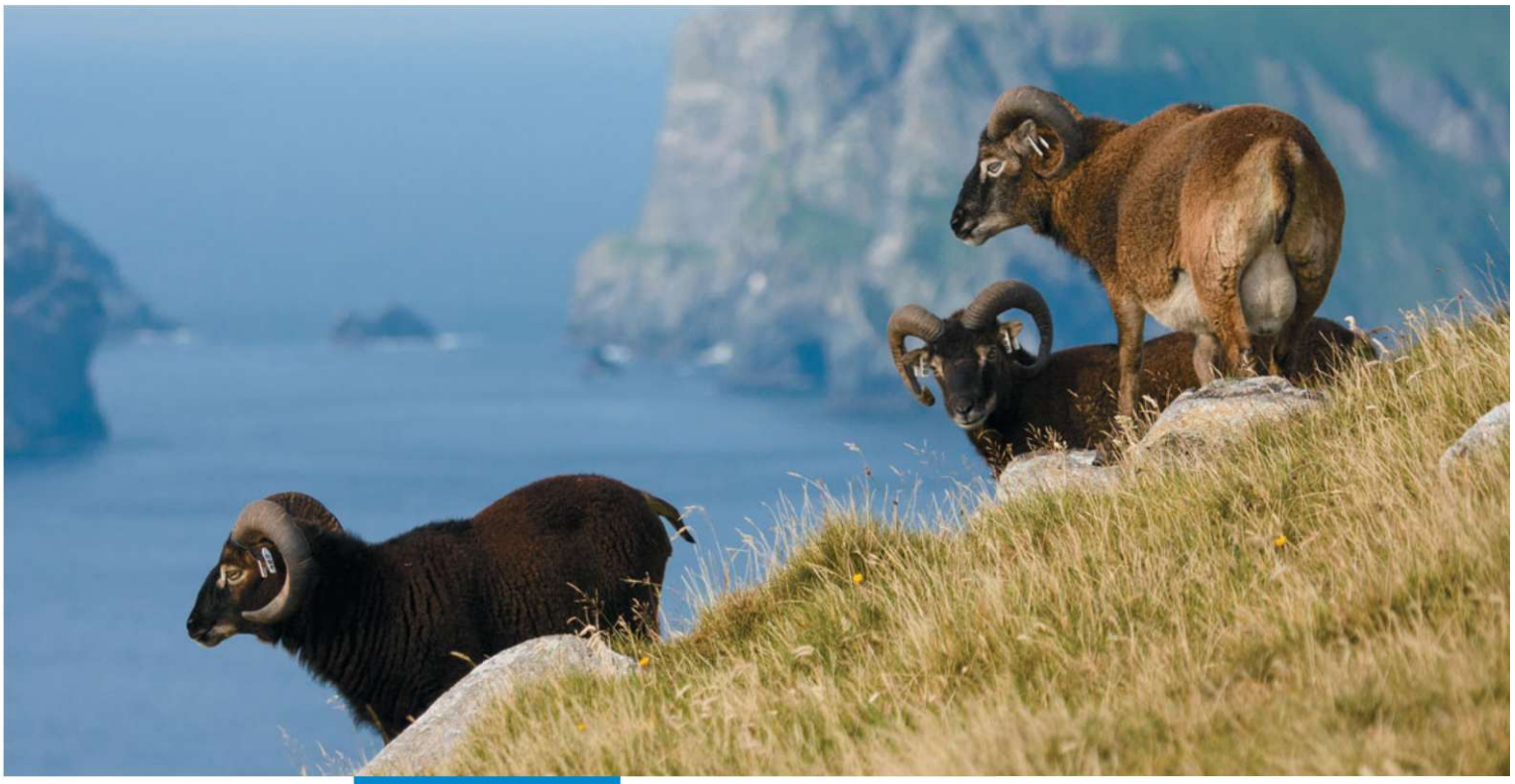


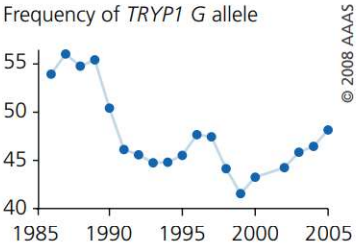
6

Mendelian Genetics in Populations I: Selection and Mutation

Darwin's theory of evolution by natural selection provides a mechanistic explanation of descent with modification that is supported by considerable evidence. However, as Darwin himself recognized, the theory is incomplete without an accurate understanding of the mechanism of inheritance (see Darwin 1868). That understanding has been provided by Mendelian and molecular genetics. With Darwin's insights and modern genetics, we have the tools we need to develop a more complete model of the mechanism of evolution.

Population genetics, the subject of this chapter (as well as Chapters 7 and 8), integrates evolution by natural selection with Mendelian genetics (for a history, see Provine 1971). The crucial insight of population genetics is that changes in the relative abundance of traits in a population can be tied to changes in the relative abundance of the genetic variants that influence them. A decline over several decades in the frequency of dark-colored Soay sheep in St. Kilda, Scotland, for example, is tied to a decline in the frequency of the dominant allele responsible for dark coloration. From a population geneticist's perspective, evolution can be defined as change across generations in the frequencies of alleles. Population genetics provides the theoretical foundation for much of our modern understanding of evolution.

Some Soay sheep are light, others dark, due to alleles of the gene for tyrosinase-related protein 1 (TRYP1). Recently, the frequency of the *G* allele, which confers dark color, has declined—and with it the frequency of dark sheep (Graph from Gratten et al. 2008; photo by Arpat Ozgul).



Scientists evaluate theories by using them to make predictions, then checking whether the predictions come true. Some of the clearest tests of theory are found in engineering. When Neil Armstrong and Buzz Aldrin traveled to the surface of the moon and back in July of 1969, they demonstrated that NASA's engineers understand a thing or two about thrust, inertia, and gravity (Figure 6.1). In this chapter, we present data from predictive tests of population genetics theory. At the end, we tell the story of a team of genetic engineers who designed and built a new gene, introduced it into a population of fruit flies, and used population genetics theory to predict the trajectory of its changing frequency 20 generations into the future.

Our first task, however, is to introduce the fundamentals of population genetics. In Section 6.1, we introduce an algebraic model that allows us to track Mendelian alleles across generations in an idealized population under simplifying assumptions. This model will show us circumstances under which evolution does not occur. In Section 6.2, we relax one of the simplifying assumptions and learn to predict how populations evolve under natural selection. In Section 6.3, we look at data that puts a variety of predictions of evolution by natural selection to the test. In Section 6.4, we relax another assumption to look at mutation as a mechanism of evolution. In Section 6.5, we close with the genetic engineering story. Throughout the chapter, we use population genetics theory to address practical issues arising from human diseases and human evolution. The first of these issues involves human evolution in response to the HIV epidemic.

Figure 6.1 Engineering success demonstrates the value of theory *Apollo 11*'s lunar module "Eagle," carrying Neil Armstrong and Buzz Aldrin, returns from the surface of the moon to dock with the command service module. The photo was taken on 21 July 1969 by command module pilot Michael Collins. Note Earth rising in the background.

6.1 Mendelian Genetics in Populations: Hardy–Weinberg Equilibrium

Most people are susceptible to HIV. Their best hope of avoiding infection is to avoid contact with the virus. There are, however, a few individuals who remain uninfected despite repeated exposure. In 1996, AIDS researchers discovered that at least some of this variation in susceptibility has a genetic basis (see Chapters 1 and 5). The gene responsible encodes a cell-surface protein called CCR5. CCR5 is the handle exploited by most sexually transmitted strains of HIV-1 as a means of binding to white blood cells. A mutant allele of the CCR5 gene, called *CCR5-Δ32,* has a 32-base-pair deletion that destroys the encoded protein's ability to function. Individuals who inherit two copies of this allele have no CCR5 on the surface of their cells and are therefore resistant to HIV-1. Given that individuals homozygous for *CCR5-Δ32* are much less likely to contract HIV, we might ask whether the global AIDS epidemic will cause an increase in the frequency of the *Δ32* allele in human populations. If so, how fast will it happen?

Before we can hope to answer such questions, we need to understand how the *CCR5-Δ32* allele would behave without the AIDS epidemic. In other words, we need to develop a null model for the behavior of genes in populations. This null model should specify, under the simplest possible assumptions, what will happen across generations to the frequencies of alleles and genotypes. The model should apply not just to humans, but to any population of organisms that are both diploid and sexual. In this first section of the chapter, we develop such a model and explore its implications. In the next section we add natural selection to the model, which will enable us to address our questions about the AIDS epidemic and the *CCR5-Δ32* allele.

Population genetics begins with a model of what happens to allele and genotype frequencies in an idealized population. Once we know how Mendelian genes behave in the idealized population, we will be able to explore how they behave in real populations.

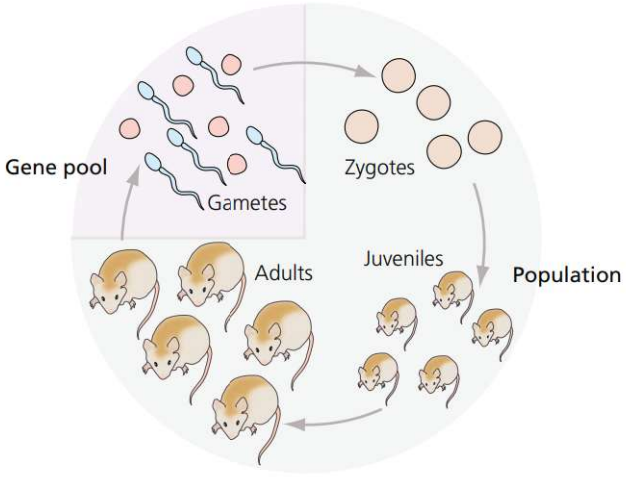


Figure 6.2 The life cycle of an idealized population The labels highlight the stages that will be important in our development of population genetics.

We develop our model by scaling Mendelian genetics up from the level of families, where the reader has used it until now, to the level of populations. We illustrate the model with an idealized population of mice (Figure 6.2). A **population** is a group of interbreeding individuals and their offspring. The crucial events in the life cycle of a population are these: Adults produce gametes, gametes combine to make zygotes, zygotes develop into juveniles, and juveniles grow up to become the next generation of adults. We want to track the fate of Mendelian genes in a population. We want to know whether particular alleles or genotypes become more common or less common across generations, and why.

Imagine that the mice in Figure 6.2 have in their genome a Mendelian locus, the A locus, with two alleles: *A* and *a*. We can begin tracking these alleles at any point in the life cycle. We then follow them through one complete turn of the cycle, from one generation to the next, to see if their frequencies change.

A Simulation

Our task of following alleles around the life cycle will be simplest if we start with the gametes produced by the adults when they mate. We will assume that the adults choose their mates at random. A useful mental trick is to picture random mating happening like this: We take all the eggs and sperm produced by all the adults in the population, dump them together in a barrel, and stir. This barrel is known as the **gene pool**. Each sperm in the gene pool swims about at random until it collides with an egg, whereupon the egg and sperm fuse to make a zygote. Something rather like this actually happens in sea urchins and other marine creatures that simply release their gametes onto the tide. For other organisms, like mice and humans, this picture is obviously a simplification.

The adults in our mouse population are diploid, so each carries two copies of the A locus. But the adults made their eggs and sperm by meiosis. Following Mendel's law of segregation, each gamete received just one copy of the A locus. Imagine that 60% of the eggs and sperm received a copy of allele *A*, and 40% received a copy of allele *a*. That is, the frequency of allele *A* in the gene pool is 0.6, and the frequency of allele *a* is 0.4.

Starting with the eggs and sperm that constitute the gene pool, our model tracks alleles through zygotes and adults and into the next generation's gene pool.

What happens when eggs meet sperm? For example, what fraction of the zygotes they produce have genotype *AA*? And once these zygotes develop into juveniles, grow up, and spawn, what are the frequencies of alleles *A* and *a* in the next generation's gene pool?

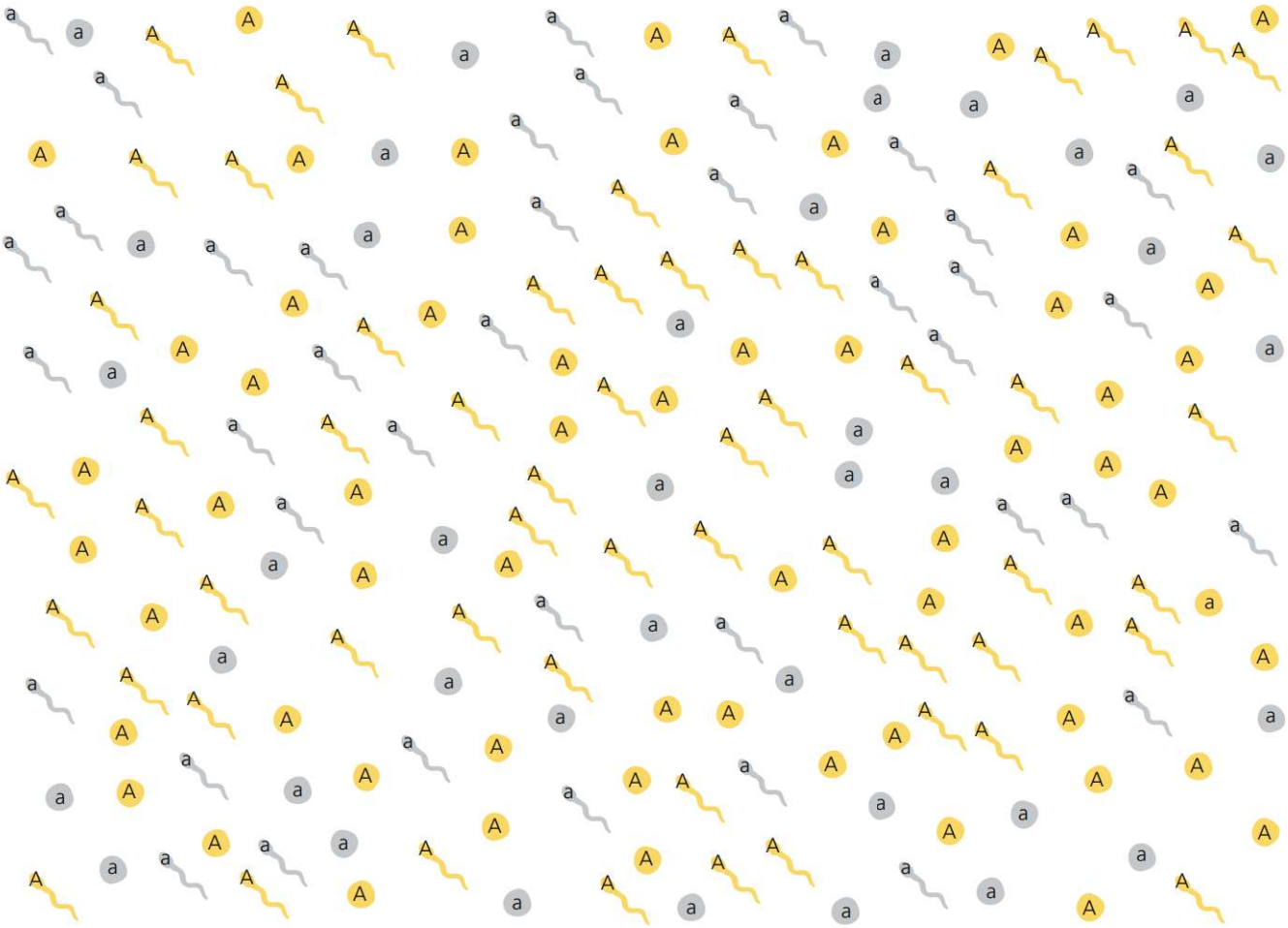


Figure 6.3 **A gene pool with frequencies of 0.6 for allele *A* and 0.4 for allele *a***

One way to find out is by simulation. We can close our eyes and put a finger down on Figure 6.3 to choose an egg. Perhaps it carries a copy of allele *A*. Now we close our eyes and put down a finger to choose a sperm. Perhaps it carries a copy of allele *a*. If we combine these gametes, we get a zygote with genotype *Aa*. We encourage the reader to carry out this process to make a large sample of zygotes—50, say, or even 100. We have paused to do so as we write. Among the 100 zygotes we made, 34 had genotype *AA*, 57 had *Aa*, and 9 had *aa*.

Now let us imagine that all these zygotes develop into juveniles, and that all the juveniles survive to adulthood. Imagine, furthermore, that when the adults reproduce, they all donate the same number of gametes to the gene pool. We can choose any number of gametes we like for the standard donation, so we will choose 10 to make the arithmetic easy. We are not worried about whether a particular adult makes eggs or sperm; instead, we are simply counting gametes:

Our 34 *AA* adults together make 340 gametes: 340 carry allele *A*; none carry allele *a*.

Our 57 *Aa* adults together make 570 gametes: 285 carry allele *A*; 285 carry allele *a*.

Our 9 *aa* adults together make 90 gametes: none carry allele *A*; 90 carry allele *a*.

Summing the gametes carrying copies of each allele, we get 625 carrying *A* and 375 carrying *a*, for a total of 1,000. The frequency of allele *A* in the new gene pool is 0.625; the frequency of allele *a* is 0.375.

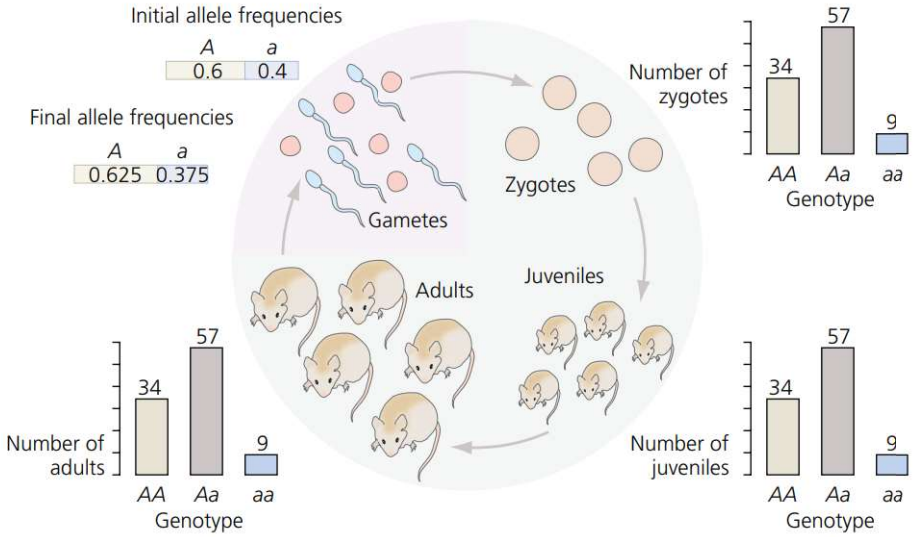


Figure 6.4 Allele and genotype frequencies throughout the life cycle in a numerical simulation We made the zygotes by picking gametes at random from the gene pool in Figure 6.3 and assumed that all the zygotes survived. The reader's results, on repeating this exercise, will likely be somewhat different.

We have followed the alleles around one complete turn of the population's life cycle and found that their ending frequencies are somewhat different from their starting frequencies (Figure 6.4). In other words, our population has evolved.

The genotype frequencies among the zygotes in the reader's sample, and the frequencies of the alleles in the reader's next generation, will almost certainly be somewhat different from ours. Indeed, we carried out the simulation two more times ourselves and got different results each time. In our second simulation, we got zygotes in proportions of 41% *AA*, 44% *Aa*, and 15% *aa*. The allele frequencies in the next generation's gene pool were 0.63 for *A* and 0.37 for *a*. In our third simulation, we got zygotes in proportions of 34% *AA*, 49% *Aa*, and 17% *aa*. The allele frequencies in the next generation were 0.585 for *A* and 0.415 for *a*.

In simulated populations, allele frequencies change somewhat across generations. This is evolution resulting from blind luck.

Our three results are not wildly divergent, but neither are they identical. In two cases the frequency of allele *A* rose, whereas in one it fell. We got different results because in each simulation, blind luck in picking gametes from the gene pool gave us a different number of zygotes with each genotype. The fact that blind luck can cause a population to evolve unpredictably is an important result of population genetics. This mechanism of evolution is called **genetic drift**. (We will return to drift in Chapter 7.) For now, however, we are interested not in whether evolution is sometimes unpredictable, but whether it is ever predictable. We want to know what would have happened in our simulations if chance had played no role.

A Numerical Calculation

We can discover the luck-free result of combining eggs and sperm to make zygotes by using a Punnett square. Punnett squares, invented by Reginald Crundall Punnett, are more typically used in Mendelian genetics to predict the genotypes among the offspring of a particular male and female. Figure 6.5, for example, shows the Punnett square for a mating between an *Aa* female and an *Aa* male. We write the genotypes of the eggs made by the female, in the proportions we expect her to make them, along the side of the square. We write the genotypes of the sperm made by the male, in the proportions we expect him to make them, along the top. Then we fill in the boxes in the square to get the genotypes of the zygotes. This Punnett square predicts that among the offspring of an *Aa* female and an *Aa* male, one-quarter will be *AA*, one-half *Aa*, and one-quarter *aa*.

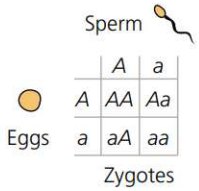


Figure 6.5 Punnett square for a cross between two heterozygotes This device makes accurate predictions about the genotype frequencies among the zygotes because the genotypes of the eggs and sperm are represented in the proportions in which the parents produce them.

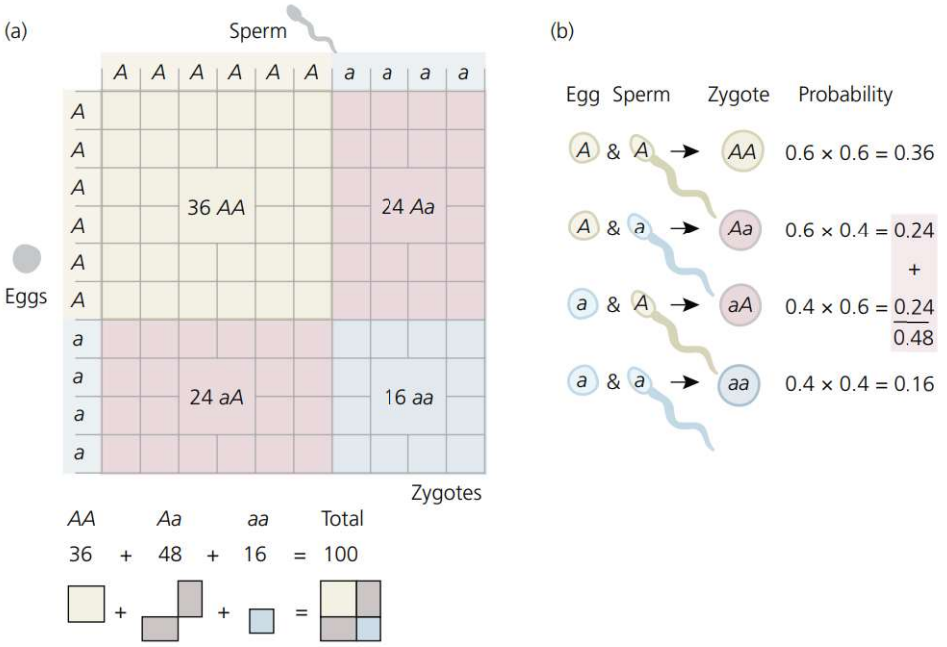


Figure 6.6 When blind luck plays no role, random mating in the gene pool of our model mouse population produces zygotes with predictable genotype frequencies (a) A Punnett square. The genotypes of the gametes are listed along the left and top edges of the box in proportions that reflect the frequencies of *A* and *a* eggs and sperm in the gene pool. The shaded areas inside the box represent the genotypes among 100 zygotes formed by random encounters between gametes in the gene pool. (b) We can also calculate genotype frequencies among the zygotes by multiplying allele frequencies. (See Computing Consequences 6.1.)

We can use the same device to predict the genotypes among the offspring of an entire population (**Figure 6.6a**). The trick is to write the egg and sperm genotypes along the side and top of the Punnett square in proportions that reflect their frequencies in the gene pool. Sixty percent of the eggs carry copies of allele *A* and 40% carry copies of allele *a*, so we have written six *A*'s and four *a*'s along the side of the square. Likewise, for the sperm, we have written six *A*'s and four *a*'s along the top. Filling in the boxes in the square, we find that among 100 zygotes in our population, we can expect 36 *AA*'s, 48 *Aa*'s, and 16 *aa*'s. Note that our population Punnett square has predicted genotype proportions different from the 1:0, 1:1, or 1:2:1 ratios that appear in single-family Punnett squares.

