5 The Theory of Natural Selection

This chapter introduces formal population genetic models. We first establish what the variables are that the models are concerned with, and the general structure of population genetic models. We look at the Hardy–Weinberg equilibrium, and see how to calculate whether a real population fits it. We then move on to models of natural selection, concentrating on the specific case of selection against a recessive homozygote. We apply the model to two examples: the peppered moth and resistance to pesticides. The second half of the chapter is mainly about how natural selection can maintain genetic polymorphism. We look at selection–mutation balance, heterozygous advantage, and frequency-dependent selection; and we finish by looking at models that include migration in a geographically subdivided population. The theory in this chapter all assumes that the population size is large enough for random effects to be ignored. Chapters 6 and 7 consider how random effects can interact with selection in small populations.

5.1 Population genetics is concerned with genotype and gene frequencies

The human genome, on current estimates, contains something like 30,000 gene loci. Let us focus on just one of them — on a locus at which there is more than one allele, because no evolutionary change can happen at a locus for which every individual in the population has two copies of the same allele. We shall be concerned in this chapter with models of evolution at a single genetic locus; these are the simplest models in population genetics. Chapters 8 and 9 discuss more complex models in which evolutionary change occurs simultaneously at more than one locus.

We define genotype frequency . . .

The theory of population genetics at one locus is mainly concerned to understand two closely connected variables: *gene frequency* and *genotype frequency*. They are easy to measure. The simplest case is one genetic locus with two alleles (*A* and *a*) and three genotypes (*AA*, *Aa*, and *aa*). Each individual has a genotype made up of two genes at the locus and a population can be symbolized like this:

Aa *AA* *aa* *aa* *AA* *Aa* *AA* *Aa*

This is an imaginary population with only eight individuals. To find the genotype frequencies we simply count the numbers of individual with each genotype. Thus:

Frequency of $AA = 3/8 = 0.375$
Frequency of $Aa = 3/8 = 0.375$
Frequency of $aa = 2/8 = 0.25$

In general we can symbolize genotype frequencies algebraically, as follows.

Genotype	*AA*	*Aa*	*aa*
Frequency	*P*	*Q*	*R*

P, *Q*, and *R* are expressed as percentages or proportions, so in our population, $P = 0.375$, $Q = 0.375$, and $R = 0.25$ (they have to add up to 1, or to 100%). They are measured simply by observing and counting the numbers of each type of organism in the population, and dividing by the total number of organisms in the population (the population size).

. . . and gene frequency

The gene frequency is likewise measured by counting the frequencies of each gene in the population. Each genotype contains two genes, and there are a total of 16 genes per locus in a population of eight individuals. In the population above,

Frequency of $A = 9/16 = 0.5625$
Frequency of $a = 7/16 = 0.4375$

Algebraically, we can define *p* as the frequency of *A*, and *q* as the frequency of *a*. *p* and *q* are usually called "gene" frequencies, but in a strict sense they are allele frequencies: they are the frequencies of the different alleles at one genetic locus. The gene frequencies can be calculated from the genotype frequencies:

$$p = P + \tfrac{1}{2}Q \qquad (5.1)$$
$$q = R + \tfrac{1}{2}Q$$

(and $p + q = 1$). The calculation of the gene frequencies from genotype frequencies is highly important. We shall make recurrent use of these two simple equations in the chapter. Although the gene frequencies can be calculated from the genotype frequencies (P, Q, R), the opposite is not true: the genotype frequencies cannot be calculated from the gene frequencies (p, q).

Now that we have defined the key variables, we can see how population geneticists analyze changes in those variables through time.

5.2 An elementary population genetic model has four main steps

Population geneticists try to answer the following question: if we know the genotype (or gene) frequencies in one generation, what will they be in the next generation? It is worth looking at the general procedure before going into particular models. The procedure is to break down the time from one generation to the next into a series of stages. We then work out how genotype frequencies are affected at each stage. We can begin at any arbitrarily chosen starting point in generation n and then follow the genotype frequencies through to the same point in generation $n + 1$. Figure 5.1 shows the general outline of a population genetics model.

We start with the frequencies of genotypes among the adults in generation n. The first step is to specify how these genotypes combine to breed (called a mating rule); the

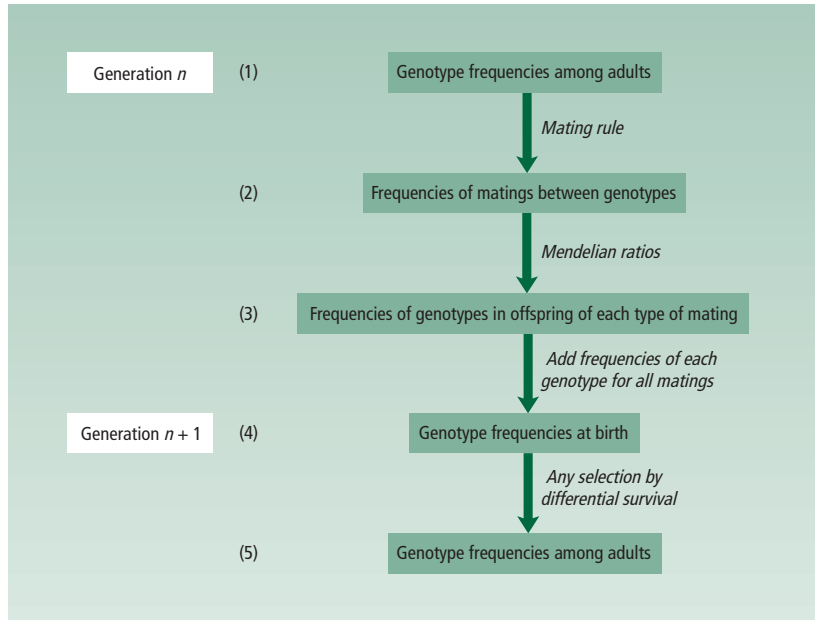

Figure 5.1
The general model of population genetics.

Population genetic models track gene frequencies over time

second step is to apply the Mendelian ratios (Chapter 2) for each type of mating; we then add the frequencies of each genotype generated from each type of mating to find the total frequency of the genotypes among the offspring, at birth, in the next generation. If the genotypes have different chances of survival from birth to adulthood, we multiply the frequency of each genotype at birth by its chance of survival to find the frequency among the adults. When the calculation at each stage has been completed, the population geneticist's question has been answered.

Natural selection can operate in two ways: by differences in survival among genotypes or by differences in fertility. There are two theoretical extremes. At one, the surviving individuals of all genotypes produce the same number of offspring, and selection operates only on survival; at the other, individuals of all genotypes have the same survival, but differ in the number of offspring they produce (that is, their fertility). Both kinds of selection probably operate in many real cases, but the models we shall consider in this chapter all express selection in terms of differences in chance of survival. This is not to suggest that selection always operates only on survival; it is to keep the models simple and consistent.

The model, in the general form of Figure 5.1, may look rather complicated. However, we can cut it down to size by making some simplifying assumptions. The first two simplifying assumptions to consider are random mating and no selection (no differences in survival between genotypes from stages 4 to 5).

5.3 Genotype frequencies in the absence of selection go to the Hardy–Weinberg equilibrium

We can stay with the case of one genetic locus with two alleles (*A* and *a*). The frequencies of genotypes *AA*, *Aa*, and *aa* are *P*, *Q*, and *R*. Our question is, if there is random mating and no selective difference among the genotypes, and we know the genotype frequencies in one generation, what will the genotype frequencies be in the next generation? The answer is called the *Hardy–Weinberg equilibrium*. Let us see what that means.

We deduce the frequencies of pairings, with random pairing . . .

Table 5.1 gives the calculation. The mating frequencies follow from the fact that mating is random. To form a pair, we pick out at random two individuals from the population. What is the chance of an $AA \times AA$ pair? Well, to produce this pair, the first individual we pick has to be an *AA* and the second one also has to be an *AA*. The chance that the first is an *AA* is simply *P*, the genotype's frequency in the population. In a large population, the chance that the second one is *AA* is also *P*.[1] The chance of drawing out two *AA* individuals in a row is therefore P^2. (The frequency of $Aa \times Aa$ and $aa \times aa$ matings are likewise Q^2 and R^2, respectively.) Similar reasoning applies for the frequencies of matings in which the two individuals have different genotypes. The chance of

[1] "Large" populations are not a separate category from "small" ones; populations come in all sizes. The random effects we consider in Chapter 6 become increasingly important as a population becomes smaller. However, one rough definition of a large population is one in which the sampling of one individual to form a mating pair does not affect the genotype frequencies in the population: if one *AA* is taken out, the frequency of *AA* in the population, and the chance of picking another *AA*, remains effectively *P*.