The Punnett square in Figure 6.6a suggests that we could also predict the genotype frequencies among the zygotes by multiplying probabilities. Figure 6.6b shows the four possible combinations of egg and sperm, the resulting zygotes, and a calculation specifying the probability of each (see also **Computing Consequences 6.1**). For example, if we look into the gene pool and pick an egg to watch, there is a 60% chance that it will have genotype *A*. When a sperm comes along to fertilize the egg, there is a 60% chance that the sperm will have genotype *A*. The probability that we will witness the production of an *AA* zygote is therefore

$$0.6 \times 0.6 = 0.36$$

If we watched the formation of all the zygotes, 36% of them would have genotype *AA*. The calculations in Figure 6.6b show that random mating in the gene pool produces zygotes in the following proportions:

AA	*Aa*	*aa*
0.36	0.48	0.16

(The *Aa* category includes heterozygotes produced by combining either an *A* egg with an *a* sperm or an *a* egg with an *A* sperm.) Notice that

$$0.36 + 0.48 + 0.16 = 1$$

This confirms that we have accounted for all of the zygotes.

COMPUTING CONSEQUENCES 6.1

Combining probabilities

The combined probability that two independent events will occur together is the product of their individual probabilities. For example, the probability that a tossed penny will come up heads is $\frac{1}{2}$. The probability that a tossed dime will come up heads is also $\frac{1}{2}$. If we toss both together, the outcome for the penny is independent of that for the dime. Thus the probability of getting heads on the penny and heads on the dime is $\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$.

The combined probability that one or the other of two mutually exclusive events will occur is the sum of their individual probabilities. When rolling a die we can get a one or we can get a two (among other possibilities), but we cannot get both a one and a two at once. The individual probability of each outcome is $\frac{1}{6}$. The combined probability of rolling either a one or a two is therefore $\frac{1}{6} + \frac{1}{6} = \frac{1}{3}$.

We now let the zygotes grow to adulthood, and we let the adults produce gametes to make the next generation's gene pool. When chance plays no role, will the frequencies of alleles *A* and *a* in the new gene pool change from one generation to the next?

If we assume, as we did before, that 100 adults each make 10 gametes, then

The 36 *AA* adults together make 360 gametes: 360 carry allele *A*; none carry allele *a*.

The 48 *Aa* adults together make 480 gametes: 240 carry allele *A*; 240 carry allele *a*.

The 16 *aa* adults together make 160 gametes: none carry allele *A*; 160 carry allele *a*.

Summing the gametes carrying each allele, we get 600 carrying copies of *A* and 400 carrying copies of *a*, for a total of 1,000. The frequency of allele *A* in the new gene pool is 0.6; the frequency of allele *a* is 0.4.

As Figure 6.7a shows graphically, we can also calculate the composition of the new gene pool using frequencies. Because adults of genotype *AA* constitute

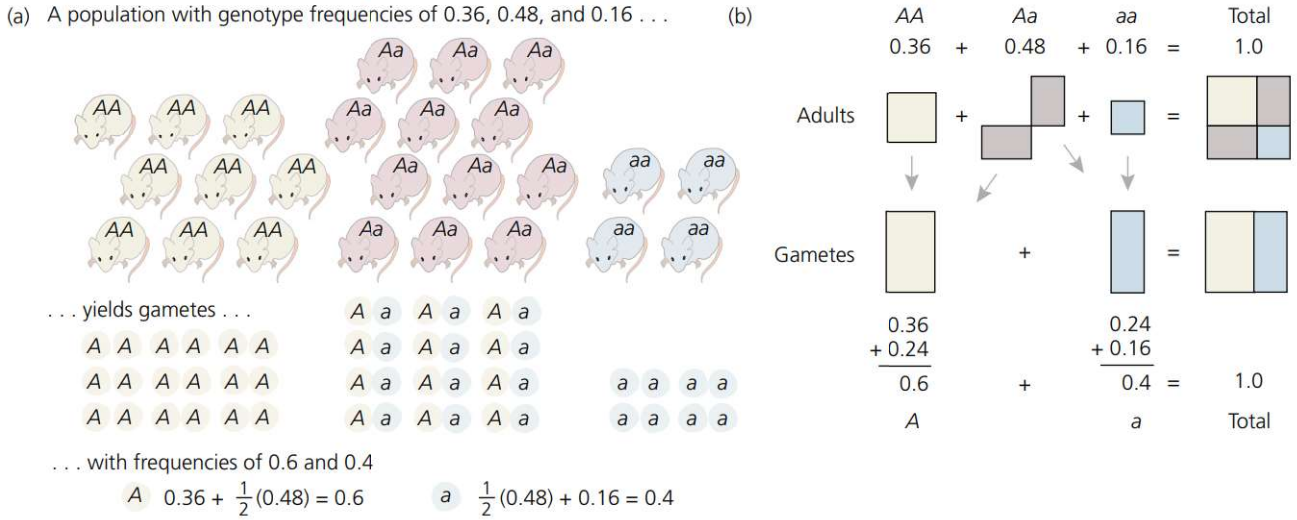


Figure 6.7 When the adults in our model mouse population make gametes, they produce a gene pool in which the allele frequencies are identical to the ones we started with a generation ago (a) Calculations using frequencies. (b) A geometrical representation. The area of each box represents the frequency of an adult or gamete genotype. Note that half the gametes produced by *Aa* adults carry allele *A*, and half carry allele *a*.

36% of the population, they will make 36% of the gametes. All of these gametes carry copies of allele *A*. Likewise, adults of genotype *Aa* constitute 48% of the population and will make 48% of the gametes. Half of these gametes carry copies of allele *A*. So the total fraction of the gametes in the gene pool that carry copies of *A* is

$$0.36 + (\tfrac{1}{2})0.48 = 0.6$$

The figure also shows a calculation establishing that the fraction of the gametes in the gene pool that carry copies of allele *a* is 0.4. Notice that

$$0.6 + 0.4 = 1$$

This confirms that we have accounted for all of the gametes. Figure 6.7b shows a geometrical representation of the same calculations.

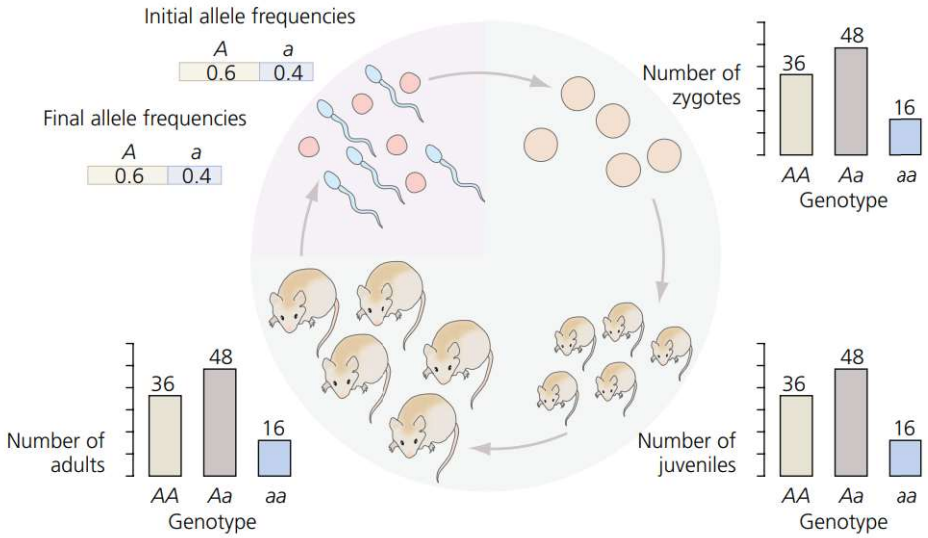


Figure 6.8 When blind luck plays no role in our model population, the allele frequencies do not change from one generation to the next We made the zygotes with the Punnett square in Figure 6.6 and assumed that all the zygotes survived.

We have come full circle (Figure 6.8). And this time, unlike in our simulations, we have arrived precisely where we began. We started with allele frequencies of 60% for *A* and 40% for *a* in our population's gene pool. We followed the alleles through zygotes, juveniles, and adults and into the next generation's gene pool. The allele frequencies in the new gene pool are still 60% and 40%. When blind luck plays no role, the allele frequencies for *A* and *a* in our population are in equilibrium: They do not change from one generation to the next. The population does not evolve.

Numerical examples show that when blind luck plays no role, allele frequencies remain constant from one generation to the next.

The first biologist to work a numerical example, tracing the frequencies of Mendelian alleles from one generation to the next in an ideal population, was G. Udny Yule in 1902. He started with a gene pool in which the frequencies of two alleles were 0.5 and 0.5 and showed that in the next generation's gene pool, the allele frequencies were still 0.5 and 0.5. The reader may want to reproduce Yule's calculations as an exercise.

Like us, Yule concluded that the allele frequencies in his imaginary population were in equilibrium. Yule's conclusion was both groundbreaking and correct, but he took it a bit too literally. He had worked only one example, and he believed that allele frequencies of 0.5 and 0.5 represented the only possible equilibrium state for a two-allele system. For example, Yule believed that if a

single copy of allele A appeared as a mutation in a population whose gene pool otherwise contained only copies of a, then the A allele would automatically increase in frequency until copies of it constituted one-half of the gene pool. Yule argued this claim during the discussion that followed a talk given in 1908 by none other than Reginald Punnett. Punnett thought that Yule was wrong, but he did not know how to prove it.

We have already demonstrated, of course, that Punnett was correct in rejecting Yule's claim. Our calculations showed that a population with allele frequencies of 0.6 and 0.4 is in equilibrium too. What Punnett wanted, however, is a general proof. This proof should show that any allele frequencies, so long as they sum to 1, will remain unchanged from one generation to the next.

Punnett took the problem to his mathematician friend Godfrey H. Hardy, who produced the proof in short order (Hardy 1908). Hardy simply repeated the calculations that Yule had performed, using variables in place of the specific allele frequencies that Yule had assumed. Hardy's calculation of the general case indeed showed that any allele frequencies can be in equilibrium.

The General Case

For our version of Hardy's general case, we again work with our imaginary mouse population. We are concerned with a single locus with two alleles: A_1 and A_2. We use capital letters with subscripts because we want our calculation to cover cases in which the alleles are codominant as well as cases in which they are dominant and recessive. The three possible diploid genotypes are A_1A_1, A_1A_2, and A_2A_2.

As in our simulations and numerical example, we will start with the gene pool and follow the alleles through one complete turn of the life cycle. The gene pool will contain some frequency of A_1 gametes and some frequency of A_2 gametes. We will call the frequency of A_1 in the gene pool p and the frequency of A_2 in the gene pool q. There are only two alleles in the population, so

$$p + q = 1$$

The first step is to let the gametes in the gene pool combine to make zygotes. Figure 6.9a shows the four possible combinations of egg and sperm, the zygotes they produce, and a calculation specifying the probability of each. For example, if we pick an egg to watch at random, the chance is p that it will have genotype A_1. When a sperm comes along to fertilize the egg, the chance is p that the sperm will have genotype A_1. The probability that we will witness the production of an A_1A_1 zygote is therefore

$$p \times p = p^2$$

If we watched the formation of all the zygotes, p^2 of them would have genotype A_1A_1. The calculations in Figure 6.9a show that random mating in our gene pool produces zygotes in the following proportions:

A_1A_1	A_1A_2	A_2A_2
p^2	$2pq$	q^2

Figure 6.9b shows a Punnett square that yields the same genotype frequencies. The Punnett square also shows geometrically that

$$p^2 + 2pq + q^2 = 1$$

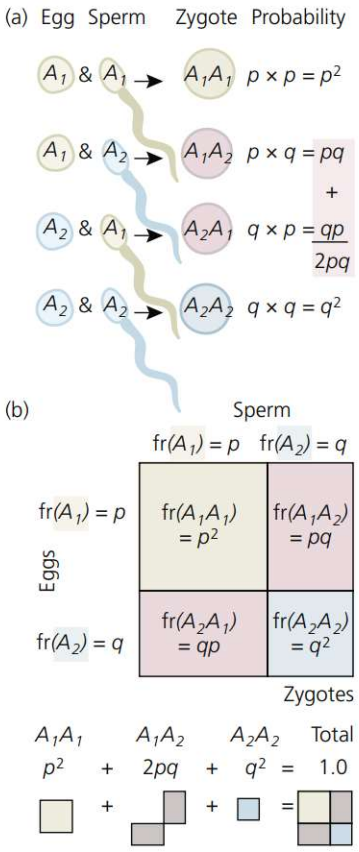


Figure 6.9 The general case for random mating in our model population (a) We can predict the genotype frequencies among the zygotes by multiplying the allele frequencies. (b) A Punnett square. The variables along the left and top edges of the box represent the frequencies of *A* and *a* eggs and sperm in the gene pool. The expressions inside the box represent the genotype frequencies among zygotes formed by random encounters between gametes in the gene pool.

This confirms that we have accounted for all the zygotes. The same result can be demonstrated algebraically by substituting $(1 - p)$ for q in the expression $p^2 + 2pq + q^2$, then simplifying.

The challenge now is to prove algebraically that there was nothing special about our numerical examples. Any allele frequencies will remain constant from generation to generation.

We have gone from the allele frequencies in the gene pool to the genotype frequencies among the zygotes. We now let the zygotes develop into juveniles, let the juveniles grow up to become adults, and let the adults produce gametes to make the next generation's gene pool.

We can calculate the frequency of allele A_1 in the new gene pool as follows. Because adults of genotype A_1A_1 constitute a proportion p^2 of the population, they will make p^2 of the gametes. All of these gametes carry copies of allele A_1. Likewise, adults of genotype A_1A_2 constitute a proportion $2pq$ of the population, and will make $2pq$ of the gametes. Half of these gametes carry copies of allele A_1. So the total fraction of the gametes in the gene pool that carry copies of A_1 is

$$p^2 + (\tfrac{1}{2})2pq = p^2 + pq$$

We can simplify the expression on the right by substituting $(1 - p)$ for q. This gives

$$\begin{aligned} p^2 + pq &= p^2 + p(1 - p) \\ &= p^2 + p - p^2 \\ &= p \end{aligned}$$

Figure 6.10 shows this calculation graphically. The figure also shows a calculation establishing that the fraction of the gametes in the gene pool that carry copies of allele A_2 is q. We assumed at the outset that p and q sum to 1, so we know that we have accounted for all the gametes.

Once again, we have come full circle and are back where we started. We started with allele frequencies of p and q in our population's gene pool. We followed the alleles through zygotes and adults and into the next generation's gene pool. The allele frequencies in the new gene pool are still p and q. The allele frequencies p and q can be stable at any values at all between 0 and 1, as long as they sum to 1. In other words, any allele frequencies can be in equilibrium, not just $p = q = 0.5$ as Yule thought.

Our model has shown that our idealized population does not evolve. This conclusion is known as the Hardy–Weinberg equilibrium principle.

This is a profound result. At the beginning of the chapter we defined evolution as change in allele frequencies in populations. The calculations we just performed show, given simple assumptions, that in populations following the rules of Mendelian genetics, allele frequencies do not change.

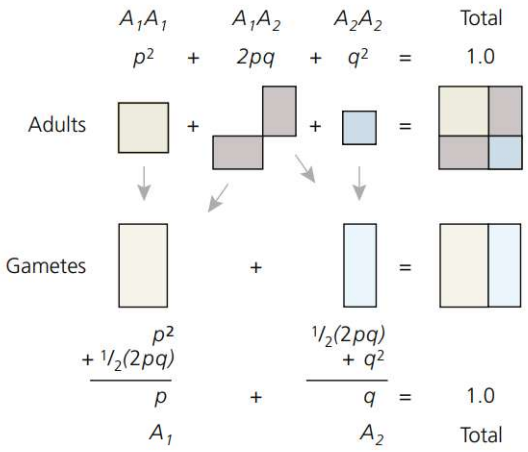


Figure 6.10 A geometrical representation of the general case for the allele frequencies produced when the adults in our model population make gametes The area of each box represents the frequency of an adult or gamete genotype.

COMPUTING CONSEQUENCES 6.2

The Hardy–Weinberg equilibrium principle with more than two alleles

Imagine a single locus with several alleles. We can call the alleles A_i, A_j, A_k, and so on, and we can represent the frequencies of the alleles in the gene pool with the variables p_i, p_j, p_k, and so on. The formation of a zygote with genotype A_iA_i requires the union of an A_i egg with an A_i sperm. Thus, the frequency of the homozygous genotype A_iA_i is p_i^2. The formation of a zygote with genotype A_iA_j requires either the union of an A_i egg with an A_j sperm, or an A_j egg with an A_i sperm. Thus, the frequency of the heterozygous genotype A_iA_j is $2p_ip_j$.

For example, if there are three alleles with frequencies p_1, p_2, and p_3 such that

$$p_1 + p_2 + p_3 = 1$$

then the genotype frequencies are given by

$$(p_1 + p_2 + p_3)^2 = p_1^2 + p_2^2 + p_3^2 + 2p_1p_2 + 2p_1p_3 + 2p_2p_3$$

and the allele frequencies do not change from generation to generation.

We have presented this result as the work of Hardy (1908). It was derived independently by Wilhelm Weinberg (1908) and has become known as the Hardy–Weinberg equilibrium principle. Some evolutionary biologists refer to it as the Hardy–Weinberg–Castle equilibrium principle, because William Castle (1903) worked a numerical example and stated the general equilibrium principle nonmathematically five years before Hardy and Weinberg explicitly proved the general case (see Provine 1971). The Hardy–Weinberg equilibrium principle yields two fundamental conclusions:

- **Conclusion 1:** The allele frequencies in a population will not change, generation after generation.
- **Conclusion 2:** If the allele frequencies in a population are given by p and q, the genotype frequencies will be given by p^2, $2pq$, and q^2.

We get an analogous result if we generalize the analysis from the two-allele case to the usual case of a population containing many alleles at a locus (see **Computing Consequences 6.2**).

What Use Is the Hardy–Weinberg Equilibrium Principle?

It may seem puzzling that in a book about evolution we have devoted so much space to a proof apparently showing that evolution does not happen. Evolution does, of course, happen—we saw it happen in this chapter in our own simulations. What makes the Hardy–Weinberg equilibrium principle useful is that it rests on a specific set of simple assumptions. When one or more of these assumptions is violated, the Hardy–Weinberg conclusions no longer hold.

We left some of the assumptions unstated when we developed our null model of Mendelian alleles in populations. We can now make them explicit. The crucial assumptions are as follows:

1. There is no selection. All members of our model population survived at equal rates and contributed equal numbers of gametes to the gene pool. When this assumption is violated—when individuals with some genotypes survive and

reproduce at higher rates than others—the frequencies of alleles may change from one generation to the next.

2. There is no mutation. In the model population, no copies of existing alleles were converted by mutation into copies of other existing alleles, and no new alleles were created. When this assumption is violated, and, for example, some alleles have higher mutation rates than others, allele frequencies may change from one generation to the next.

3. There is no migration. No individuals moved into or out of the model population. When this assumption is violated, and individuals carrying some alleles move into or out of the population at higher rates than individuals carrying other alleles, allele frequencies may change from one generation to the next.

4. There are no chance events that cause individuals with some genotypes to pass more of their alleles to the next generation than others. That is, blind luck plays no role. We saw the influence of blind luck in our simulations. We avoided it in our analysis of the general case by assuming that the eggs and sperm in the gene pool collided with each other at their actual frequencies of p and q and that no deviations were caused by chance. Another way to state this assumption is that the model population was infinitely large. When this assumption is violated, and by chance some individuals contribute more alleles to the next generation than others, allele frequencies may change from one generation to the next. This mechanism of allele frequency change is called, as we said earlier, genetic drift.

5. Individuals choose their mates at random. We explicitly set up the gene pool to let gametes find each other at random. In contrast to assumptions 1 through 4, when this assumption is violated—when, for example, individuals prefer to mate with other individuals of the same genotype—allele frequencies do not change from one generation to the next. Genotype frequencies may change, however. Such shifts in genotype frequency, in combination with a violation of one of the other four assumptions, can influence the evolution of populations.

By furnishing a list of ideal conditions under which populations will not evolve, the Hardy–Weinberg equilibrium principle identifies the set of events that can cause evolution in the real world (Figure 6.11). This is how the Hardy–Weinberg principle serves as a null model. Biologists can measure allele and genotype frequencies in nature, and determine whether the Hardy–Weinberg conclusions

The Hardy–Weinberg equilibrium principle becomes useful when we list the assumptions we made about our idealized population. By providing a set of explicit conditions under which evolution does not happen, the Hardy–Weinberg analysis identifies the mechanisms that can cause evolution in real populations.

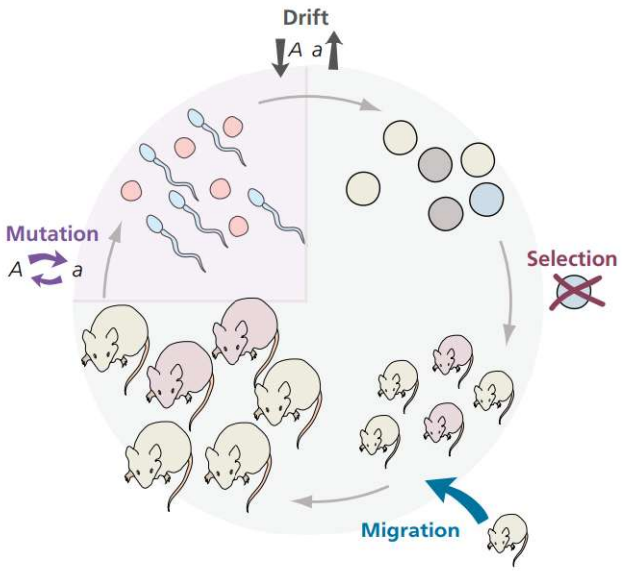


Figure 6.11 Summary of the mechanisms of evolution Four processes can cause allele frequencies to change from one generation to the next. Selection occurs when individuals with different genotypes survive or make gametes at different rates. Migration occurs when individuals move into or out of the population. Mutation occurs when mistakes during meiosis turn copies of one allele into copies of another. Genetic drift occurs when blind chance allows gametes with some genotypes to participate in more fertilizations than gametes with other genotypes.

hold. A population in which they hold is said to be in **Hardy–Weinberg equilibrium**. If a population is not in Hardy–Weinberg equilibrium—if the allele frequencies change from generation to generation or if the genotype frequencies cannot, in fact, be predicted by multiplying the allele frequencies—then one or more of the Hardy–Weinberg model's assumptions are being violated. Such a discovery does not, by itself, tell us which assumptions are being violated, but it indicates that further research may be rewarded with interesting discoveries.

In the remaining sections of Chapter 6, we consider how violations of assumptions 1 and 2 affect the two Hardy–Weinberg conclusions, and we explore empirical research on selection and mutation as mechanisms of evolution. (In Chapter 7, we consider violations of assumptions 3, 4, and 5.)

Changes in the Frequency of the *CCR5-Δ32* Allele

We began this section by asking whether we can expect the frequency of the *CCR5-Δ32* allele to change in human populations. Now that we have developed a null model for how Mendelian alleles behave in populations, we can give a partial answer. As long as individuals of all CCR5 genotypes survive and reproduce at equal rates, as long as no mutations convert some CCR5 alleles into others, as long as no one moves from one population to another, as long as populations are infinitely large, and as long as people choose their mates at random, then no, the frequency of the *CCR5-Δ32* allele will not change.

This answer is, of course, thoroughly unsatisfying. It is unsatisfying because none of the assumptions will be true in any real population. We asked the question in the first place precisely because we expect *Δ32/Δ32* individuals to survive the AIDS epidemic at higher rates than individuals with either of the other two genotypes. In the next two sections, we will see that our null model, the Hardy–Weinberg equilibrium principle, provides a framework that allows us to assess with precision the importance of differences in survival.

6.2 Selection

Our analysis in Section 6.1 was motivated by a desire to predict whether the frequency of the *CCR5-Δ32* allele will change as a result of the AIDS epidemic. We started by scaling Mendelian genetics up from single crosses to whole populations. This is the first step in integrating Mendelism with Darwin's theory of evolution by natural selection. The next step is to add differences in survival and reproductive success. Doing so makes the algebra a bit more complicated. But it also lets us glimpse the predictive strength of population genetics.

First on the list of assumptions about our idealized population was that individuals survive at equal rates and have equal reproductive success. We now explore what happens to allele frequencies when this assumption is violated.

In the model we used to derive the Hardy–Weinberg equilibrium principle, first on our list of assumptions was that all individuals survive at equal rates and contribute equal numbers of gametes to the gene pool. Systematic violations of this assumption are examples of **selection**. Selection happens when individuals with particular phenotypes survive to sexual maturity at higher rates than those with other phenotypes, or when individuals with particular phenotypes produce more offspring during reproduction than those with other phenotypes. The bottom line in either kind of selection is differential reproductive success. Some individuals have more offspring than others. Selection can lead to evolution when the phenotypes that exhibit differences in reproductive success are heritable—that is, when certain phenotypes are associated with certain genotypes.

Population geneticists often assume that phenotypes are determined strictly by genotypes. They might, for example, think of pea plants as being either tall or short, such that individuals with the genotypes *TT* and *Tt* are tall and individuals with the genotype *tt* are short. Such a view is at least roughly accurate for some traits, including the examples we use in this chapter.

When phenotypes fall into discrete classes that appear to be determined strictly by genotypes, we can think of selection as if it acts directly on the genotypes. We can then assign a particular level of lifetime reproductive success to each genotype. In reality, most phenotypic traits are not, in fact, strictly determined by genotype. Pea plants with the genotype *TT*, for example, vary in height. This variation is due to genetic differences at other loci and to differences in the environments where the pea plants grew. We will consider such complications elsewhere (see Chapter 9). For the present, however, we adopt the simple view.

When we think of selection as if it acts directly on genotypes, its defining feature is that some genotypes contribute more alleles to future generations than others. In other words, there are differences among genotypes in fitness.

Our task in this section is to incorporate selection into the Hardy–Weinberg analysis. We begin by asking whether selection can change the frequencies of alleles in the gene pool from one generation to the next. In other words, can violation of the no-selection assumption lead to a violation of conclusion 1 of the Hardy–Weinberg equilibrium principle?

Adding Selection to the Hardy–Weinberg Analysis: Changes in Allele Frequencies

We start with a numerical example showing that selection can, indeed, change the frequencies of alleles. Imagine that in our population of mice there is a locus, the B locus, that affects the probability of survival. Assume that the frequency of allele B_1 in the gene pool is 0.6 and the frequency of allele B_2 is 0.4 **(Figure 6.12)**. After random mating, we get genotype frequencies for B_1B_1, B_1B_2, and B_2B_2 of 0.36, 0.48, and 0.16. The rest of our calculations will be simpler if we give the population of zygotes a finite size, so imagine that there are 100 zygotes:

B_1B_1	B_1B_2	B_2B_2
36	48	16

These zygotes are represented by a bar graph on the upper right in the figure. We will follow the individuals that develop from these zygotes. Those that survive to adulthood will breed to produce the next generation's gene pool.

We incorporate selection by stipulating that the genotypes differ in survival. All of the B_1B_1 individuals survive, 75% of the B_1B_2 individuals survive, and 50% of the B_2B_2 individuals survive. As shown in Figure 6.12, we now have 80 adults:

B_1B_1	B_1B_2	B_2B_2
36	36	8

A numerical example shows that when individuals with some genotypes survive at higher rates than individuals with other genotypes, allele frequencies can change from one generation to the next. In other words, our model shows that natural selection causes evolution.

If we assume that each survivor donates 10 gametes to the new gene pool, then

The 36 B_1B_1 adults together make 360 gametes: 360 carry B_1; none carry B_2.
The 36 B_1B_2 adults together make 360 gametes: 180 carry B_1; 180 carry B_2.
The 8 B_2B_2 adults together make 80 gametes: none carry B_1; 80 carry B_2.

Summing the gametes carrying copies of each allele, we get 540 carrying copies of B_1 and 260 carrying copies of B_2, for a total of 800. The frequency of allele B_1

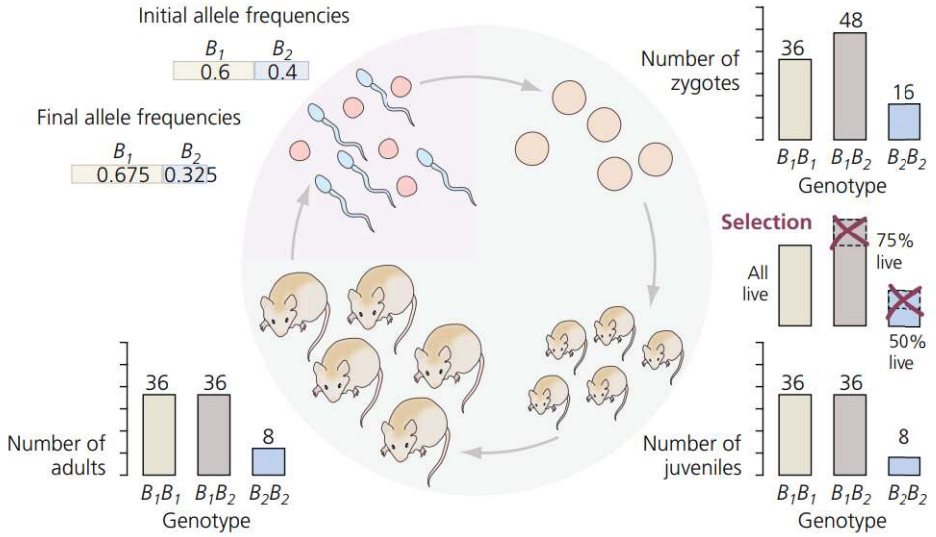


Figure 6.12 Selection can cause allele frequencies to change across generations This figure follows our model mouse population from one generation's gene pool to the next generation's gene pool. The bar graphs show the number of individuals of each genotype in the population at any given time. Selection, in the form of differences in survival among juveniles, causes the frequency of allele B_1 to increase.

in the new gene pool is $\frac{540}{800} = 0.675$; the frequency of allele B_2 is $\frac{260}{800} = 0.325$. The frequency of allele B_1 has risen by an increment of 7.5 percentage points. The frequency of allele B_2 has dropped by the same amount.

Violation of the no-selection assumption has resulted in violation of conclusion 1 of the Hardy–Weinberg analysis. The population has evolved.

We used strong selection to make a point in our numerical example. Rarely in nature are differences in survival rates large enough to cause such dramatic change in allele frequencies in a single generation. If selection continues for many generations, however, even small changes in allele frequency in each generation can add up to substantial changes over the long run.

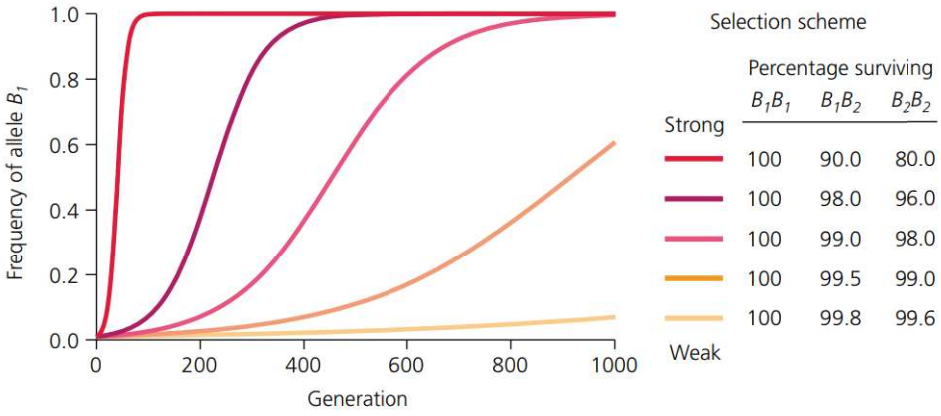


Figure 6.13 Persistent selection can produce substantial changes in allele frequencies over time Each curve shows the change in allele frequency over time under a particular selection intensity.

Figure 6.13 illustrates the cumulative change in allele frequencies that can be wrought by selection. The figure is based on a model population similar to the one we used in the preceding numerical example, except that the initial allele frequencies are 0.01 for B_1 and 0.99 for B_2. The red line shows the change in allele frequencies when the survival rates are 100% for B_1B_1, 90% for B_1B_2, and 80% for B_2B_2. The frequency of allele B_1 rises from 0.01 to 0.99 in less than 100 generations. Under weaker selection schemes, the frequency of B_1 rises more slowly, but still inexorably. (See **Computing Consequences 6.3** for a general algebraic treatment incorporating selection into the Hardy–Weinberg analysis.)

COMPUTING CONSEQUENCES 6.3

A general treatment of selection

Here we develop equations that predict allele frequencies in the next generation, given allele frequencies in this generation and fitnesses for the genotypes. We start with a gene pool in which allele A_1 is at frequency p and allele A_2 is at frequency q. We allow gametes to pair at random to make zygotes of genotypes A_1A_1, A_1A_2, and A_2A_2 at frequencies p^2, $2pq$, and q^2, respectively. We incorporate selection by imagining that A_1A_1 zygotes survive to adulthood at rate w_{11}, A_1A_2 zygotes survive at rate w_{12}, and A_2A_2 zygotes survive at rate w_{22}. All survivors produce the same number of offspring. Therefore, a genotype's survival rate is proportional to the genotype's lifetime reproductive success, or fitness. We thus refer to the survival rates as fitnesses. The average fitness for the whole population, $\overline{w}$, is given by

$$\overline{w} = p^2w_{11} + 2pqw_{12} + q^2w_{22}$$

[To see this, note that we can calculate the average of the numbers 1, 2, 2, and 3 as $\frac{(1 + 2 + 2 + 3)}{4}$ or as $(\frac{1}{4} \times 1) + (\frac{1}{2} \times 2) + (\frac{1}{4} \times 3)$. Our expression for the average fitness is of the second form: We multiply the fitness of each genotype by its frequency in the population and sum the results.]

We now calculate the genotype frequencies among the surviving adults (right before they make gametes). The new frequencies of the genotypes are

A_1A_1	A_1A_2	A_2A_2
$\frac{p^2w_{11}}{\overline{w}}$	$\frac{2pqw_{12}}{\overline{w}}$	$\frac{q^2w_{22}}{\overline{w}}$

(We have to divide by the average fitness in each case to ensure that the new frequencies still sum to 1.)

Finally, we let the adults breed, and calculate the allele frequencies in the new gene pool:

- For the A_1 allele: A_1A_1 individuals contribute $\frac{p^2w_{11}}{\overline{w}}$ of the gametes, all of them A_1, and A_1A_2 individuals contribute $\frac{2pqw_{12}}{\overline{w}}$ of the gametes, half of them A_1. The new frequency of A_1 is thus

$$\frac{p^2w_{11} + pqw_{12}}{\overline{w}}$$

- For the A_2 allele: A_1A_2 individuals contribute $\frac{2pqw_{12}}{\overline{w}}$ of the gametes, half of them A_2; A_2A_2 individuals contribute $\frac{q^2w_{22}}{\overline{w}}$ of the gametes, all of them A_2. So the new frequency of A_2 is

$$\frac{pqw_{12} + q^2w_{22}}{\overline{w}}$$

The reader should confirm that the new frequencies of A_1 and A_2 sum to 1.

It is instructive to calculate the change in the frequency of allele A_1 from one generation to the next. This value, Δp, is the new frequency of A_1 minus the old frequency of A_1:

$$\begin{aligned}\Delta p &= \frac{p^2w_{11} + pqw_{12}}{\overline{w}} - p \\ &= \frac{p^2w_{11} + pqw_{12}}{\overline{w}} - \frac{p\overline{w}}{\overline{w}} \\ &= \frac{p^2w_{11} + pqw_{12} - p\overline{w}}{\overline{w}} \\ &= \frac{p}{\overline{w}}(pw_{11} + qw_{12} - \overline{w})\end{aligned}$$

The final expression is a useful one, because it shows that the change in frequency of allele A_1 is proportional to $(pw_{11} + qw_{12} - \overline{w})$. The quantity $(pw_{11} + qw_{12} - \overline{w})$ is sometimes called the **average excess** of allele A_1. It is equal to the average fitness of allele A_1 when paired at random with other alleles $(pw_{11} + qw_{12})$ minus the average fitness of the population $(\overline{w})$. When the average excess of allele A_1 is positive, A_1 will increase in frequency. In other words, if the average A_1-carrying individual has higher-than-average fitness, then the frequency of allele A_1 will rise.

The change in the frequency of allele A_2 from one generation to the next is

$$\begin{aligned}\Delta q &= \frac{pqw_{12} + q^2w_{22}}{\overline{w}} - q \\ &= \frac{q}{\overline{w}}(pw_{12} + qw_{22} - \overline{w})\end{aligned}$$

Empirical Research on Allele Frequency Change by Selection

Douglas Cavener and Michael Clegg (1981) documented a cumulative change in allele frequencies over many generations in a laboratory-based natural selection experiment on the fruit fly (*Drosophila melanogaster*). Fruit flies, like most other animals, make an enzyme that breaks down ethanol, the poisonous active ingredient in beer, wine, and rotting fruit. This enzyme is called alcohol dehydrogenase, or ADH. Cavener and Clegg worked with populations of flies that had two alleles at the ADH locus: Adh^F and Adh^S. (The *F* and *S* refer to whether the protein encoded by the allele moves quickly or slowly through an electrophoresis gel.)

The scientists kept two experimental populations on food spiked with ethanol and two control populations of flies on normal, nonspiked food. The researchers picked the breeders for each generation at random. This is why we are calling the project a natural selection experiment: Cavener and Clegg set up different environments for their different populations, but the researchers did not themselves directly manipulate the survival or reproductive success of individual flies.

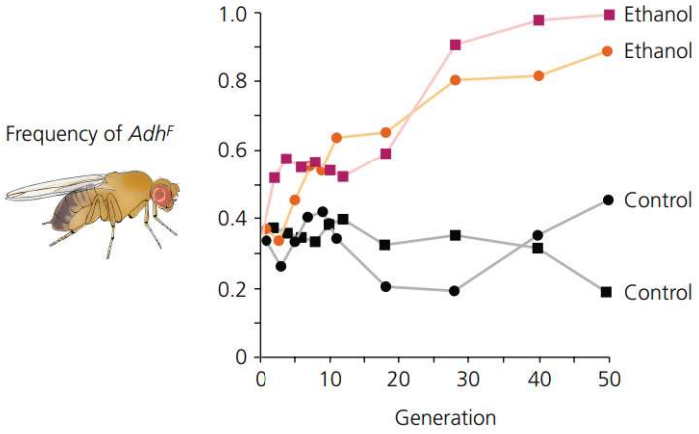


Figure 6.14 Frequencies of the allele in four populations of fruit flies over 50 generations The black squares and circles represent control populations living on normal food; the magenta squares and orange circles represent experimental populations living on food spiked with ethanol. From Cavener and Clegg (1981).

Every several generations, Cavener and Clegg took a random sample of flies from each population, determined their ADH genotypes, and calculated the allele frequencies. The results appear in Figure 6.14. The control populations showed no large or consistent long-term change in the frequency of the Adh^F allele. The experimental populations, in contrast, showed a rapid and largely consistent increase in the frequency of Adh^F (and, of course, a corresponding decrease in the frequency of Adh^S). Hardy–Weinberg conclusion 1 appears to hold true in the control populations, but is clearly not valid in the experimental populations.

Empirical research on fruit flies is consistent with our conclusion that natural selection can cause allele frequencies to change.

Can we identify for certain which of the assumptions of the Hardy–Weinberg analysis is being violated? The only difference between the two kinds of populations is that the experimentals have ethanol in their food. This suggests that it is the no-selection assumption that is being violated in the experimental populations. Flies carrying the Adh^F allele appear to have higher lifetime reproductive success (higher fitness) than flies carrying the Adh^S allele when ethanol is present in the food. Cavener and Clegg note that this outcome is consistent with the fact that alcohol dehydrogenase extracted from Adh^F homozygotes breaks down ethanol at twice the rate of alcohol dehydrogenase extracted from Adh^S homozygotes. Whether flies with the Adh^F allele have higher fitness because they have higher rates of survival or because they produce more offspring is unclear.

Adding Selection to the Hardy–Weinberg Analysis: The Calculation of Genotype Frequencies

The calculations and example we have just discussed show that selection can cause allele frequencies to change across generations. Selection invalidates conclusion 1 of the Hardy–Weinberg analysis. We now consider how selection affects conclusion 2 of the Hardy–Weinberg analysis. In a population under selection, can we still calculate the genotype frequencies by multiplying the allele frequencies?

Often, we cannot. As before, we use a population with two alleles at a locus affecting survival: B_1 and B_2. We assume that the initial frequency of each allele in the gene pool is 0.5 **(Figure 6.15)**. After random mating, we get genotype frequencies for B_1B_1, B_1B_2, and B_2B_2 of 0.25, 0.5, and 0.25. The rest of our calculations will be simpler if we give the population of zygotes a finite size, so imagine there are 100 zygotes:

B_1B_1	B_1B_2	B_2B_2
25	50	25

These zygotes are represented by a bar graph on the upper right in the figure. We will follow the individuals that develop from these zygotes. Those that survive to adulthood will breed to produce the next generation's gene pool.

As in our first selection example, we incorporate selection by stipulating that the genotypes differ in their rates of survival. This time, 60% of the B_1B_1 individuals survive, all of the B_1B_2 individuals survive, and 60% of the B_2B_2 individuals survive. As shown in Figure 6.15, there are now 80 adults in the mouse population:

B_1B_1	B_1B_2	B_2B_2
15	50	15

If we assume that each surviving adult donates 10 gametes to the next generation's gene pool, then

The 15 B_1B_1 adults together make 150 gametes: 150 carry B_1; none carry B_2.

The 50 B_1B_2 adults together make 500 gametes: 250 carry B_1; 250 carry B_2.

The 15 B_2B_2 adults together make 150 gametes: none carry B_1; 150 carry B_2.

Summing the gametes carrying each allele, we get 400 carrying B_1 and 400 carrying B_2, for a total of 800. Both alleles are still at a frequency of 0.5. Despite strong selection against homozygotes, the frequencies of the alleles have not changed; the population has not evolved.

But let us calculate frequencies of the three genotypes among the surviving adults. These frequencies are as follows:

B_1B_1	B_1B_2	B_2B_2
$\frac{15}{80} = 0.1875$	$\frac{50}{80} = 0.625$	$\frac{15}{80} = 0.1875$

These genotype frequencies reveal that violation of the no-selection assumption has resulted in violation of conclusion 2 of the Hardy–Weinberg analysis. We can no longer calculate the genotype frequencies among the adult survivors by multiplying the frequencies of the alleles. For example:

Frequency of B_1B_1		$(\textit{Frequency of } B_1)^2$
0.1875	$\neq$	$(0.5)^2 = 0.25$

Natural selection can also drive genotype frequencies away from the values predicted under the Hardy–Weinberg equilibrium principle.

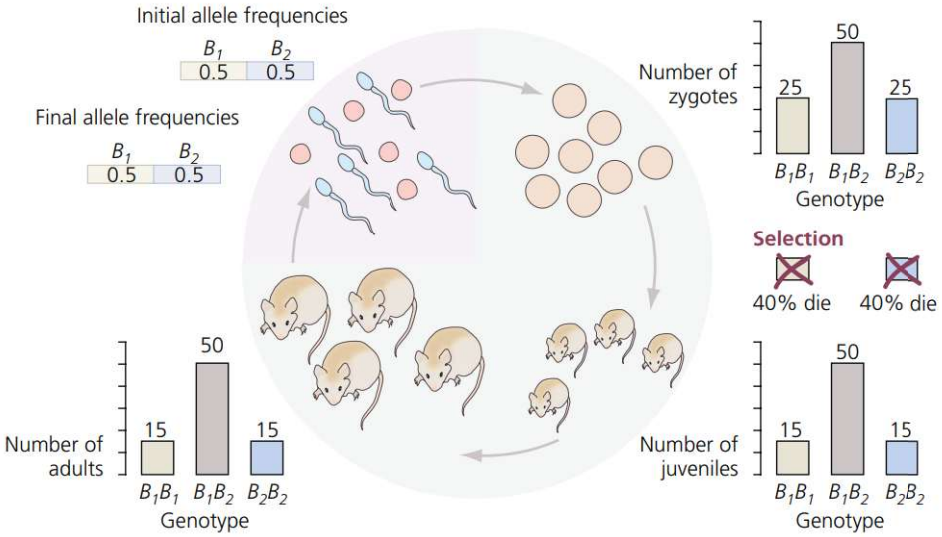


Figure 6.15 Selection can change genotype frequencies so that they cannot be calculated by multiplying the allele frequencies When 40% of the homozygotes in this population die, the allele frequencies do not change. But among the survivors, there are more heterozygotes than predicted under Hardy–Weinberg equilibrium.

We used strong selection in our numerical example to make a point. In fact, selection is rarely strong enough to produce, in a single generation, such a large violation of Hardy–Weinberg conclusion 2. Even if it does, a single bout of random mating will immediately put the genotypes back into Hardy–Weinberg equilibrium. Nonetheless, researchers sometimes find violations of Hardy–Weinberg conclusion 2 that seem to be the result of selection.

Empirical Research on Selection and Genotype Frequencies

Our example comes from research by Atis Muehlenbachs and colleagues (2008), working in the laboratory of Patrick Duffy, on genetic variation for the outcome of falciparum malaria during pregnancy. Falciparum malaria is caused by infection with the single-celled parasite *Plasmodium falciparum*. When a pregnant woman contracts the disease, the parasites invade the placenta via the mother's circulatory system (Karumanchi and Haig 2008). This triggers placental inflammation and may also interfere with placental development (Umbers et al. 2011). The potential complications include spontaneous abortion, premature delivery, low birth weight, and higher risk of infant death.

Pregnancy itself brings an increased risk of malaria infection, particularly a woman's first pregnancy (Karumanchi and Haig 2008). During a first bout of placental malaria, women develop antibodies that confer partial resistance during later pregnancies. Some 125 million women who live in areas affected by malaria become pregnant each year, and malaria infection during pregnancy is estimated to be responsible for an annual toll of 100,000 infant deaths (Umbers et al. 2011).

Muehlenbachs and colleagues (2008) suspected that the outcome of placental malaria hinges on the fetus's genotype at the locus encoding vascular endothelial growth factor receptor 1 (VEGFR1), also known as *fms*-like tyrosine kinase 1 (Flt1). Fetal cells in the placenta release a soluble form of this protein, sVEGFR1, into the mother's circulation. By interacting with vascular endothelial growth factor, VEGFR1 influences both placental development and inflammation.

Copies of the gene for VEGFR1 vary in the length of a two-nucleotide repeat in a region that is transcribed to mRNA but not translated. Alleles cluster into a short group (*S* alleles) and a long group (*L* alleles). Cultured cord blood cells with genotypes *SS* and *SL* produce more VEGFR1 than do *LL* cells.

COMPUTING CONSEQUENCES 6.4

Statistical analysis of allele and genotype frequencies using the χ^2 (chi-square) test

Here we use data from Muehlenbachs and colleagues (2008) to illustrate a method for determining whether genotype frequencies deviate from Hardy–Weinberg equilibrium. The researchers surveyed Tanzanian infants born to first-time mothers during malaria season. The genotype counts (provided by Atis Muehlenbachs and Patrick Duffy, personal communication) were

SS	*SL*	*LL*
16	50	10

The analysis proceeds in five steps:

1. Calculate the allele frequencies. The sample of 76 infants is also a sample of 152 gene copies. All 32 copies carried by the *SS* infants are *S*, as are 50 of the copies carried by the *SL* infants. Thus, the frequency of *S* is

$$\frac{32 + 50}{152} = 0.54$$

The frequency of *L* is

$$\frac{50 + 20}{152} = 0.46$$

2. Calculate the genotype frequencies expected under Hardy–Weinberg equilibrium. If the frequencies of two alleles are p and q, then the expected frequencies of the genotypes are p^2, $2pq$, and q^2. The expected frequencies among the infants are thus

SS	*SL*	*LL*
$0.54^2 = 0.29$	$2 \cdot 0.54 \cdot 0.46 = 0.5$	$0.46^2 = 0.21$

3. Calculate the expected number of infants of each genotype under Hardy–Weinberg equilibrium. This is simply the expected frequency of each genotype multiplied by the total number of infants, 76:

SS	*SL*	*LL*
$0.29 \cdot 76 = 22$	$0.5 \cdot 76 = 38$	$0.21 \cdot 76 = 16$

These expectations are different from the numbers observed (16, 50, and 10). The actual sample contains more heterozygotes and fewer homozygotes. Is it plausible that a difference this large between expectation and reality could arise by chance? Or is the difference statistically significant? Our null hypothesis is that the difference is simply due to chance.

Working with newborn babies of first-time mothers in Muheza, Tanzania, where malaria is a perennial scourge, Muehlenbachs and colleagues (2008) tested their hypothesis in part by using the Hardy–Weinberg equilibrium principle.

The researchers first determined the allele frequencies among 163 infants born from October through April, when the rate of placental malaria was at its annual low. The frequencies were

S	*L*
0.555	0.445

If the population of infants was in Hardy–Weinberg equilibrium, then multiplying these allele frequencies will allow us to predict the genotype frequencies:

SS	*SL*	*LL*
$0.555^2 = 0.308$	$2 \cdot 0.555 \cdot 0.445 = 0.494$	$0.445^2 = 0.198$

These predicted frequencies are, in fact, close to the actual genotype frequencies among the off-season infants:

SS	*SL*	*LL*
$\frac{49}{163} = 0.301$	$\frac{83}{163} = 0.509$	$\frac{31}{163} = 0.190$

The true frequency of heterozygotes is slightly higher than predicted, and the frequencies of homozygotes are slightly lower, but the discrepancies are modest. The infants thus conform to conclusion 2 of the Hardy–Weinberg analysis.