Table 5.1
Calculations needed to derive the Hardy–Weinberg ratio for one locus and two alleles, *A* and *a*. (Frequency of *AA* = *P*, of *Aa* = *Q*, and of *aa* = *R*.) The table shows the frequencies of different matings if the genotypes mate randomly, and the genotype proportions among the progeny of the different matings.

Mating type	Frequency of mating	Offspring genotype proportions
$AA \times AA$	P^2	1 *AA*
$AA \times Aa$	PQ	$\frac{1}{2}$ *AA* : $\frac{1}{2}$ *Aa*
$AA \times aa$	PR	1 *Aa*
$Aa \times AA$	QP	$\frac{1}{2}$ *AA* : $\frac{1}{2}$ *Aa*
$Aa \times Aa$	Q^2	$\frac{1}{4}$ *AA* : $\frac{1}{2}$ *Aa* : $\frac{1}{4}$ *aa*
$Aa \times aa$	QR	$\frac{1}{2}$ *Aa* : $\frac{1}{2}$ *aa*
$aa \times AA$	RP	1 *Aa*
$aa \times Aa$	RQ	$\frac{1}{2}$ *Aa* : $\frac{1}{2}$ *aa*
$aa \times aa$	R^2	1 *aa*

picking an *AA* and then an *Aa* (to produce an $AA \times Aa$ pair), for example, is *PQ*; the chance of picking an *AA* and then an *aa* is *PR*; and so on.

. . . and use Mendel's rules to deduce the genotype frequencies in the offspring

The genotypic proportions in the offspring of each type of mating are given by the Mendelian ratios for that cross. We can work out the frequency of a genotype in the next generation by addition. We look at which matings generate the genotype, and add the frequencies generated by all the matings. Let us work it out for the genotype *AA*. *AA* individuals, Table 5.1 shows, come from $AA \times AA$, $AA \times Aa$ (and $Aa \times AA$), and $Aa \times Aa$ matings. We can ignore all the other types of mating. $AA \times AA$ matings have frequency P^2 and produce all *AA* offspring, $AA \times Aa$ and $Aa \times AA$ matings each have frequency PQ and produce 50% *AA* offspring, and $Aa \times Aa$ matings have frequency Q^2 and produce 25% *AA* offspring. The frequency of *AA* in the next generation,[2] P', is then:

$$P' = P^2 + \tfrac{1}{2}PQ + \tfrac{1}{2}PQ + \tfrac{1}{4}Q^2 \qquad (5.2)$$

This can be rearranged to:

$$P' = (P + \tfrac{1}{2}Q)\,(P + \tfrac{1}{2}Q)$$

We have seen that $(P + \frac{1}{2}Q)$ is simply the frequency of the gene *A*, *p*. Therefore:

$$P' = p^2$$

[2] Population geneticists conventionally symbolize the frequency of variables one generation on by writing a prime. If *P* is the frequency of genotype *AA* in one generation, P' is its frequency in the next; if *p* is the frequency of an allele in one generation, p' is its frequency in the next generation. We shall follow this convention repeatedly in this book.

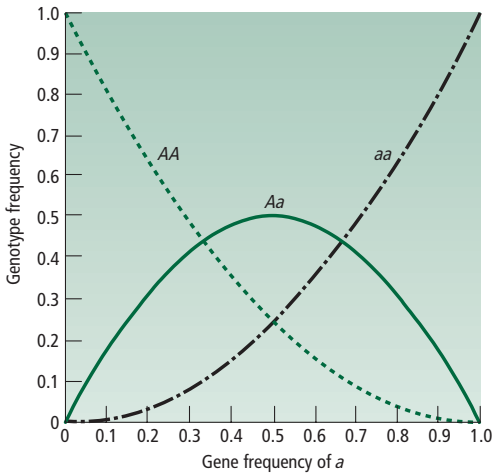

Figure 5.2
Hardy–Weinberg frequencies of genotypes *AA*, *Aa*, and *aa* in relation to the frequency of the gene *a* (q).

The result is the Hardy–Weinberg equilibrium

The frequency of genotype *AA* after one generation of random mating is equal to the square of the frequency of the *A* gene. Analogous arguments show that the frequencies of *Aa* and *aa* are $2pq$ and q^2. The Hardy–Weinberg frequencies are then:

Genotype	AA : Aa : aa
Frequency	p^2 : $2pq$: q^2

Figure 5.2 shows the proportions of the three different genotypes at different frequencies of the gene *a*; heterozygotes are most frequent when the gene frequency is 0.5.