4. Calculate a test statistic. We will use one devised in 1900 by Karl Pearson. It is called chi-square (χ^2).

$$\chi^2 = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$$

where the symbol $\sum$ indicates a sum taken across all the classes considered. In our data there are three classes: the three genotypes. For our data set

$$\chi^2 = \frac{(16 - 22)^2}{22} + \frac{(50 - 38)^2}{38} + \frac{(10 - 16)^2}{16} = 7.68$$

5. Determine whether the test statistic is significant. χ^2 is defined such that it gets larger as the difference between the observed and expected values gets larger. How likely is it that we could get a χ^2 as large as 7.68 by chance? Most statistical textbooks have a table giving the answer. In Zar's (1996) book. it is called "Critical values of the chi-square distribution."

To use this table, we need to calculate a number called the degrees of freedom for the test statistic. This value for χ^2 is the number of classes minus the number of independent values calculated from the data for use in determining the expected values. For our χ^2 there are three classes: the genotypes. We calculated two values from the data for use in determining the expected values: the total number of individuals, and the frequency of allele *S*. (We also calculated the frequency of *L*, but it is not independent of the frequency of *S*, because the two must sum to 1.) Thus the number of degrees of freedom is 1. (Another formula for calculating the degrees of freedom in χ^2 tests for Hardy–Weinberg equilibrium is

$$df = k - 1 - m$$

where k is the number of classes and m is the number of independent allele frequencies estimated.)

According to the table, the critical value of χ^2 for one degree of freedom and $P = 0.05$ is 3.841. This means there is a 5% chance under the null hypothesis of getting $\chi^2 \geq 3.841$. The probability under the null hypothesis of getting $\chi^2 \geq 7.68$ is therefore (considerably) less than 5%. We reject the null hypothesis and assert that our χ^2 is statistically significant at $P < 0.05$. (In fact, $P < 0.006$.)

The χ^2 test tells us that among infants born during malaria season, the alleles of the gene for VEGFR1 are not in Hardy–Weinberg equilibrium. This indicates that one or more assumptions of the Hardy–Weinberg analysis has been violated. By itself, however, it does not tell us which are being violated, or how.

Muehlenbachs and colleagues then determined the allele frequencies among 76 infants born from May through September, when the rate of placental malaria was at its annual high. The frequencies were nearly the same as among the off-season newborns:

S	*L*
0.539	0.461

If this segment of the population was, like their off-season counterparts, in Hardy–Weinberg equilibrium, then multiplying the allele frequencies will again allow us to predict the genotype frequencies:

SS	*SL*	*LL*
$0.539^2 = 0.291$	$2 \cdot 0.539 \cdot 0.461 = 0.497$	$0.461^2 = 0.213$

This time the predicted values are a poor fit to the actual frequencies:

SS	*SL*	*LL*
$\frac{16}{76} = 0.211$	$\frac{50}{76} = 0.658$	$\frac{10}{76} = 0.132$

There are substantially more heterozygotes than expected, and substantially fewer homozygotes. This discrepancy between prediction and data is statistically significant (see **Computing Consequences 6.4**). The genotypes of the infants born during peak malaria season are in violation of Hardy–Weinberg conclusion 2.

The discovery that genotype frequencies in a population are not in Hardy–Weinberg equilibrium may be a clue that natural selection is at work.

On this and other evidence, Muehlenbachs and colleagues (2008) believe the best explanation for the missing homozygotes is that they did not survive fetal development. A fetus's chance of surviving depends on both its own genotype and whether its mother contracts malaria (Figure 6.16). If the mother does not contract malaria, *SS* infants do somewhat better than others. If, however, the mother does contract malaria, *SL* infants do substantially better than others. Overall, when malaria is common, heterozygotes survive at the highest rate. Consistent with this explanation, where malaria is absent, *S* alleles occur at high frequency.

Figure 6.16 Probability of fetal survival as a function of genotype and placental malaria Inferred from the patterns in maternal and newborn genotype frequencies in Muehlenbachs et al. (2008).

Changes in the Frequency of the *CCR5-Δ32* Allele Revisited

We are now in a position to give a more satisfying answer to the question we raised at the beginning of Section 6.1: Will the AIDS epidemic cause the frequency of the *CCR5-Δ32* allele to increase in human populations? The AIDS epidemic could, in principle, cause the frequency of the allele to increase rapidly, but at present it appears that it will probably not do so in any real population. This conclusion is based on the three model populations depicted in Figure 6.17 (see **Computing Consequences 6.5** for the algebra). Each model is based on different assumptions about the initial frequency of the *CCR5-Δ32* allele and the prevalence of HIV infection. Each graph shows the predicted change in the frequency of the *Δ32* allele over 40 generations, or approximately 1,000 years.

The model population depicted in Figure 6.17a offers a scenario in which the frequency of the *Δ32* allele could increase rapidly. In this scenario, the initial frequency of the *CCR5-Δ32* allele is 20%. One-quarter of the individuals with genotype +/+ or +/*Δ32* contract AIDS and die without reproducing, whereas all of the *Δ32*/*Δ32* individuals survive. The 20% initial frequency of *Δ32* is approximately equal to the highest frequency reported for any population, a sample of Ashkenazi Jews studied by Martinson et al. (1997). The mortality rates approximate the situation in Botswana, Namibia, Swaziland, and Zimbabwe, where up to 25% of individuals between the ages of 15 and 49 are infected with HIV (UNAIDS 1998). In this model population, the frequency of the *Δ32* allele increases by as much as a few percentage points each generation. By the end of 40 generations, the allele is at a frequency of virtually 100%. Thus, in a human population that combined the highest reported frequency of the *Δ32* allele with the highest reported rates of infection, the AIDS epidemic could cause the frequency of the allele to increase rapidly.

At present, however, no known population combines a high frequency of the *Δ32* allele with a high rate of HIV infection. In northern Europe, many populations have *Δ32* frequencies between 0.1 and 0.2 (Martinson et al. 1997; Stephens et al. 1998), but HIV infection rates are under 1% (UNAIDS 1998). A model population reflecting these conditions is depicted in Figure 6.17b. The initial frequency of the *Δ32* allele is 0.2, and 0.5% of the +/+ and +/*Δ32* individuals contract AIDS and die without reproducing. The frequency of the *Δ32* allele hardly changes at all. Selection is too weak to cause appreciable evolution in such a short time.

In parts of sub-Saharan Africa, as many as a quarter of all individuals of reproductive age are infected with HIV. However, the *Δ32* allele is virtually absent (Martinson et al. 1997). A model population reflecting this situation is depicted in Figure 6.17c. The initial frequency of the *Δ32* allele is 0.01, and 25% of the +/+ and +/*Δ32* individuals contract AIDS and die without reproducing. Again, the frequency of the *Δ32* allele hardly changes at all. When the *Δ32* allele is at low

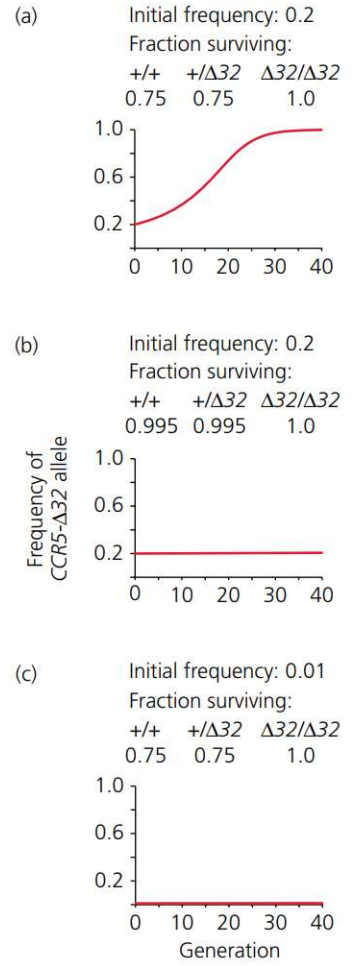


Figure 6.17 Predicted change in allele frequencies at the CCR5 locus under different scenarios (a) If the initial frequency of *CCR5-Δ32* is high and many people become infected with HIV, allele frequencies can change rapidly. (b) In Europe allele frequencies are high, but infection rates are low. (c) In parts of Africa the infection rates are high, but allele frequencies are low.

COMPUTING CONSEQUENCES 6.5

Predicting the frequency of the *CCR5-Δ32* allele in future generations

Let q_g be the frequency of the *CCR5-Δ32* allele in the present generation. Based on Computing Consequences 6.3, we can write an equation predicting the frequency of the allele in the next generation, given estimates of the survival rates (fitnesses) of individuals with each genotype. The equation is

$$q_{g+1} = \frac{(1 - q_g)q_g w_{+\Delta} + q_g^2 w_{\Delta\Delta}}{(1 - q_g)^2 w_{++} + 2(1 - q_g)q_g w_{+\Delta} + q_g^2 w_{\Delta\Delta}}$$

where q_{g+1} is the frequency of the *Δ32* allele in the next generation, w_{++} is the fitness of individuals homozygous for the normal allele, $w_{+\Delta}$ is the fitness of heterozygotes, and $w_{\Delta\Delta}$ is the fitness of individuals homozygous for the *CCR5-Δ32* allele.

After choosing a starting value for the frequency of the *Δ32* allele, we plug it and the estimated fitnesses into the equation to generate the frequency of the *Δ32* allele after one generation. We then plug this resulting value into the equation to get the frequency of the allele after two generations, and so on.

frequency, most copies are in heterozygotes. Because heterozygotes are susceptible to infection, these copies are hidden from selection.

The analysis we have just described is based on a number of simplifying assumptions. We have assumed, for example, that all HIV-infected individuals die without reproducing. In fact, however, many HIV-infected individuals have children. We have also assumed that the death rate is the same in heterozygotes as in +/+ homozygotes. In reality, although heterozygotes are susceptible to HIV infection, they appear to progress more slowly to AIDS (Dean et al. 1996). As a result, the fitness of heterozygotes may actually be higher than that of +/+ homozygotes. We challenge the reader to explore the evolution of human populations under a variety of selection schemes, to see how strongly our simplifying assumptions affect the predicted course of evolution. For analyses of more complex models of human evolution in response to selection imposed by AIDS, see models by Schliekelman et al. (2001; but also Ramaley et al. 2002), by Sullivan et al. (2001), and by Cromer et al. (2010).

Our exploration of natural selection has given us tools we can use to predict the future of human populations.

6.3 Patterns of Selection: Testing Predictions of Population Genetics Theory

In the 1927 case of *Buck v. Bell*, the United States Supreme Court upheld the state of Virginia's sterilization statute by a vote of eight to one. Drafted on the advice of eugenicists, the law was intended to improve the genetic quality of future generations by allowing the forced sterilization of individuals afflicted with hereditary forms of insanity, feeblemindedness, and other mental defects. The court's decision in *Buck v. Bell* reinvigorated a compulsory sterilization movement dating from 1907 (Kevles 1995). By 1940, 30 states had enacted sterilization laws, and by 1960 over 60,000 people had been sterilized without their consent (Reilly 1991; Lane 1992). In hindsight, the evidence that these individuals suffered from

hereditary diseases was weak. But what about the evolutionary logic behind compulsory sterilization? If the genetic assumptions had been correct, would sterilization have been an effective means of reducing the incidence of undesirable traits?

Before we try to answer this question, it will be helpful to address a more general one. How well does the theory of population genetics actually work? We developed this theory in Sections 6.1 and 6.2. The final product is a model of how allele frequencies change in response to natural selection (Figures 6.12 and 6.13, Computing Consequences 6.3 and 6.5). If our model is a good one, it should accurately predict the direction and rate of allele frequency change under a variety of selection schemes. It should work, for example, whether the allele favored by selection is dominant or recessive, common or rare. It should work whether selection favors heterozygotes or homozygotes. It should even predict what will happen when a particular allele is favored by selection under some circumstances and disfavored in others.

In this section, we will find out how well our model works. Using the theory we have developed to predict the course of evolution under different patterns of selection, we compare our predictions to empirical data from experimental populations. We then return to our question about the effectiveness of eugenic sterilization in changing the composition of populations.

Selection on Recessive and Dominant Alleles

For our first test, we focus on whether our theory accurately predicts changes in the frequencies of recessive and dominant alleles. Our example comes from the work of Peter Dawson (1970). Dawson had been studying a laboratory colony of flour beetles **(Figure 6.18)** and had identified a gene we will call the l locus. This locus has two alleles: + and *l*. Individuals with genotype +/+ or +/*l* are phenotypically normal, whereas individuals with genotype *l*/*l* do not survive. In other words, *l* is a recessive lethal allele.

Figure 6.18 Flour beetles, *Tribolium castaneum* Courtesy of Susan J. Brown, Professor/Kansas State University, Kansas.

Dawson collected heterozygotes from his beetle colony and used them to establish two new experimental populations. Because all the founders were heterozygotes, the initial frequencies of the two alleles were 0.5 in both populations. Because *l*/*l* individuals have zero fitness, Dawson expected his populations to evolve toward ever lower frequencies of the *l* allele and ever higher frequencies of the + allele. He let his two populations evolve for a dozen generations, each generation measuring the frequencies of the two alleles.

Dawson used the equations derived in Computing Consequences 6.3 and the method described in Computing Consequences 6.5 to make a quantitative prediction of the course of evolution in his populations. We can reproduce this prediction with a straightforward numerical calculation like the ones we performed in Figures 6.12 and 6.13. Imagine a gene pool in which alleles + and *l* are both at a frequency of 0.5. If we combine gametes at random to make 100 zygotes, we get the three genotypes in the following numbers:

+/+	+/*l*	*l*/*l*
25	50	25

Now we imagine that all the *l*/*l* individuals die and that everyone else survives to breed. Finally, imagine that each of the survivors donates 10 gametes to the new gene pool:

The 25 +/+ survivors together make 250 gametes: 250 carry +; none carry *l*.

The 50 +/*l* survivors together make 500 gametes: 250 carry +; 250 carry *l*.

This gives us 500 copies of the + allele and 250 copies of the *l* allele for a total of 750. In this new gene pool, the frequency of the + allele is 0.67, and the frequency of the *l* allele is 0.33. We have gone from the gene pool in generation zero to the gene pool in generation one. The frequency of the + allele has risen, and the frequency of the *l* allele has fallen.

To get from generation one's gene pool to generation two's gene pool, we just repeat the exercise. We combine the gametes in generation one's gene pool at random to make 100 zygotes—45 +/+, 44 +/*l*, and 11 *l*/*l*—and so on. The only problem with using pencil-and-paper numerical calculations to predict evolution is that chasing the alleles around and around the life cycle all the way to generation 12 is a tedious job.

With a computer, however, predicting how Dawson's population will evolve is quick and easy. We can use a spreadsheet application to set up the required calculations ourselves (see Computing Consequences 6.3 and 6.5), or we can use any of a variety of population genetics programs that are already set up to do the calculations for us. Such programs take starting allele frequencies and genotype fitnesses as input and use the model we have developed in this chapter to produce predicted allele frequencies in future generations as output. We encourage the reader to get one of these programs and experiment with it.

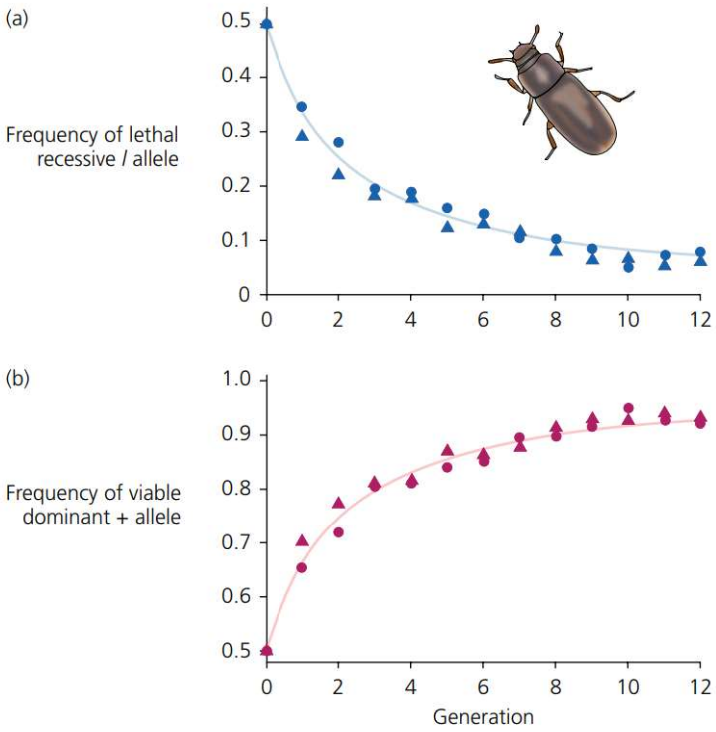


Figure 6.19 Evolution in laboratory populations of flour beetles (a) The decline in frequency of a lethal recessive allele (blue symbols) matches the theoretical prediction (blue curve) almost exactly. As the allele becomes rare, the rate of evolution slows dramatically. (b) This graph plots the increase in frequency of the corresponding dominant allele. Redrawn from Dawson (1970).

The prediction for Dawson's experiment appears as a curve in each of the graphs in Figure 6.19. The curve in the top graph predicts the falling frequency of the *l* allele; equivalently, the curve in the bottom graph predicts the rising frequency of the + allele. Our theory predicts that evolution will be rapid at first but will slow as the experiment proceeds.

Dawson's data appear in the graphs as colored circles and triangles. They match our theoretical predictions closely. This tight fit between prediction and data may seem unsurprising, even mundane. It should not. It should be astonishing. We

Empirical research on flour beetles shows that predictions made with population genetics models are accurate, at least under laboratory conditions.

COMPUTING CONSEQUENCES 6.6

An algebraic treatment of selection on recessive and dominant alleles

Here we develop equations that illuminate the differences between selection on recessive versus dominant alleles. Imagine a single locus with two alleles. Let p be the frequency of the dominant allele A, and let q be the frequency of the recessive allele a.

Selection on the recessive allele

Let the fitnesses of the genotypes be given by

w_{AA}	w_{Aa}	w_{aa}
1	1	$1 - s$

where s, called the **selection coefficient**, represents the strength of selection against homozygous recessives relative to the other genotypes. (Selection in favor of homozygous recessives can be accommodated by choosing a negative value for s.)

Based on Computing Consequences 6.3, the following equation gives the frequency of allele a in the next generation, q', given the frequency of a in this generation and the fitnesses of the three genotypes:

$$q' = \frac{pqw_{Aa} + q^2w_{aa}}{\overline{w}} = \frac{pqw_{Aa} + q^2w_{aa}}{p^2w_{AA} + 2pqw_{Aa} + q^2w_{aa}}$$

Substituting the fitness values from the table above, and $(1 - q)$ for p, then simplifying, gives

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

If a is a lethal recessive, then s is equal to 1. Substituting this value into the preceding equation gives

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

A little experimentation shows that once a recessive lethal allele becomes rare, further declines in frequency are slow. For example, if the frequency of allele a in this generation is 0.01, then in the next generation its frequency will be approximately 0.0099.

Selection on the dominant allele

Let the fitnesses of the genotypes be given by

w_{AA}	w_{Aa}	w_{aa}
$1 - s$	$1 - s$	1

where s, the selection coefficient, represents the strength of selection against genotypes containing the dominant allele relative to homozygous recessives. (Selection in favor of genotypes containing the dominant allele can be accommodated by choosing a negative value of s.)

Based on Computing Consequences 6.3, we can write an equation that predicts the frequency of allele A in the next generation, p', given the frequency of A in this generation and the fitness of the three genotypes:

$$p' = \frac{p^2w_{AA} + pqw_{Aa}}{\overline{w}} = \frac{p^2w_{AA} + pqw_{Aa}}{p^2w_{AA} + 2pqw_{Aa} + q^2w_{aa}}$$

Substituting the fitnesses from the table, and $(1 - p)$ for q, then simplifying, gives

$$p' = \frac{p(1 - s)}{1 - 2sp + sp^2}$$

If A is a lethal dominant, s is equal to 1. Substituting this value into the foregoing equation shows that a lethal dominant is eliminated from a population in a single generation.

Selection on recessive alleles versus selection on dominant alleles

Selection on recessive alleles and selection on dominant alleles are opposite sides of the same coin. Selection against a recessive allele is selection in favor of the dominant allele, and vice versa.

used a simple model of the mechanism of evolution combining the fundamental insights of Gregor Mendel with those of Charles Darwin to predict how a population would change over 12 generations. If the creatures in question had been humans instead of flour beetles, it would have meant forecasting events that will happen in 300 years. And Dawson's data show that our prediction was not just reasonably accurate, but spot on. If we had a theory that worked like that for picking stocks or racehorses—well, we could have retired years ago. Our model has passed its first test.

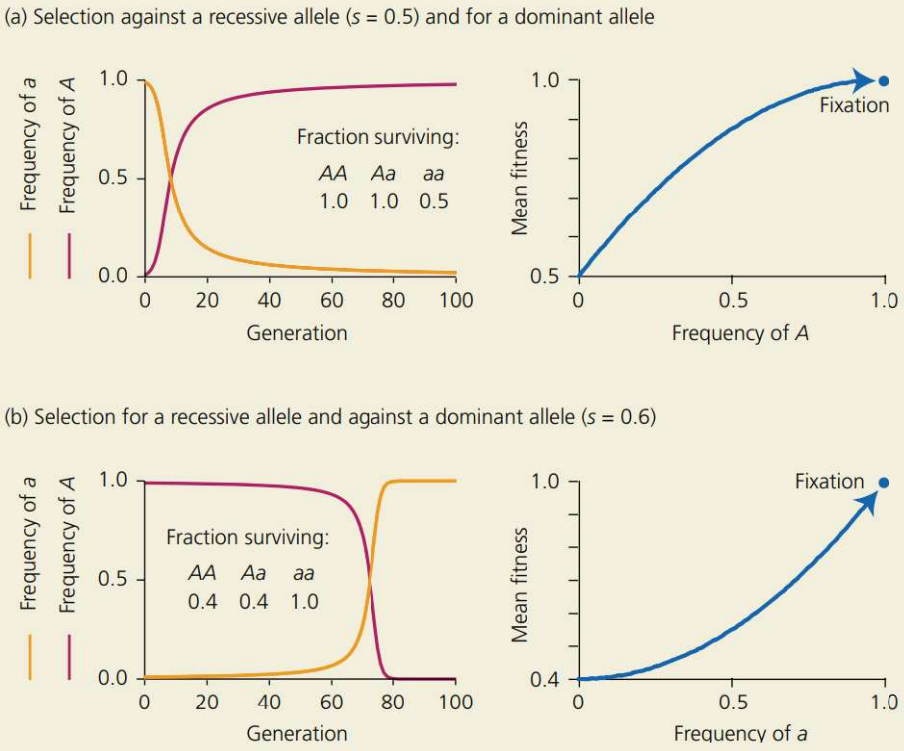


Figure 6.20 Evolution in model populations under selection on recessive and dominant alleles Graphs on the left show changes in allele frequencies over time. Graphs on the right show adaptive landscapes: Changes in population mean fitness as a function of allele frequencies.

Figure 6.20a (left) shows 100 generations of evolution in a model population with selection against a recessive allele and in favor of the dominant allele. At first, the allele frequencies change rapidly. As the recessive allele becomes rare, however, the rate of evolution slows dramatically. When the recessive allele is rare, most copies in the population are in heterozygous individuals, where they are effectively hidden from selection.

The figure also shows (right) the mean fitness of the population (see Computing Consequences 6.3) as a function of the frequency of the dominant allele. As the dominant allele goes from rare to common, the mean fitness of the population rises. Mean fitness is maximized when the favored allele reaches a frequency of 100%. Graphs of mean fitness as a function of allele frequency are often referred to as adaptive landscapes.

Figure 6.20b (left) shows 100 generations of evolution in a model population with selection in favor of a recessive allele and against the dominant allele. At first, the allele frequencies change slowly. The recessive allele is rare, most copies present are in heterozygotes, and selection cannot see it. However, as the recessive allele becomes common enough that a substantial fraction of homozygotes appear, the rate of evolution increases dramatically. Once the pace of evolution accelerates, the favorable recessive allele quickly achieves a frequency of 100%. That is, the recessive allele becomes fixed in the population.

The figure also shows (right) the mean fitness of the population (see Computing Consequences 6.3) as a function of the frequency of the recessive allele. As the recessive allele goes from rare to common, the mean fitness of the population rises. Mean fitness is maximized when the favored allele reaches a frequency of 100%.

An algebraic treatment of selection on recessive and dominant alleles appears in **Computing Consequences 6.6.** Even without the algebra, we can draw some important conclusions by reflecting further on Dawson's experiment.

Dawson's experiment shows that dominance and allele frequency interact to determine the rate of evolution. When a recessive allele is common (and a dominant allele is rare), evolution by natural selection is rapid. In contrast, when a recessive allele is rare, and a dominant allele is common, evolution by natural selection is slow. The Hardy–Weinberg equilibrium principle explains why.

First imagine a recessive allele that is common: Its frequency is, say, 0.95. The dominant allele thus has a frequency of 0.05. By multiplying the allele frequencies, we can calculate the genotype frequencies:

AA	*Aa*	*aa*
$0.05^2 = 0.0025$	$2 \cdot 0.05 \cdot 0.95 = 0.095$	$0.95^2 = 0.9025$

Roughly 10% of the individuals in the population have the dominant phenotype, while 90% have the recessive phenotype. Both phenotypes are reasonably well represented, and if they differ in fitness, then the allele frequencies in the next generation may be substantially different.

Natural selection is most potent as a mechanism of evolution when it is acting on common recessive alleles (and rare dominant alleles). When a recessive allele is rare, most copies are hidden in heterozygotes and protected from selection.

Now imagine a recessive allele that is rare: Its frequency is 0.05. The dominant allele thus has a frequency of 0.95. The genotype frequencies are

AA	*Aa*	*aa*
$0.95^2 = 0.9025$	$2 \cdot 0.95 \cdot 0.05 = 0.095$	$0.05^2 = 0.0025$

Approximately 100% of the population has the dominant phenotype, while approximately 0% has the recessive phenotype. Even if the phenotypes differ greatly in fitness, there are so few of the minority phenotype that there will be little change in allele frequencies in the next generation. In a random mating population, most copies of a rare recessive allele are phenotypically hidden inside heterozygous individuals and thereby immune from selection.

As a final consideration in our discussion of dominant and recessive alleles, note that selection may favor or disfavor both kinds of variants. We emphasize this point because many people new to population genetics expect that dominant alleles are necessarily beneficial and thus tend to rise in frequency. While it is certainly true that some dominant alleles are beneficial, many others are deleterious. For example, Eileen Shore and colleagues (2006) identified a dominant mutation, located in a gene encoding a receptor for bone morphogenic protein, as the cause of fibrodysplasia ossificans progressiva, a rare and severely disabling condition in which skeletal muscle and connective tissue transform inexorably into bone. In all, some 30% of the alleles known to cause human diseases are autosomal dominants (López-Bigas et al. 2006). The terms *dominant* and *recessive* describe the relationship between genotype and phenotype, not the relationship between genotype and fitness.

Selection on Heterozygotes and Homozygotes

In our next two tests, we focus on whether our model can accurately predict what happens when selection favors heterozygotes or homozygotes. Both tests will use data on laboratory populations of fruit flies (*Drosophila melanogaster*).

Selection Favoring Heterozygotes

Our first example comes from research by Terumi Mukai and Allan Burdick (1959). Like Dawson, Mukai and Burdick studied evolution at a single locus with two alleles. We will call the alleles *V*, for viable, and *L* for lethal. This is because flies with genotype *VV* or *VL* are alive, whereas flies with genotype *LL* are dead. The researchers used heterozygotes as founders to establish two experimental populations with initial allele frequencies of 0.5. They let the populations evolve for 15 generations, each generation measuring the frequency of allele *V*.

So far, Mukai and Burdick's experiment sounds just like Dawson's. If it is, then our theory predicts that *V* will rise in frequency—rapidly at first, then more

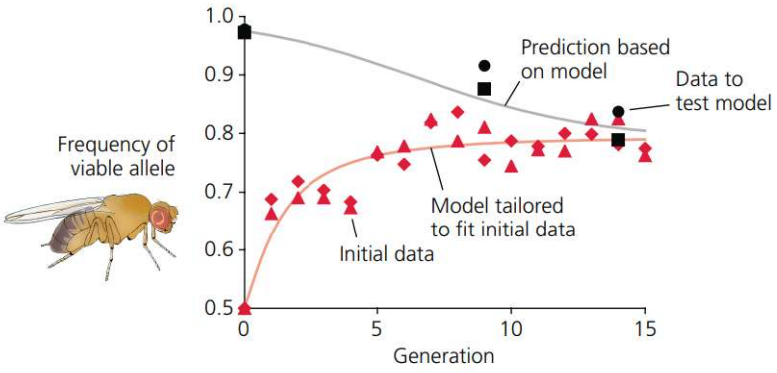


Figure 6.21 Evolution in four laboratory populations of fruit flies When homozygous, one allele is viable and the other lethal. Nonetheless, populations with a frequency of 0.5 for both alleles (red) evolved toward an intermediate equilibrium. The black populations represent a test of the hypothesis that heterozygotes enjoy the highest fitness. From data in Mukai and Burdick (1959).

slowly. By generation 15 it should reach a frequency of over 94%. But that is not what happened.

Mukai and Burdick's data appear in Figure 6.21, represented by the red symbols. As expected, the frequency of V increased rapidly over the first few generations. However, in both populations the rate of evolution slowed long before the viable allele approached a frequency of 0.94. Instead, V seemed to reach an equilibrium, or unchanging state, at a frequency of about 0.79.

How could this happen? An equilibrium frequency of 0.79 for the viable allele means that the lethal allele has an equilibrium frequency of 0.21. How could natural selection maintain a lethal allele at such a high frequency in this population? Mukai and Burdick argue that the most plausible explanation is **heterozygote superiority**, also known as **overdominance**. Under this hypothesis, heterozygotes have higher fitness than either homozygote. At equilibrium, the selective advantage enjoyed by the lethal allele when it is in heterozygotes exactly balances the obvious disadvantage it suffers when it is in homozygotes.

A little experimentation with a computer should allow the reader to confirm that Mukai and Burdick's hypothesis explains their data nicely. The red curve in Figure 6.21 represents evolution in a model population in which the fitnesses of the three genotypes are as follows:

VV	VL	LL
0.735	1.0	0

This theoretical curve matches the data closely.

Note that in this case the fit between theory and data does not represent a rigorous test of our model. That is because we examined the data first, then tweaked the fitnesses in the model to make its prediction fit. That is a bit like shooting at a barn and then painting a target around the bullet hole. Mukai and Burdick's flies did, however, provide an opportunity for a test of our model that is rigorous. And Mukai and Burdick performed it.

Research on fruit flies shows that natural selection can act to maintain two alleles at a stable equilibrium. One way this can happen is when heterozygotes have superior fitness.

The researchers established two more experimental populations, this time with the initial frequency of the viable allele at 0.975. If the genotype fitnesses are, indeed, those required to make our model fit the red data points in Figure 6.21, then this time our model predicts that the frequency of the V allele should fall. As before, it should ultimately reach an equilibrium near 0.79. The predicted fall toward equilibrium is shown by the blue curve in Figure 6.21. Mukai and Burdick's data appear in the figure as blue symbols. The data match the prediction closely. Our model has passed its second test.

Mukai and Burdick's flies have shown us something new. In all our previous examples, selection has favored one allele or the other. Under such circumstances

COMPUTING CONSEQUENCES 6.7

Stable equilibria with heterozygote superiority and unstable equilibria with heterozygote inferiority

Here we develop algebraic and graphical methods for analyzing evolution at loci with overdominance and underdominance. Imagine a population in which allele A_1 is at frequency p and allele A_2 is at frequency q. In Computing Consequences 6.3, we developed an equation describing the change in p from one generation to the next under selection:

$$\Delta p = \frac{p}{\overline{w}}(pw_{11} + qw_{12} - \overline{w})$$

$$= \frac{p}{\overline{w}}(pw_{11} + qw_{12} - p^2w_{11} - 2pqw_{12} - q^2w_{22})$$

Substituting $(1 - q)$ for p in the first and third terms in the expression in parentheses gives

$$\Delta p = \frac{p}{\overline{w}}[(1 - q)w_{11} + qw_{12} - (1 - q)^2w_{11} - 2pqw_{12} - q^2w_{22}]$$

which, after simplifying and factoring out q, becomes

$$\Delta p = \frac{pq}{\overline{w}}(w_{12} + w_{11} - qw_{11} - 2pw_{12} - qw_{22})$$

Now, by definition, the frequency of allele A_1 is at equilibrium when $\Delta p = 0$. The equation above shows that $\Delta p = 0$ when $p = 0$ or $q = 0$. These two equilibria are unsurprising. They occur when one allele or the other is absent from the population. The equation also gives a third condition for equilibrium, which is

$$w_{12} + w_{11} - qw_{11} - 2pw_{12} - qw_{22} = 0$$

Substituting $(1 - p)$ for q and solving for p gives

$$\hat{p} = \frac{w_{22} - w_{12}}{w_{11} - 2w_{12} + w_{22}}$$

where $\hat{p}$ is the frequency of allele A_1 at equilibrium. Finally, let the genotype fitnesses be as follows:

A_1A_1	A_1A_2	A_2A_2
$1 - s$	1	$1 - t$

Positive values of the selection coefficients s and t represent overdominance; negative values represent underdominance. Substituting the fitnesses into the previous equation and simplifying gives

$$\hat{p} = \frac{t}{s + t}$$

For example, when $s = 0.4$ and $t = 0.6$, heterozygotes have superior fitness, and the equilibrium frequency for allele A_1 is 0.6. When $s = -0.4$ and $t = -0.6$, heterozygotes have inferior fitness, and the equilibrium frequency for allele A_1 is also 0.6.

Another useful method for analyzing equilibria is to plot Δp as a function of p. Figure 6.20a shows such a plot for the two numerical examples we just calculated. Both curves show that $\Delta p = 0$ when $p = 0$, $p = 1$, or $p = 0.6$.

The curves in Figure 6.22a also allow us to determine whether an equilibrium is stable or unstable. Look at the red curve; it describes a locus with heterozygote superiority. Notice that when p is greater than 0.6, Δp is negative. This means that when the frequency of allele A_1 exceeds its equilibrium value, the population will move back toward equilibrium in the next generation. Likewise, when p is less than 0.6, Δp is positive. When

our model predicts that sooner or later the favored allele will reach a frequency of 100%, and the disfavored allele will disappear. By keeping a population at an equilibrium in which both alleles are present, however, heterozygote superiority can maintain genetic diversity indefinitely. For an algebraic treatment of heterozygote superiority, see **Computing Consequences 6.7**.

Selection Favoring Homozygotes

Our second example comes from work by G. G. Foster and colleagues (1972). These researchers set up experiments to demonstrate how populations evolve when heterozygotes have lower fitness than either homozygote. Foster and colleagues used fruit flies with compound chromosomes.

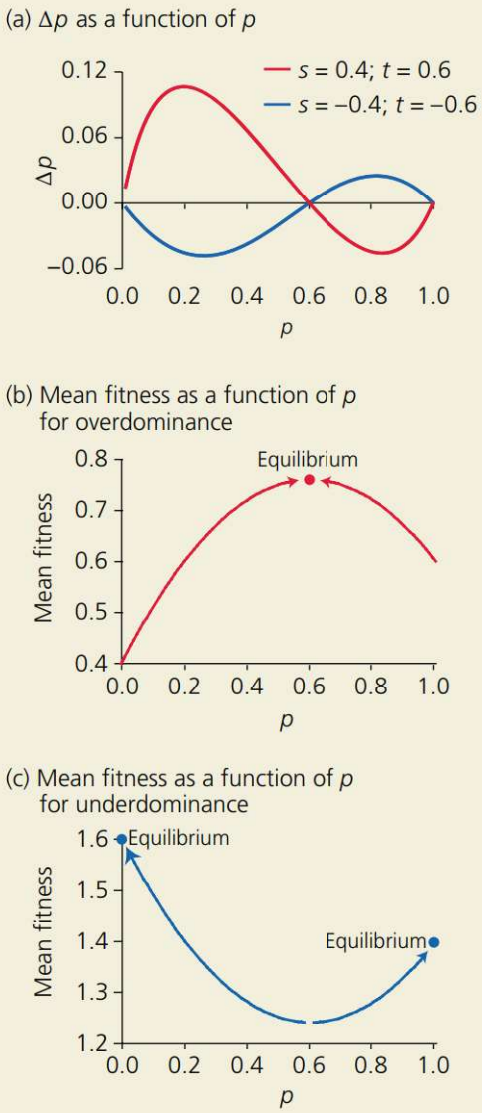


Figure 6.22 A graphical analysis of stable and unstable equilibria at loci with overdominance and underdominance (a) A plot of Δp as a function of p. (b) and (c) Adaptive landscapes.

the frequency of allele A_1 is below its equilibrium value, the population will move back toward equilibrium in the next generation. The "internal" equilibrium for a locus with heterozygote superiority is stable.

Figure 6.22b shows an adaptive landscape for a locus with heterozygote superiority. The graph plots population mean fitness as a function of the frequency of allele A_1. Mean fitness is low when A_1 is absent, and relatively low when A_1 is fixed. As the allele frequency moves from either direction toward its stable equilibrium, the population mean fitness rises to a maximum.

Now, look at the blue curve in Figure 6.22a. It describes a locus with heterozygote inferiority. If p rises even slightly above 0.6, p will continue to rise toward 1.0 in subsequent generations; if p falls even slightly below 0.6, p will continue to fall toward 0 in subsequent generations. The internal equilibrium for a locus with heterozygote inferiority is unstable.

Figure 6.22c shows an adaptive landscape for a locus with heterozygote inferiority. Population mean fitness is lowest when the frequency of allele A_1 is at its unstable internal equilibrium. As the allele frequency moves away from this equilibrium in either direction, mean fitness rises.

A comparison of the adaptive landscape in Figure 6.22c with those in Figure 6.22b and Figure 6.20 offers a valuable insight. As a population evolves in response to selection, the mean fitness of the individuals in the population tends to rise. Selection does not, however, always maximize mean fitness in a global sense. Depending on the initial allele frequencies, the population depicted in Figure 6.22c may evolve toward either fixation or loss of A_1. If the allele becomes fixed, the population will be at a stable equilibrium, but the population's mean fitness will be substantially lower than it would be if the allele were lost.

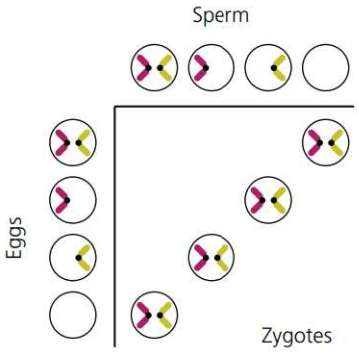
Compound chromosomes are homologous chromosomes that have swapped entire arms, so that one homolog has two copies of one arm, and the other homolog has two copies of the other arm **(Figure 6.23a and b, next page)**. During meiosis, compound chromosomes may or may not segregate. As a result, four kinds of gametes are produced in equal numbers: gametes with both homologous chromosomes, gametes with just one member of the pair, gametes with the other member of the pair, and gametes with neither member of the pair (Figure 6.23c). When two flies with compound chromosomes mate with each other, one-quarter of their zygotes have every chromosome arm in the correct dose and are thus viable (Figure 6.23d). The other three-quarters have too many or too few of copies of one or both chromosome arms and are thus inviable. When a fly

(a) A normal pair of homologous chromosomes (each has one magenta arm and one yellow arm).

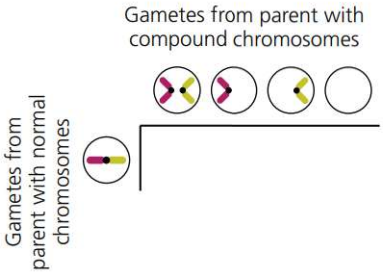
(b) A pair of compound chromosomes (one has two magenta arms, the other has two yellow arms).

(c) Gametes made by an individual with compound chromosomes may contain both chromosomes, one, or neither.

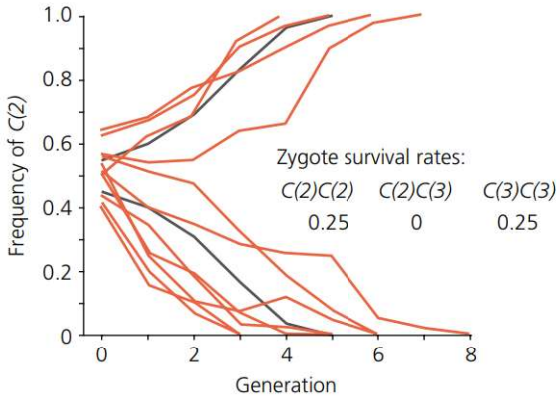
(d) When individuals with compound chromosomes mate, one quarter of their zygotes are viable.



(e) When an individual with compound chromosomes mates with an individual with normal chromosomes, none of their zygotes are viable.



(f) Evolution in 11 populations of *Drosophila melanogaster* containing compound second chromosomes [*C(2)*] and compound third chromosomes [*C(3)*]. The initial frequency of *C(2)* ranged from 0.4 to 0.65.



(g) Evolution in 13 populations of *Drosophila melanogaster* containing a mixture of compound second chromosomes [*C(2)*] and normal second chromosomes [*N(2)*]. The initial frequency of *C(2)* ranged from 0.71 to 0.96.

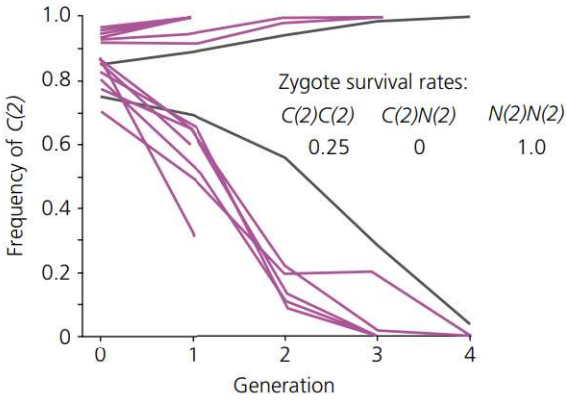


Figure 6.23 An experiment designed to show how populations evolve when heterozygotes have lower fitness than either homozygote (a–e) The experimental design makes clever use of compound chromosomes. (f and g) The data (orange and purple) match the theoretical predictions (gray). Redrawn with permission from Foster et al. (1972).

From "Chromosome rearrangements for the control of insect pests." *Science* 176:875–880. Reprinted with permission from AAAS.

with compound chromosomes mates with a fly with normal chromosomes, none of the zygotes they make are viable (Figure 6.23e).

Foster and colleagues established two sets of laboratory populations. In the first set of populations, some of the founders had compound second chromosomes [*C(2)*] and others had compound third chromosomes [*C(3)*]. Note that if two flies with compound second chromosomes mate, one-quarter of their offspring survive. Likewise, if two flies with compound third chromosomes mate, one-quarter of their offspring survive. But if a fly with compound second chromosomes (and

It is also possible for heterozygotes to have inferior fitness.

normal third chromosomes) mates with a fly with compound third chromosomes (and normal second chromosomes), none of their offspring survive. For purposes of analysis, then, we can treat the second and third chromosome as though they are alleles of a single locus. Thus the founders consisted of *C(2)C(2)* homozygotes and *C(3)C(3)* homozygotes. Based on the zygote viabilities we just described, the fitnesses of the possible offspring genotypes in the mixed population are

$C(2)C(2)$	$C(2)C(3)$	$C(3)C(3)$
0.25	0	0.25

In other words, the genotypes exhibit strong **underdominance**.

The algebraic analysis described in Computing Consequences 6.7 predicts that such a mixed population will be in genetic equilibrium, with both alleles present, when the frequency of *C(2)* is exactly 0.5. This equilibrium is unstable, however. If the frequency of *C(2)* ever gets above 0.5, then it should quickly rise to 1.0. Likewise, if the frequency of *C(2)* ever dips below 0.5, it should quickly fall to zero. Experimentation with a computer should allow the reader to reproduce this behavior.

Intuitively, the reason for the behavior is as follows. Heterozygotes are inviable, so the adults in the population are all homozygotes. Imagine first a situation in which *C(2)C(2)* individuals are common and *C(3)C(3)* individuals are rare. If the flies mate at random, then most matings will involve *C(2)C(2)* flies mating with each other, or *C(2)C(2)* flies mating with *C(3)C(3)* flies. Only rarely will *C(3)C(3)* flies mate with their own kind. Consequently, most *C(3)C(3)* flies will have zero reproductive success, and the frequency of *C(2)* will climb toward 1.0. Now imagine that *C(3)C(3)* individuals are common and *C(2)C(2)* individuals are rare. Under random mating, most matings involve *C(3)C(3)* flies mating with each other, or *C(3)C(3)* flies mating with *C(2)C(2)* flies. As a result, most of the *C(2)C(2)* flies will have zero reproductive success, and the frequency of *C(2)* will fall toward zero.

Foster and colleagues set up 11 mixed populations, with *C(2)* frequencies ranging from about 0.4 to about 0.65, then monitored their evolution for up to eight generations. Predictions for the evolution of populations with initial C(2) frequencies of 0.45 and 0.55 appear as gray lines in the graph in Figure 6.23f. The data from Foster et al.'s flies appear as orange lines. There is some deviation between prediction and result, probably due to genetic drift. That is, in a few of the experimental populations the frequency of *C(2)* started above 0.5 but ultimately fell to zero. In all 11 populations, however, once the frequency of *C(2)* had moved substantially away from 0.5, it continued moving in the same direction until it hit zero or one.

In the researchers' second set of populations, some founders had compound second chromosomes [*C(2)*] and others had normal second chromosomes [*N(2)*]. If two flies with compound second chromosomes mate, one-quarter of their offspring are viable. If a fly with compound second chromosomes mates with a fly with normal second chromosomes, none of their offspring is viable. If two flies with normal second chromosomes mate, all of their offspring are viable. Again, for purposes of analysis, we can treat each chromosome as though it were a single allele. Thus the founders consisted of *C(2)C(2)* homozygotes and *N(2)N(2)* homozygotes. The fitnesses of the genotypes in the mixed population are

$C(2)C(2)$	$C(2)N(2)$	$N(2)N(2)$
0.25	0	1.0

As in the first set of populations, the genotypes exhibit strong underdominance. This time, however, one kind of homozygote has much higher fitness than the other.

The algebraic analysis described in Computing Consequences 6.7 predicts an unstable equilibrium when the frequency of *C(2)* is exactly 0.8. If the frequency of *C(2)* ever gets above 0.8, then it should quickly rise to 1.0. Likewise, if the frequency of *C(2)* ever dips below 0.8, it should quickly fall to zero. Experimentation with a computer should allow the reader to reproduce this prediction.

The intuitive explanation is as follows. Heterozygotes are inviable, so the adults in the population are all homozygotes. Imagine first that *C(2)C(2)* individuals are common and *N(2)N(2)* individuals are rare. If the flies mate at random, then almost all matings will involve *C(2)C(2)* flies mating with each other, or *C(2)C(2)* flies mating with *N(2)N(2)* flies. Only very rarely will *N(2)N(2)* flies mate with their own kind. Consequently, most *N(2)N(2)* flies will have zero reproductive success, and the frequency of *C(2)* will climb to 1.0. Now imagine that there are enough *N(2)N(2)* flies present that appreciable numbers of them do mate with each other. These matings will produce four times as many offspring as matings between *C(2)C(2)* flies. Consequently, the frequency of *N(2)* will climb to 1.0 and the frequency of *C(2)* will fall to zero.

Foster and colleagues set up 13 mixed populations, with *C(2)* frequencies ranging from 0.71 to 0.96, then monitored their evolution for up to four generations. Predictions for the evolution of populations with initial *C(2)* frequencies of 0.75 and 0.85 appear as gray lines in the graph in Figure 6.23g. The data appear as purple lines. Qualitatively, the outcome matches the theoretical prediction nicely. In populations with higher initial *C(2)* frequencies, *C(2)* quickly rose to fixation, while in populations with lower initial *C(2)* frequencies, *C(2)* was quickly lost. The exact location of the unstable equilibrium turned out to be approximately 0.9 instead of 0.8. Foster and colleagues note that their *C(2)C(2)* flies carried recessive genetic markers, bred into them to allow for easy identification. They suggest that these markers reduced the relative fitness of the *C(2)C(2)* flies below the value of 0.25 inferred solely on the basis of their compound chromosomes.

Our model's predictions were not as accurate for Foster et al.'s experiments as they were for Dawson's and Mukai and Burdick's. Nonetheless, the model performed well. It predicted something we have not seen before: an unstable equilibrium above which the frequency of an allele would rise and below which it would fall. It predicted that the unstable equilibrium would be higher in Foster et al.'s second set of populations than in their first. And its predictions about the rate of evolution were roughly correct. Our model has passed its third test.

When heterozygotes have inferior fitness, one allele tends to go to fixation while the other allele is lost. However, different populations may lose different alleles.

Foster et al.'s experiments demonstrate that heterozygote inferiority leads to a loss of genetic diversity within populations. By driving different alleles to fixation in different populations, however, heterozygote inferiority may help maintain genetic diversity among populations.

Frequency-Dependent Selection

For our fourth and final test of population genetics theory, we will see whether our model can predict the evolutionary outcome when the fitness of individuals with a particular phenotype depends on their frequency in the population. Our example, from the work of Luc Gigord, Mark Macnair, and Ann Smithson (2001), concerns a puzzling color polymorphism in the Elderflower orchid (*Dactylorhiza sambucina*).

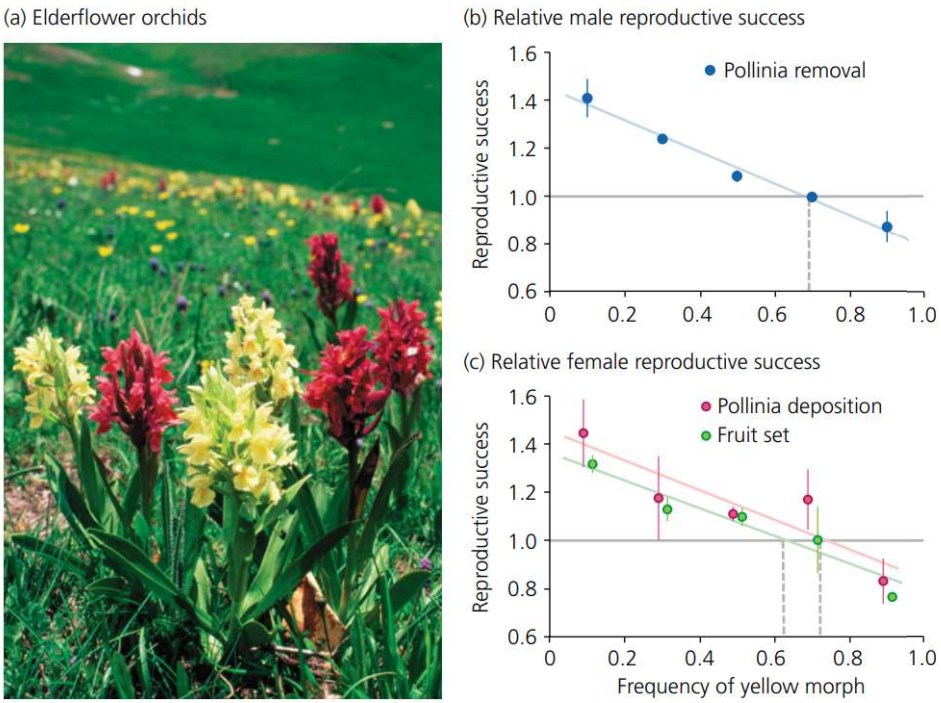


Figure 6.24 Frequency-dependent selection in Elderflower orchids (a) A mixed population. Some plants have yellow flowers, others have purple flowers. (b) Through male function, yellow flowers have higher fitness than purple flowers when yellow is rare, but lower fitness than purple flowers when yellow is common. (c) Through female function, yellow flowers have higher fitness than purple flowers when yellow is rare, but lower fitness than purple flowers when yellow is common. The dashed vertical lines show the predicted frequency of yellow flowers, which matches the frequency in natural populations. From Gigord et al. (2001).

Elderflower orchids come in yellow and purple (Figure 6.24a). Populations typically include both colors, though yellow is usually more common. The flowers attract bumblebees, which are the orchid's main pollinators. But the bees that visit Elderflower orchids are always disappointed. To the bees the orchid's colorful flowers appear to advertise a reward, but in fact they offer nothing. The puzzle Gigord and colleagues wanted to solve was this: How can two distinct deceptive advertisements persist together in Elderflower orchid populations?

The researchers' hypothesis grew from earlier observations by Smithson and Macnair (1997). When naive bumblebees visit a stand of orchids to sample the flowers, they tend to alternate between colors. If a bee visits a purple flower first and finds no reward, it looks next in a yellow flower. Finding nothing there either, it tries another purple one. Disappointment sends it back to a yellow, and so on, until the bee gives up and leaves. Because bumblebees tend to visit equal numbers of yellow and purple flowers, orchids with the less common of the two colors receive more visits per plant. If more pollinator visits translate into higher reproductive success, then the rare-color advantage could explain why both colors persist. Selection by bumblebees favors yellow until it becomes too common, then it favors purple. This is an example of **frequency-dependent selection**.

To test their hypothesis, Gigord and colleagues collected and potted wild orchids, then placed them in the orchids' natural habitat in 10 experimental arrays of 50 plants each. The frequency of yellow flowers varied among arrays, with two arrays at each of five frequencies: 0.1, 0.3, 0.5, 0.7, and 0.9. The researchers monitored the orchids for removal of their own pollinia (pollen-bearing structures), for deposition of pollinia from other individuals, and for fruit set. From their data, Gigord and colleagues estimated the reproductive advantage of yellow flowers, relative to purple, via both male and female function.

The resulting estimates of relative reproductive success, plotted as a function of the frequency of yellow flowers, appear in Figure 6.24b and c. Consistent with the researchers' hypothesis, yellow-flowered orchids enjoyed higher reproductive

Selection can also maintain two alleles in a population if each allele is advantageous when it is rare.

success than purple-flowered plants when yellow was rare and suffered lower reproductive success when yellow was common.

Gigord and colleagues calculated the relative reproductive success of yellow orchids as

$$RRS_y = \frac{2(RS_y)}{RS_y + RS_p}$$

where RS_y and RS_p are the absolute reproductive success of yellow and purple orchids. The relationship between relative reproductive success via male function and the frequency of yellow flowers is given by the best-fit line in Figure 6.21b. It is

$$RRS_y = -0.66F_y + 1.452$$

where F_y is the frequency of yellow flowers.

We can incorporate this relationship into a population genetics model. We might imagine, for example, that flower color is determined by two alleles at a single locus and that yellow is recessive to purple. We set the starting frequency of the yellow allele to an arbitrary value. We assign fitnesses to the three genotypes as we have before, except that the fitnesses change each generation with the frequency of yellow flowers. When we use a computer to track the evolution of our model population, we discover that the frequency of the yellow allele moves rapidly to equilibrium at an intermediate value. This value is precisely the allele frequency at which yellow flowers have a relative fitness of 1. We get the same result if we imagine that yellow flowers are dominant. Again the equilibrium value for the yellow allele is the frequency at which yellow and purple flowers have equal fitness.

The dashed vertical lines in Figure 6.24b and c indicate the predicted equilibrium frequencies Gigord and colleagues calculated for each of their fitness measures. The predictions are 61%, 69%, and 72% yellow flowers. The researchers surveyed 20 natural populations in the region where they had placed their experimental arrays. The actual frequency of yellow flowers, 69 ± 3%, is in good agreement with the predicted frequency. Our model has passed its fourth test.

Gigord et al.'s study of Elderflower orchids demonstrates that frequency-dependent selection can have an effect similar to heterozygote superiority. Both patterns of selection can maintain genetic diversity in populations.

Compulsory Sterilization

We can use population genetics models to evaluate whether eugenic sterilization could have accomplished the aims of its proponents, had their assumptions about the heritability of traits been correct. The answer depends on the frequency of the alleles in question, and on the criteria for success.

The theory of population genetics, despite its simplifying assumptions, allows us to predict the course of evolution. Our four tests show that the model we have developed works remarkably well. So long as we know the starting allele frequencies and genotype fitnesses, the model can predict how allele frequencies will change, under a variety of selection schemes, many generations into the future. The requisite knowledge is easiest to get, of course, for experimental populations living under controlled conditions in the lab. But Gigord et al.'s study of Elderflower orchids shows that the model can even make fairly accurate predictions in natural populations. Given its success in the four tests, it is reasonable to use our model to consider the evolutionary consequences of a eugenic sterilization program. The proponents of eugenic sterilization sought to reduce the fitness of particular genotypes to zero and thereby to reduce the frequency of alleles responsible for undesirable phenotypes. Would their plan have worked?

The phenotype that caught the eugenicists' attention perhaps more than any other was feeblemindedness. The Royal College of Physicians in England defined a feebleminded individual as "One who is capable of earning his living under favorable circumstances, but is incapable from mental defect existing from birth or from an early age (a) of competing on equal terms with his normal fellows or (b) of managing himself and his affairs with ordinary prudence" (see Goddard 1914). Evidence presented in 1914 by Henry H. Goddard, who was the director of research at the Training School for Feebleminded Girls and Boys in Vineland, New Jersey, convinced many eugenicists that strength of mind behaved like a simple Mendelian trait (see Paul and Spencer 1995). Normal-mindedness was believed to be dominant and feeblemindedness recessive.

A recessive genetic disease is not a promising target for a program that would eliminate it by sterilizing affected individuals. As Figures 6.19a and 6.20a show, rare recessive alleles decline in frequency slowly, even under strong selection. On the other hand, eugenicists did not believe that feeblemindedness was especially rare (Paul and Spencer 1995). Indeed, they believed that feeblemindedness was alarmingly common and increasing in frequency. Edward M. East (1917) estimated the frequency of feeblemindedness at three per thousand. Henry H. Goddard reported a frequency of 2% among New York schoolchildren. Tests of American soldiers during World War I suggested a frequency of nearly 50% among white draftees.

We will assume a frequency for feeblemindedness of 1% and reproduce a calculation reported by R. C. Punnett (1917) and revisited by R. A. Fisher (1924). Let f be the purported allele for feeblemindedness, with frequency q. If 1% of the population has genotype ff, then, by the Hardy–Weinberg equilibrium principle, the initial frequency of f is

$$q = \sqrt{0.01} = 0.1$$

If all affected individuals are sterilized, then the fitness of genotype ff is zero (or, equivalently, the selection coefficient for genotype ff is 1). Using the equation developed in Computing Consequences 6.6, we can calculate the value of q in successive generations, and from q we can calculate the frequency of genotype ff.

The result appears in Figure 6.25. Over 10 generations, about 250 years, the frequency of affected individuals declines from 0.01 to 0.0025.

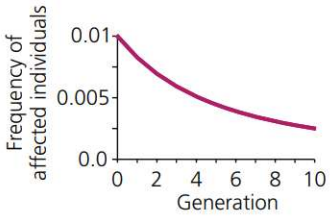


Figure 6.25 Predicted evolution due to sterilization The graph shows the change in the frequency of homozygotes for a putative allele for feeblemindedness under a eugenic sterilization program that prevents homozygous recessive individuals from reproducing.

Whether geneticists saw this calculation as encouraging or discouraging depended on whether they saw the glass as partially empty or partially full. Some looked at the numbers, saw that it would take a very long time to completely eliminate feeblemindedness, and argued that compulsory sterilization was such a hopelessly slow solution that it was not worth the effort. Others, such as Fisher, dismissed this argument as "anti-eugenic propaganda." Fisher noted that after just one generation, the frequency of affected individuals would drop from 100 per 10,000 to 82.6 per 10,000. "In a single generation," he wrote, "the load of public expenditure and personal misery caused by feeblemindedness . . . would be reduced by over 17 percent." Fisher also noted that most copies of the allele for feeblemindedness are present in heterozygous carriers rather than affected individuals. Along with East, Punnett, and others, Fisher called for research into methods for identifying carriers.

While their evolutionary logic was sound, the eugenicists' models were built on dubious genetic hypotheses. It is not entirely fair to use modern standards to criticize Goddard's research on the genetics of feeblemindedness. Mendelian

genetics was in its infancy. Still, looking back after nearly a century, we can see that Goddard's evidence was deeply flawed. We will consider three problems.

First, the individuals whose case studies he reports are a highly diverse group. Some have Down syndrome; some have other developmental challenges. At least one is deaf and appears to be the victim of a woefully inadequate education. Some appear to have been deposited at Goddard's training school by widowed fathers who felt that children from a prior marriage were a liability in finding a new wife. Some may just have behaved differently than the directors of the school thought they should. Concluding the first case report in his book, Goddard writes of a 16-year-old who has been at the school for seven years:

> Gertrude is a good example of that type of girl who, loose in the world, makes so much trouble. Her beauty and attractiveness and relatively high [intelligence] would enable her to pass almost anywhere as a normal child and yet she is entirely incapable of controlling herself and would be led astray most easily. It is fortunate for society that she is cared for as she is.

Second, Goddard's methods for collecting data were prone to distortion. He sent caseworkers to collect pedigrees from the families of the students at the training school. The caseworkers relied on hearsay and subjective judgments to assess the strength of mind of family members—many of whom were long since deceased.

Third, Goddard's method of analysis stacked the cards in favor of his conclusion. He first separated his 327 cases into various categories: definitely hereditary cases; probably hereditary cases; cases caused by accidents; and cases with no assignable cause. He apparently placed cases in his "definitely hereditary" group only when they had siblings, recent ancestors, or other close kin also classified as feebleminded. When he later analyzed the data to determine whether feeblemindedness was a Mendelian trait, Goddard analyzed only the data from his "definitely hereditary" group. Given how he had filtered the data ahead of time, it is not too surprising that he concluded that feeblemindedness is Mendelian.

Although feeblemindedness is not among them, many genetic diseases are now known to be inherited as simple Mendelian traits. Yet eugenic sterilization has few advocates. One reason is that most serious genetic diseases are recessive and very rare; sterilization of affected individuals would have little impact on the frequency at which new affected individuals are born. A second reason is that mainstream attitudes about reproductive rights have changed to favor individual autonomy over societal mandates (Paul and Spencer 1995). A third reason is that, as we discuss in the next section, a growing list of disease alleles are suspected or known to be maintained in populations by heterozygote superiority. It would be futile and possibly ill advised to try to reduce the frequency of such alleles by preventing affected individuals from reproducing.

6.4 Mutation

Cystic fibrosis is among the most common serious genetic diseases among people of European ancestry, affecting approximately 1 newborn in 2,500. Cystic fibrosis is inherited as an autosomal recessive trait. Affected individuals suffer chronic infections with the bacterium *Pseudomonas aeruginosa* and ultimately sustain severe lung damage (Pier et al. 1997). At present, most individuals with cystic fibrosis

live into their thirties or forties (Elias et al. 1992), but until recently few survived to reproductive age. Although cystic fibrosis was lethal for most of human history, in some populations as many as 4% of individuals are carriers. How can alleles that cause a lethal genetic disease remain this common?

Our consideration of heterozygote superiority in the previous section hinted at one possible answer. Another potential answer is that new disease alleles are constantly introduced into populations by mutation. Before we can evaluate the relative merits of these two hypotheses for explaining the persistence of any particular disease allele, we need to discuss mutation in more detail.

Elsewhere in the book (see Chapter 5), we presented mutation as the source of all new alleles and genes. In its capacity as the ultimate source of all genetic variation, mutation provides the raw material for evolution. Here, we consider the importance of mutation as a mechanism of evolution. How rapidly does mutation cause allele frequencies to change over time? How strongly does mutation affect the conclusions of the Hardy–Weinberg analysis?

Adding Mutation to the Hardy–Weinberg Analysis: Mutation as an Evolutionary Mechanism

Second on the list of assumptions for the Hardy–Weinberg equilibrium principle was that there are no mutations. We now explore what happens to allele frequencies when this assumption is violated.

Mutation by itself is generally not a rapid mechanism of evolution. To see why, return to our model population of mice. Imagine a locus with two alleles, A and a, with initial frequencies of 0.9 and 0.1. A is the wild-type allele, and a is a recessive loss-of-function mutation. Furthermore, imagine that copies of A are converted by mutation into new copies of a at the rate of 1 copy per 10,000 per generation. This is a very high mutation rate, but it is within the range of mutation rates known. Back-mutations that restore function are much less common than loss-of-function mutations, so we will ignore mutations that convert copies of a into new copies of A. Finally, imagine that all mutations happen while the adults are making gametes to contribute to the gene pool.

Figure 6.26 follows the allele and genotype frequencies through one turn of the life cycle. Among the zygotes, juveniles, and adults, the genotypes are in Hardy–Weinberg proportions:

AA	Aa	aa
0.81	0.18	0.01

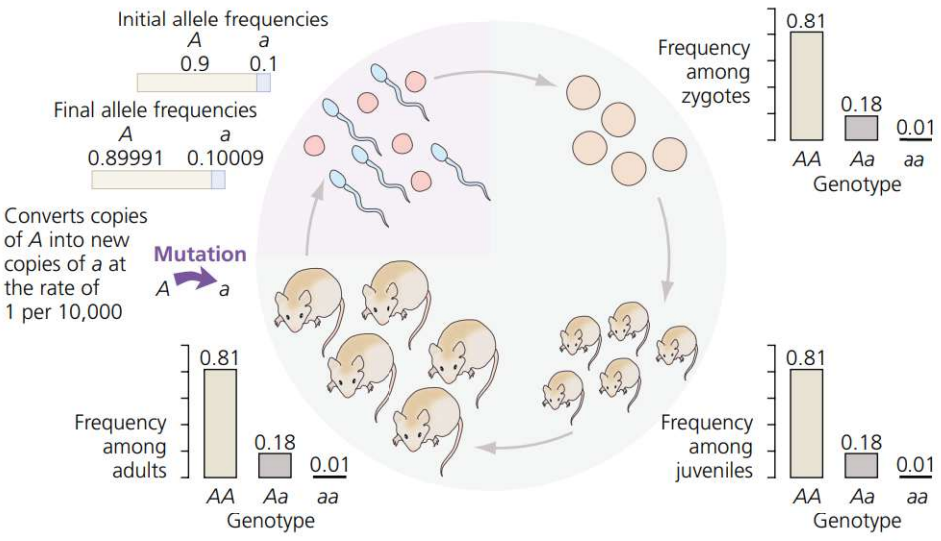


Figure 6.26 Mutation is a weak mechanism of evolution In a single generation in our model population, mutation produces virtually no change in allele and genotype frequencies.

COMPUTING CONSEQUENCES 6.8

A mathematical treatment of mutation as an evolutionary mechanism

Imagine a single locus with two alleles: a wild-type allele, A, and a recessive loss-of-function mutation, a. Let μ be the rate of mutation from A to a. Assume that the rate of back-mutation from a to A is negligible. If the frequency of A in this generation is p, then its frequency in the next generation is given by

$$p' = p - \mu p$$

If the frequency of a in this generation is q, then its frequency in the next generation is given by

$$q' = q + \mu p$$

The change in p from one generation to the next is

$$\Delta p = p' - p$$

which simplifies to

$$\Delta p = -\mu p$$

After n generations, the frequency of A is approximately

$$p_n = p_0 e^{-\mu n}$$

where p_n is the frequency of A in generation n, p_0 is the frequency of A in generation 0, and e is the base of the natural logarithms.

Readers familiar with calculus can derive the last equation as follows. First, assume that a single generation is an infinitesimal amount of time, so that we can rewrite the equation $\Delta p = -\mu p$ as

$$\frac{dp}{dg} = -\mu p$$

Now divide both sides by p, and multiply both sides by dg to get

$$\left(\frac{1}{p}\right) dp = -\mu dg$$

Finally, integrate the left side from frequency p_0 to p_n and the right side from generation 0 to n:

$$\int_{p_0}^{p_n} \left(\frac{1}{p}\right) dp = \int_0^n -\mu dg$$

and solve for p_n.

Now the adults make gametes. Were it not for mutation, the allele frequencies in the new gene pool would be

A	a
0.9	0.1

But mutation converts 1 of every 10,000 copies of allele A into a new copy of allele a. The frequency of A after mutation is given by the frequency before mutation minus the fraction lost to mutation; the frequency of a after mutation is given by the frequency before mutation plus the fraction gained by mutation. That is,

A	a
$0.9 - (0.0001 \cdot 0.9) = 0.89991$	$0.1 + (0.0001 \cdot 0.9) = 0.10009$

The new allele frequencies are almost identical to the old allele frequencies. As a mechanism of evolution, mutation has had almost no effect.

Almost no effect is not the same as exactly no effect. Could mutation of A into a, occurring at the rate of 1 copy per 10,000 every generation for many generations, eventually result in an appreciable change in allele frequencies? The graph in Figure 6.27 provides the answer (see Computing Consequences 6.8 for a mathematical treatment). After 1,000 generations, the frequency of allele A in our model population will be about 0.81. Mutation can cause substantial change in allele frequencies, but it does so slowly.

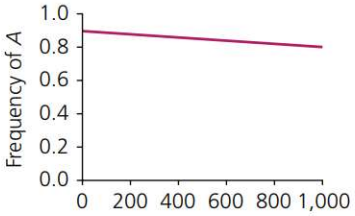


Figure 6.27 Over very long periods of time, mutation can eventually produce appreciable changes in allele frequency

As mutation rates go, the value we used in our model, 1 per 10,000 per generation, is very high. For most genes, mutation is an even less efficient mechanism of allele frequency change.

Hardy–Weinberg analysis shows that mutation is a weak mechanism of evolution.

Mutation and Selection

Although mutation alone usually does not cause appreciable changes in allele frequencies, this does not mean that mutation is unimportant in evolution. In combination with selection, mutation is a crucial piece of the evolutionary process. This point is demonstrated by an experiment conducted by Mingcai Zhang, Priti Azad, and Ron Woodruff (2011), who showed that fruit fly populations with virtually no initial genetic variation accumulate novel alleles quickly enough to allow rapid adaptive evolution.

Zhang and colleagues began with a stock of *Drosophila melanogaster* that had been propagated by single-pair sibling matings for over 150 generations. As we will see later (Chapter 7), this kind of intense inbreeding results in rapid loss of genetic variation. Screening at loci that are highly variable in most populations confirmed that all flies in the inbred stock were essentially genetically identical.

The researchers next reared larvae from their inbred stock on food spiked with table salt (NaCl) at concentrations ranging from 1% to 6%. At least a few larvae survived to adulthood at concentrations up to 4%, but all the larvae died at 5%.

Then Zhang and colleagues used flies from the inbred stock as founders for six experimental populations, which they maintained for 30 generations. They kept the population sizes large, establishing each generation with 200 pairs of randomly chosen adults from the previous generation. The researchers kept two of the populations under benign conditions, and four on food spiked with salt. They distributed the salty food in patches with different concentrations, but all of it was stressful for the flies. The conditions of the experiment allowed the populations to evolve by natural selection to adapt to their new environments, but the populations would do so only if they accrued genetic variation via mutation.

Finally, the researchers assessed the salt tolerance of the thirtieth generation in each of the six populations. Survival data for larvae reared on food spiked with 5% salt appear in Figure 6.28. The original inbred stock appears first, as a reminder that for the founding flies, 5% salt was 100% lethal. The unstressed lines appear next. Even though theses lines had evolved under benign conditions, both included a few individuals that could survive in 5% salt. This result demonstrates the accumulation of genetic variation by mutation in the absence of selection. The salt-stressed lines appear last. They contained higher proportions of individuals that could survive 5% salt. This result demonstrates adaptive evolution as a result of mutation and natural selection in combination. Further evidence that the stressed lines harbored alleles for salt tolerance at higher frequency than the unstressed lines came when Zhang and colleagues attempted to rear larvae on 6% salt. A few individuals from each of the salt-stressed lines survived, but all the flies from the unstressed lines died.

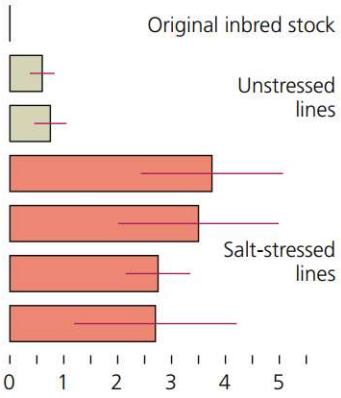


Figure 6.28 Adaptive evolution resulting from natural selection on novel mutations Bars show the salt tolerance of flies in the thirtieth generation of fruit fly lineages founded from a single genetically homogenous stock. Prepared from data in Zhang et al. (2011).

Research with fruit flies illustrates that while mutation itself is only a weak mechanism of evolution, it nonetheless supplies the raw material on which natural selection acts.

The experiment by Zhang and colleagues reinforces one of the messages we discussed earler (Chapter 5). Without mutation, evolution would eventually grind to a halt. Mutation is the ultimate source of genetic variation.

Mutation–Selection Balance

Unlike the mutations that allowed the evolution of increased salt tolerance in Zhang et al.'s fruit fly populations, many mutations are at least mildly deleterious.

COMPUTING CONSEQUENCES 6.9

Allele frequencies under mutation–selection balance

Here we derive equations for predicting the equilibrium frequencies of deleterious alleles under mutation–selection balance. Imagine a single locus with two alleles, A_1 and A_2, with frequencies p and q. A_1 is the wild type; A_2 is deleterious. Let μ be the rate at which copies of A_1 are converted into copies of A_2 by mutation. Assume that the rate of back-mutation is negligible.

Selection continuously removes copies of A_2 from the population, while mutation continuously creates new copies. We want to calculate the frequency of A_2 at which these processes cancel each other. Following Felsenstein (1997), we will perform our calculation in a roundabout way. We will develop an equation in terms of p that describes mutation–selection balance for allele A_1. Then we will solve the equation for q to get the equilibrium frequency of A_2. This approach may seem perverse, but it greatly simplifies the algebra.

Mutation–selection balance for a deleterious recessive

Imagine that A_2 is a deleterious recessive allele, such that the genotype fitnesses are given by

w_{11}	w_{12}	w_{22}
1	1	$1 - s$

where the selection coefficient s gives the strength of selection against A_2.

First, we write an equation for $p^\star$, the frequency of allele A_1 after selection has acted, but before mutations occur. From Computing Consequences 6.3, this is

$$p^\star = \frac{p^2 w_{11} + pq w_{12}}{p^2 w_{11} + 2pq w_{12} + q^2 w_{22}}$$

Substituting the fitnesses from the table above, and $(1 - p)$ for q, then simplifying gives

$$p^\star = \frac{p}{1 - s(1 - p)^2}$$

Next we write an expression for p', the frequency of allele A_1 after mutations occur. Mutations convert a fraction μ of the copies of A_1 into copies of A_2, leaving behind a fraction $(1 - \mu)$. Thus

$$p' = (1 - \mu)p^\star = \frac{(1 - \mu)p}{1 - s(1 - p)^2}$$

Finally, when mutation and selection are in balance, p' is equal to p, the frequency of allele A_1 that we started with:

$$\frac{(1 - \mu)p}{1 - s(1 - p)^2} = p$$

This simplifies to

$$(1 - p)^2 = \frac{\mu}{s}$$

Substituting q for $(1 - p)$ and solving for q yields an equation for $\hat{q}$, the equilibrium frequency of allele A_2 under mutation–selection balance:

$$\hat{q} = \sqrt{\frac{\mu}{s}}$$

If A_2 is a lethal recessive, then $s = 1$, and the equilibrium frequency of A_2 is equal to the square root of the mutation rate.

Mutation–selection balance for a lethal dominant allele

Imagine that A_2 is a lethal dominant allele, such that the genotype fitnesses are given by

w_{11}	w_{12}	w_{22}
1	0	0

Now the expression for $p^\star$ simplifies to

$$p^\star = 1$$

which makes sense because, by definition, selection removes all copies of the lethal dominant A_2 from the population. Now the expression for p' is

$$p' = 1 - \mu$$

and the equilibrium condition is

$$1 - \mu = p$$

Substituting $(1 - q)$ for p and simplifying gives

$$\hat{q} = \mu$$

In other words, the equilibrium frequency of A_2 is equal to the mutation rate.

Selection tends to eliminate such mutations from populations. Deleterious alleles persist, however, because they are continually created anew. When the rate at which copies of a deleterious allele are being eliminated by selection is exactly equal to the rate at which new copies are being created by mutation, the frequency of the allele is at equilibrium. This situation is called **mutation–selection balance**.

What is the frequency of the deleterious allele at equilibrium? If the allele is recessive, its equilibrium frequency, $\hat{q}$, is given by

$$\hat{q} = \sqrt{\frac{\mu}{s}}$$

where μ is the mutation rate, and s, the selection coefficient, is a number between 0 and 1 expressing the strength of selection against the allele (see **Computing Consequences 6.9** for a derivation). This equation captures with economy what intuition tells us about mutation–selection balance. If the selection coefficient is small (the allele is only mildly deleterious) and the mutation rate is high, then the equilibrium frequency of the allele will be relatively high. If the selection coefficient is large (the allele is highly deleterious) and the mutation rate is low, then the equilibrium frequency of the allele will be low.

Research by Brunhilde Wirth and colleagues (1997) on patients with spinal muscular atrophy provides an example. Spinal muscular atrophy is a neurodegenerative disease characterized by weakness and wasting of the muscles that control voluntary movement. It is caused by deletions in a locus on chromosome 5 called the telomeric survival motor neuron gene (*telSMN*). In some cases, the disease may be exacerbated by additional mutations in a nearby gene. Spinal muscular atrophy is, after cystic fibrosis, the second most common lethal autosomal recessive disease in Caucasians (McKusick et al. 1999).

Collectively, the loss-of-function alleles of *telSMN* have a frequency of about 0.01 in the Caucasian population. Wirth and colleagues estimate that the selection coefficient is about 0.9. With such strong selection against them, we would expect that disease-causing alleles would slowly but inexorably disappear from the population. How, then, do they persist at a frequency of 1 in 100?

At the same time selection removes deleterious alleles from a population, mutation constantly supplies new copies. In some cases, this balance between mutation and selection may explain the persistence of deleterious alleles in populations.

One possibility is that the disease alleles are being kept in the population by a balance between mutation and selection. If we substitute the allele frequency and selection coefficient for $\hat{q}$ and s in the equation above, and then solve for μ, we find that this scenario requires a mutation rate of about 9.0×10^{-5} $(= 0.9 \times 10^{-4})$ mutations per *telSMN* allele per generation.

Wirth and colleagues analyzed the chromosomes of 340 individuals with spinal muscular atrophy, and the chromosomes of their parents and other family members. They found that 7 of the 340 affected individuals carried a new mutation not present in either parent. These numbers allowed the scientists to estimate directly the mutation rate at the *telSMN* locus (see **Computing Consequences 6.10**). Their estimate is 1.1×10^{-4}. This directly measured mutation rate is in good agreement with the rate predicted under the hypothesis of a mutation–selection balance. Wirth and colleagues conclude that mutation–selection balance provides a sufficient explanation for the persistence of spinal muscular atrophy alleles.

It is possible that the parents of spinal muscular atrophy patients are unusually susceptible to mutations in *telSMN*. Ideally, we would determine the mutation rate by comparing the genotypes of parents and offspring in the general population. However, such a study would require an extremely large sample size.

COMPUTING CONSEQUENCES 6.10

Estimating mutation rates for recessive alleles

Here, we present the method used by Brunhilde Wirth and colleagues (1997) to estimate mutation rates for recessive alleles. The key information required is the fraction of affected individuals that carry a brand-new mutant allele. With modern molecular techniques, this fraction can be obtained by direct examination of the chromosomes of affected individuals and their relatives.

Let q be the frequency of recessive loss-of-function allele *a*. Ignoring the extremely rare individuals with two new mutant copies, there are two ways to be born with genotype *aa*:

1. An individual can be the offspring of two carriers. The probability of this outcome for a given birth is the product of (a) the probability that an offspring of two carriers will be affected; (b) the probability that the mother is a carrier; and (c) the probability that the father is a carrier. This probability is given by

$$\left[\frac{1}{4}\right] \times [2q(1-q)] \times [2q(1-q)]$$

2. An individual can be the offspring of one carrier and one homozygous dominant parent and can receive allele *a* from the affected parent and a new mutant copy of *a* from the unaffected parent. The probability of this outcome for a given birth is the product of (a) the probability that an offspring of one carrier will receive that carrier's mutant allele; (b) the probability that the mother is a carrier; (c) the probability that the father is the homozygous dominant; and (d) the mutation rate plus the same probability for the scenario in which the father is the carrier and the mother is the homozygous dominant:

$$\left\{\left[\frac{1}{2}\right] \times [2q(1-q)] \times [(1-q)^2] \times [\mu]\right\} + \left\{\left[\frac{1}{2}\right] \times [2q(1-q)] \times [(1-q)^2] \times [\mu]\right\} = [2q(1-q)] \times [(1-q)^2] \times [\mu]$$

With these probabilities, we can write an expression for r, the fraction of affected individuals that carry one new mutant allele. This is the second probability divided by the sum of the second probability and the first. Simplified just a bit, we have

$$r = \frac{2q(1-q)(1-q)^2\mu}{2q(1-q)(1-q)^2\mu + q(1-q)q(1-q)}$$

Simplifying further yields

$$r = \frac{2(1-q)\mu}{2(1-q)\mu + q}$$

Finally, assume that q is small, so that $(1-q)$ is approximately equal to one. This assumption gives

$$r = \frac{2\mu}{2\mu + q}$$

which can be solved for μ:

$$\mu = \frac{rq}{2-2r}$$

The mutation rate for spinal muscular atrophy

In Caucasian populations, spinal muscular atrophy affects about 1 infant in 10,000, implying that the frequency of the mutant allele is

$$q = \sqrt{0.0001} = 0.01$$

Wirth and colleagues examined the chromosomes of 340 affected patients and their family members. The researchers discovered that seven of their patients had a new mutant allele not present in either parent. Thus

$$r = \frac{7}{340} = 0.021$$

Substituting these values for q and r into the equation for μ gives the estimate

$$\mu = \frac{(0.021)(0.01)}{2 - 2(0.021)} = 0.00011$$

The mutation rate for cystic fibrosis

In Caucasian populations, cystic fibrosis affects about 1 infant in 2,500. Wirth and colleagues cite data from other authors to establish that only 2 of about 30,000 cystic fibrosis patients studied proved to have a new mutant allele not present in either parent. These figures give an estimated mutation rate of

$$\mu = 6.7 \times 10^{-7}$$

Are the Alleles That Cause Cystic Fibrosis Maintained by a Balance between Mutation and Selection?

Cystic fibrosis is caused by recessive loss-of-function mutations in a locus on chromosome 7 that encodes a protein called the cystic fibrosis transmembrane conductance regulator (CFTR). CFTR is a cell-surface protein expressed in the mucous membrane lining the intestines and lungs. Gerald Pier and colleagues (1997) demonstrated that one of CFTR's key functions is to enable cells of the lung lining to ingest and destroy *Pseudomonas aeruginosa* bacteria (see also Campodónico et al. 2008). Figure 6.29 shows a snapshot of this process (Bajmoczi et al. (2009). In individuals with cystic fibrosis, *P. aeruginosa* cause chronic lung infections that eventually lead to severe lung damage (Figure 6.30).

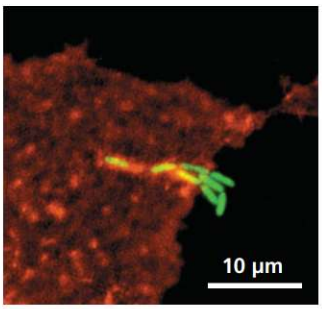


Figure 6.29 A lung epithelial cell ingesting *Pseudomonas aeruginosa* The lung cell is red; the bacteria are green. The bacteria on the left are already inside the cell. They are surrounded by halos of fluorescently labeled CFTR. From Bajmoczi et al. (2009).

Selection against the alleles that cause cystic fibrosis appears to be strong. Until recently, few affected individuals survived to reproductive age; those that do survive are often infertile. And yet the alleles that cause cystic fibrosis have a collective frequency of approximately 0.02 among people of European ancestry.

Could cystic fibrosis alleles be maintained at a frequency of 0.02 by mutation–selection balance? If we assume a selection coefficient of 1 and use the equation derived in Computing Consequences 6.9, the mutation rate creating new disease alleles would have to be 4×10^{-4}. The actual mutation rate for cystic fibrosis alleles appears to be considerably lower than that: approximately 6.7×10^{-7} (see Computing Consequences 6.10). We can conclude that a steady supply of new mutations cannot, by itself, explain the maintenance of cystic fibrosis alleles at a frequency of 0.02.

In other cases, the frequency of a deleterious allele may be too high to explain by mutation–selection balance. This may be a clue that heterozygotes have superior fitness.

Our discussion of heterozygote superiority suggests an alternative explanation (Figure 6.21 and Computing Consequences 6.7). Perhaps the fitness cost suffered by cystic fibrosis alleles when they are in homozygotes is balanced by a fitness advantage they enjoy when they are in heterozygotes.

Gerald Pier and colleagues (1998) hypothesized that cystic fibrosis heterozygotes might be resistant to typhoid fever and therefore have superior fitness. Typhoid fever is caused by *Salmonella typhi* bacteria (also known as *Salmonella enterica* serovar *typhi*). The bacteria initiate an infection by crossing the layer of epithelial cells that line the gut. Pier and colleagues suggested that *S. typhi* bacteria infiltrate the gut by exploiting the CFTR protein as a point of entry. If so, then heterozygotes, which have fewer copies of CFTR on the surface of their cells, should be less vulnerable to infiltration.

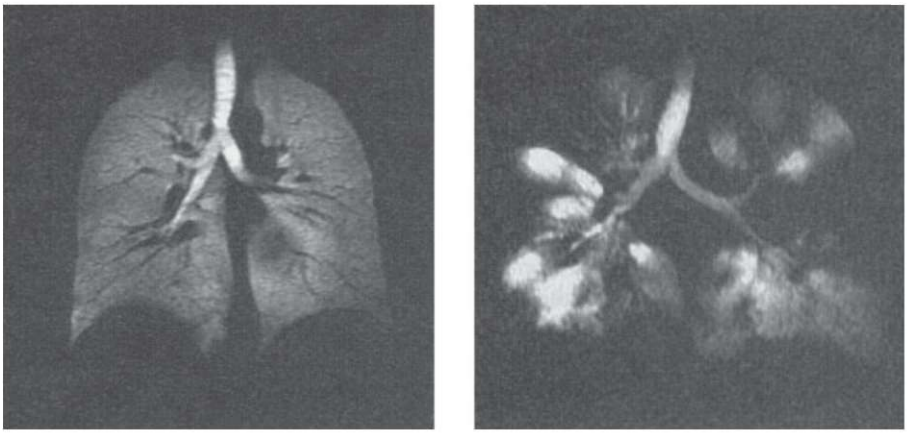


Figure 6.30 Lung damage in cystic fibrosis Left, normal lungs. Right, lungs ravaged by the bacterial infections that accompany cystic fibrosis. Photos by James R. MacFall, Duke University Medical Center.

Pier and colleagues tested their hypothesis by constructing mouse cells with three different CFTR genotypes: homozygous wild-type cells; heterozygotes with one functional CFTR allele and one allele containing the most common human cystic fibrosis mutation, a single-codon deletion called *ΔF508;* and homozygous *ΔF508* cells. The researchers exposed these cells to *S. typhi*, then measured the number of bacteria that got inside cells of each genotype. The results were dramatic (Figure 6.31a). As the researchers predicted, homozygous *ΔF508* cells were almost totally resistant to infiltration by *S. typhi*, while homozygous wild-type cells were highly vulnerable. Heterozygous cells were partially resistant; they accumulated 86% fewer bacteria than did the wild-type cells. These results are consistent with the hypothesis that cystic fibrosis disease alleles are maintained in human populations because heterozygotes have superior fitness during typhoid fever epidemics.

Also consistent with the hypothesis are two more recent discoveries by Pier and coworkers. First, Jeffrey Lyczak and Pier (2002) found that *S. typhi* bacteria manipulate the gut cells of their hosts, causing the cells to display more CFTR protein on their membranes and easing the bacteria's entry. This helps explain why cells that cannot make CFTR are resistant to invasion. Second, Lyczak, Carolyn Cannon, and Pier (2002), using data compiled from the literature, found an apparent association across 11 European countries between the severity of typhoid fever outbreaks and the frequency a generation later, among CFTR mutations, of the *ΔF508* allele (Figure 6.31b).

Pier et al.'s research serves as another example in which an evolutionary analysis has proved valuable in biomedical research.

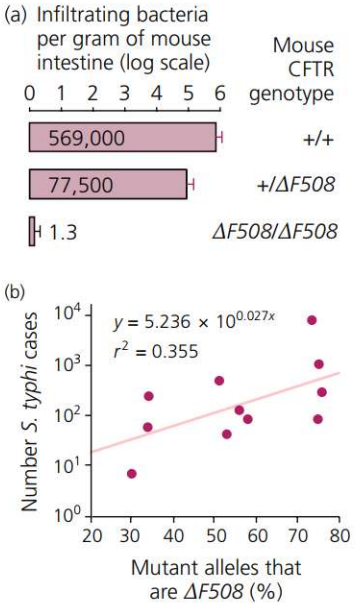


Figure 6.31 Heterozygotes for the *ΔF508* allele are resistant to typhoid fever (a) The rate at which *Salmonella typhi* infiltrate cultured mouse cells with different CFTR genotypes. From Pier et al. (1998). (b) The severity of typhoid fever outbreaks in 11 European countries versus the frequency of *ΔF508*, among cystic fibrosis mutations, in the generation born following the outbreak. From Lyczak et al. (2002).

6.5 An Engineering Test of Population Genetics Theory

Chun-Hong Chen and colleagues (2007), working in the laboratory of Bruce Hay, sought methods to confer malaria resistance on free-living mosquitoes. Their concern was not for the mosquitoes, but for people. If the mosquitoes are resistant to malaria, they cannot transmit the disease to humans.

The task the researchers had set for themselves was one of evolutionary engineering. Genetic variants that make mosquitoes resistant to malaria were already known. The challenge was to ensure that the resistance genes, once introduced into a wild population, would rise to high frequency. Chen and colleagues had an idea for how to do this, which they put to the test in laboratory populations of fruit flies.

The researchers designed a new gene that they expected would carry a strong selective advantage. The gene was a synthetic example of a kind of genetic element, called a *Medea*, that also occurs naturally. *Medea* is an acronym for Maternal-effect dominant embryonic arrest. It is also the name of the title character in a play by Euripides about a mother who murders her own children.

Chen's synthetic *Medea* includes two sets of instructions (Figure 6.32a). One causes mothers that carry the element to infuse their eggs with a poison. The other allows embryos that carry the element to make an antidote. If mother and baby both carry the gene, the baby lives (Figure 6.32b). If mother carries the gene but the baby does not, the baby dies (Figure 6.32c).

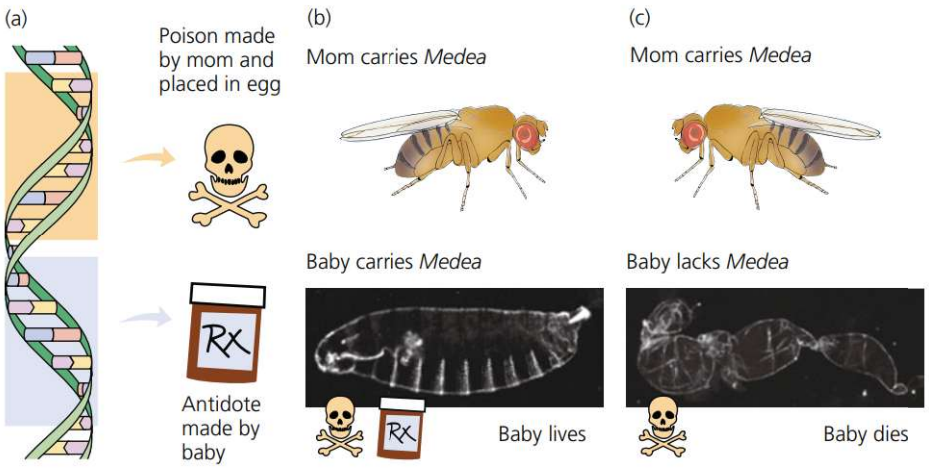


Figure 6.32 An artificial *Medea* gene (a) The gene encodes both a poison and its antidote. (b) Mothers carrying *Medea* make the poison and put it into their eggs. If the baby carries *Medea*, it makes the antidote and lives. (c) If the baby does not carry *Medea*, it dies. Photos from Kambris et al. (2003).

An embryonic fly's fate is determined by the genotypes of both its mother and its father. We will call the two alleles a fly can carry *M* (for *Medea*) and + (for wild type, or lack of *Medea*). Punnett squares predict that if mom is a heterozygote and dad is a ++ homozygote, half the babies will die. If mom and dad are both heterozygotes, a quarter of the babies will die **(Figure 6.33)**.

Mom	Dad	Offspring	Predicted	Actual
+*M*	++	+ + / + ~~+~~ / *M* *M*+ *M*+	50% die	51.7 % die
+*M*	+*M*	+ *M* / + ~~+~~ +*M* / *M* *M*+ *MM*	25% die	25.7 % die

Figure 6.33 Punnett squares for crosses in which the mother carries *Medea* and some of the offspring do not. Data from Chen et al. (2007).

For all other matings, all the babies live. In experimental matings, Chen and colleagues found these predictions to be accurate.

We claimed earlier that Chen and colleagues expected their synthetic Medea gene to carry a strong selective advantage. This claim may be counterintuitive for a gene that causes mothers to kill their offspring. But note that the offspring that die all lack *Medea*. The selective advantage accrues not to the individuals that carry the gene, but to the gene itself. *Medea* is a selfish allele that, given the chance, kills individuals that do not carry it. If introduced into a population, *Medea* should inexorably rise in frequency.

Chen and colleagues introduced their synthetic *Medea* into laboratory fruit fly populations at a frequency of 0.25. They used the basic population genetics theory we have discussed in this chapter to predict the trajectory of the gene's rise in frequency. They assumed an infinitely large population with random mating and no mutation or migration. They assumed no fitness costs of *Medea* other than the mortality inflicted by the maternal poison on embryos lacking the gene. The only complication in Chen's model, compared to the models we have used throughout the chapter, is that an individual's fitness depends not just on its own genotype but also on the genotypes of its parents. That meant Chen had to track genotype frequencies and account for the progeny from random matings, rather

An attempt to design a gene that would rise to high frequency in a predictable way in an insect population provides a strong test of our understanding of the mechanisms of evolution.

COMPUTING CONSEQUENCES 6.11

Predicting the frequency of *Medea* across generations

Here we provide an overview of how we can predict changes in the frequency of Chen et al.'s *Medea* gene. Because an individual's fitness depends on both its own genotype and the genotypes of its parents, we have to keep track of genotype frequencies and matings.

Let P, Q, and R be the frequencies of genotypes ++, +M, and MM. Random mating results in the nine types of matings shown in Figure 6.34. Each type of mating occurs at the frequency shown in the upper left of its square. For example, matings between ++ females and ++ males, shown in square (1), occur at frequency P^2.

We want to know the frequency of each genotype in the next generation. Consider genotype ++. Zygotes with this genotype are conceived in matings (1), (2), (4), and (5). All the zygotes conceived in mating type (1) are ++, as are half of those conceived in mating types (2) and (4) and a quarter of those conceived in mating type (5). The frequency of genotype ++ at conception is thus

$$P^2 + \tfrac{1}{2}PQ + \tfrac{1}{2}QP + \tfrac{1}{4}Q^2$$

Before the zygotes develop into larvae and hatch, however, all the ++ offspring conceived in mating types (4) and (5) die. The frequency of ++ individuals in the next generation is the number of surviving ++ individuals divided by the total number of survivors,

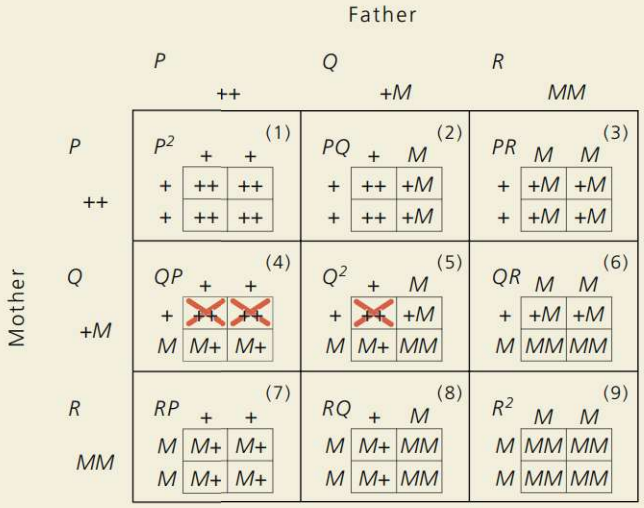


Figure 6.34 A Punnett square of Punnett squares

which is given by

$$\frac{P^2 + \tfrac{1}{2}PQ}{P^2 + PQ + PR + \tfrac{1}{2}QP + \tfrac{3}{4}Q^2 + QR + RP + RQ + R^2}$$

We leave it to the reader to write expressions for the frequency of +M and MM individuals in the next generation.

Once we have a generation's genotype frequencies, we can readily calculate the allele frequencies.

than simply track allele frequencies and account for the progeny from random unions of gametes (**Computing Consequences 6.11**).

Chen and colleagues' prediction, and the data for 7 populations, some maintained for 15 generations and some for 20, appear in Figure 6.35. Their predic-

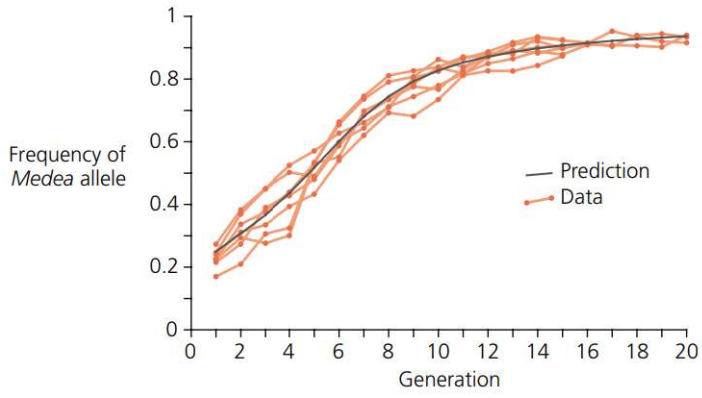


Figure 6.35 Predicted and actual evolution of laboratory populations harboring an engineered gene The prediction is in gray; the data are in orange. From Chen et al. (2007).

tion, as we have come to expect, was spot on. Take a moment to reflect on what Chen and colleagues accomplished. They designed and built a new gene, accurately predicted the effect it would have on the individuals that carry it, introduced the gene into populations, and accurately predicted how the populations would change over the course of 20 generations. The success of this evolutionary engineering project demonstrates that, like NASA's engineers know something about physics, we know a thing or two about inheritance and descent with modification.

Population genetics is a theory that works.

In the next chapter, we will continue to use the Hardy–Weinberg equilibrium principle to explore additional mechanisms of evolution.

SUMMARY

Population genetics represents a synthesis of Mendelian genetics and Darwinian evolution and is concerned with the mechanisms that cause allele frequencies to change from one generation to the next. The Hardy–Weinberg equilibrium principle is a null model that provides the conceptual framework for population genetics. It shows that under simple assumptions—no selection, no mutation, no migration, no genetic drift, and random mating—allele frequencies do not change. Furthermore, genotype frequencies can be calculated from allele frequencies.

When any one of the first four assumptions is violated, allele frequencies may change across generations. Selection, mutation, migration, and genetic drift are thus the four mechanisms of evolution. Nonrandom mating does not cause allele frequencies to change and is thus not a mechanism of evolution. It can, however, alter genotype frequencies and thereby affect the course of evolution.

Population geneticists can measure allele and genotype frequencies in real populations. Thus, biologists can test whether allele frequencies are stable across generations and whether the genotype frequencies conform to Hardy–Weinberg expectations. If either of the conclusions of the Hardy–Weinberg analysis is violated, then one or more of the assumptions does not hold. The nature of the deviation from Hardy–Weinberg expectations does not, by itself, identify the faulty assumption. We can, however, often infer which mechanisms of evolution are at work based on other characteristics of the populations under study.

Selection occurs when individuals with some genotypes are more successful at getting copies of their genes into future generations than are individuals with other genotypes. Selection is a powerful mechanism of evolution. It can cause allele frequencies to change from one generation to the next and can take genotype frequencies away from Hardy–Weinberg equilibrium. Some patterns of selection tend to drive alleles to fixation or to loss; other patterns of selection serve to maintain allelic diversity in populations. Population genetics theory allows us to make accurate predictions about both the direction and the rate of evolution under a variety of patterns of selection.

Alone, mutation is a weak evolutionary mechanism. Mutation does, however, provide the genetic variation that is the raw material for evolution. In some cases, a steady supply of new mutant alleles can counterbalance selection against those same alleles and thereby serve to hold allele frequencies at equilibrium.

QUESTIONS

1. List the five conditions that must be true for a population to be in Hardy–Weinberg equilibrium. Why is it useful to know the conditions that prevent evolution? For each condition, specify whether violation of that assumption results in changes in genotype frequencies, allele frequencies, or both.

2. Why was it important that G. H. Hardy used variables in his mathematical treatment of changes in population allele frequencies across generations? Would it have been equally useful to simply work several more examples with different specific allele frequencies?

3. Name the phenomenon being described in each of these (hypothetical) examples, and describe how it is likely to affect allele frequencies in succeeding generations.
   a. A beetle species is introduced to an island covered with dark basaltic rock. On this dark background, dark beetles, *TT* or *Tt*, are much more resistant to predation than are light-colored beetles, *tt*. The dark beetles have a large selective advantage. Both alleles are relatively common in the group of beetles released on the new island.
   b. Another beetle population, this time consisting of mostly light beetles and just a few dark beetles, is introduced onto a different island with a mixed substrate of light sand, vegetation, and black basalt. On this island, dark beetles have only a small selective advantage.
   c. A coral-reef fish has two genetically determined types of male. One kind of male is much smaller than the other, and sneaks into larger males' nests to fertilize their females' eggs. When small males are rare, they have a selective advantage over large males. However, if there are too many small males, large males switch to a more aggressive strategy of nest defense, and small males lose their advantage.
   d. In a tropical plant, *CC* and *Cc* plants have red flowers and cc plants have yellow flowers. However, *Cc* plants have defective flower development and produce very few flowers.
   e. In a species of bird, individuals with genotype *MM* are susceptible to avian malaria, *Mm* birds are resistant to avian malaria, and *mm* birds are resistant to avian malaria, but the *mm* birds are also vulnerable to avian pox.
4. In Muehlenbachs et al.'s study of placental malaria, why was it important that they studied infants born during both high and low malaria season? Can you think of any other possible explanations for their data?
5. Black color in horses is governed primarily by a recessive allele at the A locus. *AA* and *Aa* horses are nonblack colors such as bay, while *aa* horses are black all over. (Other loci can override the effect of the A locus, but we will ignore that complication.) In an online conversation, one person asked why there are relatively few black horses of the Arabian breed. One response was, "Black is a rare color because it is recessive. More Arabians are bay or gray because those colors are dominant." Discuss the merits and/or problems with this argument. (Assume that the *A* and *a* alleles are in Hardy–Weinberg equilibrium, which was probably true at the time of this discussion.) Generally, what does the Hardy–Weinberg model show us about the impact that an allele's dominance or recessiveness has on its frequency?
6. In humans, the COL1A1 locus codes for a certain collagen protein found in bone. The normal allele at this locus is denoted with *S*. A recessive allele *s* is associated with reduced bone mineral density and increased risk of fractures in both *Ss* and *ss* women. A study of 1,778 women showed that 1,194 were *SS*, 526 were *Ss*, and 58 were *ss* (Uitterlinden et al. 1998). Are these two alleles in Hardy–Weinberg equilibrium in this population? How do you know? What information would you need to determine whether the alleles will be in Hardy–Weinberg equilibrium in the next generation?
7. We used Figure 6.14 as an example of how the frequency of an allele (in fruit flies) does not change in unselected (control) populations but does change in response to selection. However, look again at the unselected control lines in Figure 6.14. The frequency of the allele in the two control populations did change a little, moving up and down over time. Which assumption of the Hardy–Weinberg model is most probably being violated? If this experiment were repeated, what change in experimental design would reduce this deviation from Hardy–Weinberg equilibrium?
8. Most animal populations have a 50:50 ratio of males to females. This does not have to be so; it is theoretically possible for parents to produce predominantly male offspring or predominantly female offspring. Imagine a population with a male-biased sex ratio, say, 70% males and 30% females. Which sex will have an easier time finding a mate? As a result, which sex will probably have higher average fitness? Which parents will have higher fitness—those that produce mostly males or those that produce mostly females? Now imagine the same population with a female-biased sex ratio, and answer the same questions. What sort of selection is probably maintaining the 50:50 sex ratio seen in most populations?
9. Discuss how each of the following recent developments—resulting from improvements in medicine,

technology, pubic health, and from evolution—may affect the frequency of alleles that cause cystic fibrosis (CF).

a. Many women with CF now survive long enough to have children. (CF causes problems with reproductive ducts, but many CF women can bear children nonetheless. CF men are usually sterile.)

b. Typhoid fever in developed nations has declined to very low levels since 1900.

c. In some populations, couples planning to have children are now routinely screened for the most common CF alleles.

d. Drug-resistant typhoid fever has recently appeared in several developing nations.

10. Kerstin Johannesson and colleagues (1995) studied two populations of a marine snail living in the intertidal zone on the shore of Ursholmen Island. Each year, the researchers determined the allele frequencies for the enzyme aspartate aminotransferase (don't worry about what this enzyme does). Their data are shown in the graphs in Figure 6.36. The first year of the study was 1987. In 1988, a bloom of toxic algae (tan bars) killed all of the snails in the intertidal zone across the entire island. That is why there are no data for 1988 and 1989. Although the snails living in the intertidal zone were exterminated by the bloom, snails of the same species living in the splash zone just above the intertidal survived unscathed. By 1990, the intertidal zone had been recolonized by splash-zone snails. Your challenge in this question is to develop a coherent explanation for the data in the graphs. In each part, be sure to name the evolutionary mechanism involved (selection, mutation, migration, or drift).

a. Why was the frequency of the Aat^{120} allele higher in both populations in 1990 than it was in 1987? Name the evolutionary mechanism, and explain.

b. Why did the allele frequency decline in both populations from 1990 through 1993? Name the evolutionary mechanism, and explain.

c. Why are the curves traced by the 1990–1993 data for the two populations generally similar but not exactly identical? Name the evolutionary mechanism, and explain.

d. Predict what would happen to the allele frequencies if we followed these two populations for another 100 years (assuming there are no more toxic algal blooms). Explain your reasoning.

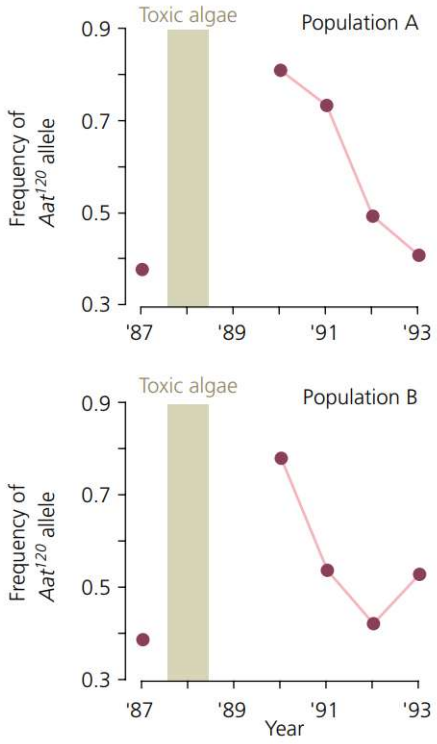


Figure 6.36 Changes over time in the frequency of an allele in two intertidal populations of a marine snail
From Johannesson et al. (1995).

EXPLORING THE LITERATURE

11. The photo and graph on the first page of this chapter document the evolution of coat color in a population of Soay sheep. The cause of this evolutionary change been the subject of some controversy. See the following papers:

Ozgul, A., S. Tuljapurkar, et al. 2009. The dynamics of phenotypic change and the shrinking sheep of St. Kilda. *Science* 325: 464–467.

Maloney, S. K., A. Fuller, and D. Mitchell. 2009. Climate change: Is the dark Soay sheep endangered? *Biology Letters* 5: 826–829.

Gratten, J., A. J. Wilson, et al. 2010. No evidence for warming climate theory of coat colour change in Soay sheep: A comment on Maloney et al. *Biology Letters* 6: 678–679.

Maloney, S. K., A. Fuller, and D. Mitchell. 2010. A warming climate remains a plausible hypothesis for the decrease in dark Soay sheep. *Biology Letters* 6: 680–681.

For an analysis of the evolution of coat pattern in the same sheep population, see:

Gratten, J., J. G. Pilkington, et al. 2012. Selection and microevolution of coat pattern are cryptic in a wild population of sheep. *Molecular Ecology* 21: 2977–2990

12. Often, the first step in a study of genetic variation is to evaluate deviations from Hardy–Weinberg equilibrium. Read the following paper to explore how careful examination of Hardy–Weinberg equilibrium is necessary for assessing gene–disease associations in humans.

Trikalinos, T. A., G. Salanti, et al. 2006. Impact of violations and deviations in Hardy–Weinberg equilibrium on postulated gene-disease associations. *American Journal of Epidemiology* 163: 300–309.

13. In the Elderflower orchid, we saw that frequency-dependent selection tends to maintain the presence of both yellow and purple flowers in mixed populations. See the following references for additonal cases of possible frequency-dependent selection. How strong is the evidence in each example?

Cox, C. L., and A. R. Davis Rabosky. In press. Spatial and temporal drivers of phonotypic diversity in polymorphic snakes. *American Naturalist*. (These snakes appear in Figure 5.1, page 148.)

Eizaguirre, C., T. L. Lenz, et al. 2012. Rapid and adaptive evolution of MHC genes under parasite selection in experimental vertebrate populations. *Nature Communications* 3: 621.

Faurie, C., and M. Raymond. 2005. Handedness, homicide and negative frequency-dependent selection. *Proceedings of the Royal Society of London B* 272: 25–28.

Hori, M. 1993. Frequency-dependent natural selection in the handedness of scale-eating cichlid fish. *Science* 260: 216–219.

Sinervo, B., and C. M. Lively. 1996. The rock-paper-scissors game and the evolution of alternative male strategies. *Nature* 380: 240–243.

14. The version of the adaptive landscape presented in Computing Consequences 6.6 and 6.7, in which the landscape is a plot of mean fitness as a function of allele frequency, is actually somewhat different from the original version of the concept that Sewall Wright presented in 1932. Furthermore, there is even a third common interpretation of the adaptive landscape idea. For a discussion of the differences among the three versions, see Chapter 9 in

Provine, W. B. 1986. *Sewall Wright and Evolutionary Biology*. Chicago: University of Chicago Press.

For Sewall Wright's response to Provine's history, see:

Wright, S. 1988. Surfaces of selective value revisited. *American Naturalist* 131: 115–123.

Wright's original 1932 paper is reprinted in Chapter 11 of:

Wright, S. 1986. *Evolution: Selected Papers,* ed. W. B. Provine. Chicago: University of Chicago Press.

15. If you have access to the earliest volumes of the *Journal of Heredity*, read:

Bell, Alexander Graham. 1914. How to improve the race. *Journal of Heredity* 5: 1–7.

Keep in mind that population genetics was in its infancy; Mendelism had yet to be integrated with natural selection. What was accurate and inaccurate in Bell's understanding of the mechanisms of evolution? Would the policy Bell advocated actually have accomplished his aims? Why or why not? If so, would it have done so for the reasons Bell thought it would?

16. For an example in which strong natural selection caused rapid change in allele frequencies in wild populations, see:

Rank, N. E., and E. P. Dahlhoff. 2002. Allele frequency shifts in response to climate change and physiological consequences of allozyme variation in a montane insect. *Evolution* 56: 2278–2289.

17. For an example in which strong selection for insecticid resistance caused rapid change in allele frquencies, see:

Mathias, D. K., E. Ochomo, et al. 2011. Spatial and temporal variation in the kdr allele L1014S in Anopheles gambiae s.s. and phenotypic variability in susceptibility to insecticides in Western Kenya. *Malaria Journal* 10: 10.

18. For another example of a human population taken out of Hardy–Weinberg equlibrium, apparently by strong selection, see:

Mead, S., M. P. H. Stumpf, et al. 2003. Balancing selection at the prion protein gene consistent with prehistoric kurulike epidemics. *Science* 300: 640–643.

Hedrick, P. W. 2003. A heterozygote advantage. *Science* 302: 57.

Mead, S., J. Whitfield, et al. 2008. Genetic susceptibility, evolution and the kuru epidemic. *Philosophical Transactions of the Royal Society B* 363: 3741–3746.

19. Patients with cystic fibrosis (CF) are chronically infected with *Pseudomonas aeruginosa* bacteria. Their immune systems are engaged in a constant battle with the bacteria. In addition, they take powerful antibiotics to help keep the bacterial populations under control. Consider the consequences for the bacteria. How would you expect a *P. aeruginosa* population to evolve in the environment found inside a CF patient's lungs? What novel traits would you expect to appear? Make some predictions, then see the following paper (we are withholding the full title to avoid giving too much away):

Oliver, A., R. Cantón, et al. 2000. High frequency of . . . in cystic fibrosis lung infection. *Science* 288: 1251–1253.

20. As discussed in this chapter, the chemokine receptor CCR5 is the major means by which HIV gains entry to human white blood cells. CCR5 is also important in susceptibility to other important diseases. One example is described in the following article. Consider how CCR5's multiple role in different emerging diseases may affect its evolution, and the implications for medical treatments.

Glass, W. G., D. H. McDermott, et al. 2006. CCR5 deficiency increases risk of symptomatic West Nile virus infection. *Journal of Experimental Medicine* 203: 35–40.

CITATIONS

Much of the population genetics material in this chapter is modeled after presentations in the following:

Crow, J. F. 1983. *Genetics Notes*. Minneapolis, MN: Burgess Publishing.

Felsenstein, J. 1997. *Theoretical Evolutionary Genetics*. Seattle, WA: ASUW Publishing, University of Washington.

Felsenstein, J. 2011. *Theoretical Evolutionary Genetics*. Seattle, WA: http://evolution.genetics.washington.edu/pgbook/pgbook.html.

Griffiths, A. J. F., J. H. Miller, et al. 1993. *An Introduction to Genetic Analysis*. New York: W. H. Freeman.

Templeton, A. R. 1982. Adaptation and the integration of evolutionary forces. In *Perspectives on Evolution*, ed. R. Milkman. Sunderland, MA: Sinauer, 15–31.

Here is the list of all other citations in this chapter:

Bajmoczi, M., M. Gadjeva, et al. 2009. Cystic fibrosis transmembrane conductance regulator and caveolin-1 regulate epithelial cell internalization of *Pseudomonas aeruginosa*. *American Journal of Physiology—Cell Physiology* 297: C263–C277.

Campodónico, V. L., M. Gadjeva, et al. 2008. Airway epithelial control of *Pseudomonas aeruginosa* infection in cystic fibrosis. *Trends in Molecular Medicine* 14: 120–133.

Castle, W. E. 1903. The laws of heredity of Galton and Mendel, and some laws governing race improvement by selection. *Proceedings of the American Academy of Arts and Sciences* 39: 223–242.

Cavener, D. R., and M. T. Clegg. 1981. Multigenic response to ethanol in *Drosophila melanogaster*. *Evolution* 35: 1–10.

Chen, C.-H., H. Huang, et al. 2007. A synthetic maternal-effect selfish genetic element drives population replacement in *Drosophila*. *Science* 316: 597–600.

Cromer, D., S. M. Wolinsky, and A. R. McLean. 2010. How fast could HIV change gene frequencies in the human population? *Proceedings of the Royal Society B* 277: 1981–1989.

Darwin, C. R. 1868. *The Variation of Animals and Plants Under Domestication*. London: John Murray.

Dawson, P. S. 1970. Linkage and the elimination of deleterious mutant genes from experimental populations. *Genetica* 41: 147–169.

Dean, M., M. Carrington, et al. 1996. Genetic restriction of HIV-1 infection and progression to AIDS by a deletion allele of the CKR5 structural gene. *Science* 273: 1856–1862.

East, E. M. 1917. Hidden feeblemindedness. *Journal of Heredity* 8: 215–217.

Elias, S., M. M. Kaback, et al. 1992. Statement of the American Society of Human Genetics on cystic fibrosis carrier screening. *American Journal of Human Genetics* 51: 1443–1444.

Fisher, R. A. 1924. The elimination of mental defect. *Eugenics Review* 16: 114–116.

Foster, G. G., M. J. Whitten, et al. 1972. Chromosome rearrangements for the control of insect pests. *Science* 176: 875–880.

Gigord, L. D. B., M. R. Macnair, and A. Smithson. 2001. Negative frequency-dependent selection maintains a dramatic flower color polymorphism in the rewardless orchid *Dactylorhiza sambucina* (L.) Soò. *Proceedings of the National Academy of Sciences, USA* 98: 6253–6255.

Goddard, H. H. 1914. *Feeblemindedness: Its Causes and Consequences*. New York: Macmillan.

Gratten, J., A. J. Wilson, et al. 2008. A localized negative genetic correlation constrains microevolution of coat color in wild sheep. *Science* 319: 318–320.

Hardy, G. H. 1908. Mendelian proportions in a mixed population. *Science* 28: 49–50.

Johannesson, K., B. Johannesson, and U. Lundgren. 1995. Strong natural selection causes microscale allozyme variation in a marine snail. *Proceedings of the National Academy of Sciences, USA* 92: 2602–2606.

Kambris, Z., H. Bilak, et al. 2003. *DmMyD88* controls dorsoventral patterning of the *Drosophila* embryo. *EMBO Reports* 4: 64–69.

Karumanchi, S. A., and D. Haig. 2008. Flt1, pregnancy, and malaria: Evolution of a complex interaction. *Proceedings of the National Academy of Sciences, USA* 105: 14243–14244.

Kevles, D. J. 1995. *In the Name of Eugenics: Genetics and the Uses of Human Heredity*. Cambridge, MA: Harvard University Press.

Lane, H. 1992. *The Mask of Benevolence: Disabling the Deaf Community*. New York: Vintage Books.

López-Bigas, N., B. J. Blencowe, and C. A. Ouzounis. 2006. Highly consistent patterns for inherited human diseases at the molecular level. *Bioinformatics* 22: 269–277.

Lyczak, J. B., C. L. Cannon, and G. B. Pier. 2002. Lung infections associated with cystic fibrosis. *Clinical Microbiology Reviews* 15: 194–222.

Lyczak, J. B., and G. B. Pier. 2002. *Salmonella enterica* serovar *typhi* modulates cell surface expression of its receptor, the cystic fibrosis transmembrane conductance regulator, on the intestinal epithelium. *Infection and Immunity* 70: 6416–6423.

Martinson, J. J., N. H. Chapman, et al. 1997. Global distribution of the CCR5 gene 32-base-pair deletion. *Nature Genetics* 16: 100–103.

McKusick, Victor A., et al. 1999. Spinal muscular atrophy I. Record 253300 in *Online Mendelian Inheritance in Man*. Center for Medical Genetics, Johns Hopkins University (Baltimore, MD) and National Center for Biotechnology Information, National Library of Medicine (Bethesda, MD). Available at http://www.ncbi.nlm.nih.gov/omim/.

Muehlenbachs, A., M. Fried, et al. 2008. Natural selection of *FLT1* alleles and their association with malaria resistance *in utero*. *Proceedings of the National Academy of Sciences, USA* 105: 14488–14491.

Mukai, T., and A. B. Burdick. 1959. Single gene heterosis associated with a second chromosome recessive lethal in *Drosophila melanogaster*. *Genetics* 44: 211–232.

Paul, D. B., and H. G. Spencer. 1995. The hidden science of eugenics. *Nature* 374: 302 304.

Pier, G. B., M. Grout, and T. S. Zaidi. 1997. Cystic fibrosis transmembrane conductance regulator is an epithelial cell receptor for clearance of *Pseudomonas aeruginosa* from the lung. *Proceedings of the National Academy of Sciences, USA* 94: 12088–12093.

Pier, G. B., M. Grout, et al. 1998. *Salmonella typhi* uses CFTR to enter intestinal epithelial cells. *Nature* 393: 79–82.

Provine, W. B. 1971. *The Origins of Theoretical Population Genetics*. Chicago: University of Chicago Press.

Punnett, R. C. 1917. Eliminating feeblemindedness. *Journal of Heredity* 8: 464–465.

Ramaley, P. A., N. French, et al. 2002. Chemokine-receptor genes and AIDS risk. *Nature* 471: 140.

Reilly, P. 1991. *The Surgical Solution: A History of Involuntary Sterilization in the United States*. Baltimore, MD: Johns Hopkins University Press.

Schliekelman, P., C. Garner, and M. Slatkin. 2001. Natural selection and resistance to HIV. *Nature* 411: 545–546.

Shore, E. M., M. Xu, et al. 2006. A recurrent mutation in the BMP type I receptor ACVR1 causes inherited and sporadic fibrodysplasia ossificans progressiva. *Nature Genetics* 38: 525–527.

Smithson, A., and M. R. Macnair. 1997. Negative frequency-dependent selection by pollinators on artificial flowers without rewards. *Evolution* 51: 715–723.

Stephens, J. C., D. E. Reich, et al. 1998. Dating the origin of the AIDS-resistance allele by the coalescence of haplotypes. *American Journal of Human Genetics* 62: 1507–1515.

Sullivan, A. D., J. Wigginton, and D. Kirschner. 2001. The coreceptor mutation influences the dynamics of HIV epidemics and is selected for by HIV. *Proceedings of the National Academy of Sciences, USA* 98: 10214–10219.

Uitterlinden A. G., H. Burger, et al. 1998. Relation of alleles of the collagen type IA1 gene to bone density and the risk of osteoporotic fractures in postmenopausal women. *New England Journal of Medicine* 338: 1016–1021.

Umbers, A. J., E. H. Aitken, and S. J. Rogerson. 2011. Malaria in pregnancy: Small babies, big problem. *Trends in Parasitology* 27: 168–175.

UNAIDS. 1998. AIDS epidemic update: December 1998. (Geneva, Switzerland). Available at http://www.unaids.org.

United States Supreme Court. 1927. *Buck v. Bell*, 274 U.S. 200.

Weinberg, W. 1908. Ueber den nachweis der vererbung beim menschen. *Jahreshefte des Vereins für Vaterländische Naturkunde in Württemburg* 64: 368–382. English translation in Boyer, S. H. 1963. *Papers on Human Genetics*. Englewood Cliffs, NJ: Prentice Hall.

Wirth, B., T. Schmidt, et al. 1997. De novo rearrangements found in 2% of index patients with spinal muscular atrophy: Mutational mechanisms, parental origin, mutation rate, and implications for genetic counseling. *American Journal of Human Genetics* 61: 1102–1111.

Yule, G. U. 1902. Mendel's laws and their probable relations to intraracial heredity. *New Phytologist* 1: 193–207; 222–238.

Zar, J. H. 1996. *Biostatistical Analysis*. 3rd ed. Upper Saddle River, NJ: Prentice Hall.

Zhang, M. C., P. Azad, and R. C. Woodruff. 2011. Adaptation of *Drosophila melanogaster* to increased NaCl concentration due to dominant beneficial mutations. *Genetica* 139: 177–186.