The Hardy–Weinberg genotype frequencies are reached after a single generation of random mating from any initial genotype frequencies. Imagine, for example, two populations with the same gene frequency but different genotype frequencies. One population has 750 *AA*, 0 *Aa*, and 250 *aa*; the other has 500 *AA*, 500 *Aa*, and 0 *aa*. $p = 0.75$ and $q = 0.25$ in both. After one generation of random mating, the genotype frequencies in both will become 563 *AA*, 375 *Aa*, and 62 *aa* if the population size remains 1,000. (Fractions of an individual have been rounded to make the numbers add to 1,000. The proportions are 9/16, 6/16, and 1/16.) After reaching those frequencies immediately, in one generation, the population stays at the Hardy–Weinberg equilibrium for as long as the population size is large, there is no selection, and mating is random.

As we saw in Section 5.1, it is not in general possible to calculate the genotype frequencies in a generation if you only know the gene frequencies. We can now see that it is possible to calculate, from gene frequencies alone, what the genotype frequencies will be in the next generation, provided that mating is random, there is no selection, and the population is large. If the gene frequencies in this generation are p and q, in the next generation the genotype will have Hardy–Weinberg frequencies.

The proof of the Hardy–Weinberg theorem we have worked through was long-winded. We worked though it all in order to illustrate the general model of population genetics in its simplest case. However, for the particular case of the Hardy–Weinberg equilibrium, a more elegant proof can be given in terms of gametes.

A simpler proof of the Hardy–Weinberg equilibrium

Diploid organisms produce haploid gametes. We could imagine that the haploid gametes are all released into the sea, where they combine at random to form the next generation. This is called random union of gametes. In the "gamete pool" A gametes will have frequency p and a gametes frequency q. Because they are combining at random, an a gamete will meet an A gamete with chance p and an a gamete with chance q. From the a gametes, Aa zygotes will therefore be produced with frequency pq and aa gametes with frequency q^2. A similar argument applies for the A gametes (which have frequency p): they combine with a gametes with chance q, to produce Aa zygotes (frequency pq) and A gametes with chance p to form AA zygotes (frequency p^2). If we now add up the frequencies of the genotypes from the two types of gamete, the Hardy–Weinberg genotype frequencies emerge. We have now derived the Hardy–Weinberg theorem for the case of two alleles; the same argument easily extends to three or more alleles (Box 5.1).

(Some people may be puzzled by the 2 in the frequency of the heterozygotes. It is a simple combinatorial probability. Imagine flipping two coins and asking what the chances are of flipping two heads, or two tails, or one head and one tail. The chance of two heads is $(1/2)^2$ and of two tails $(1/2)^2$; the chance of a head and a tail is $2 \times (1/2)^2$, because a tail then a head, and a head then a tail, both give one head and one tail. The head is analogous to allele A, the tail to a; two heads to producing an AA genotype, and one head and one tail to a heterozygote Aa. The coin produces heads with probability 1/2, and is analogous to a gene frequency of $p = 1/2$. The frequency $2pq$ for heterozygotes is analogous to the chance of one head and one tail, $2 \times (1/2)^2$. The 2 arises because there are two ways of obtaining one head and one tail. Likewise there are two ways of producing an Aa heterozygote: either the A gene can come from the father and the a from the mother, or the a gene from the father and the A from the mother. The offspring is Aa either way.)

Box 5.1
The Hardy–Weinberg Theorem for Three Alleles

We can call the three alleles A_1, A_2, and A_3, and define their gene frequencies as p, q, and r, respectively. We form new zygotes by sampling two successive gametes from a large pool of gametes. The first gamete we pick could be A_1, A_2, or A_3. If we first pick (with chance p) an A_1 allele from the gamete pool, the chance that the second allele is another A_1 allele is p, the chance that it is an A_2 allele is q, and the chance that it is an A_3 allele is r: from these three, the frequencies of A_1A_1, A_1A_2, and A_1A_3 zygotes are p^2, pq, and pr.

Now suppose that the first allele we picked out had been an A_2 (which would happen with chance q). The chances that the second allele would again be A_1, A_2, or A_3would be p, q, and r, respectively, giving A_1A_2, A_2A_2, and A_2A_3 zygotes in frequency pq, q^2, and qr.

Finally, if we had picked (with chance r) an A_3 allele, we produce A_1A_3, A_2A_3, and A_3A_3 zygotes in frequency pr, qr, and r^2.

The only way to form the homozygotes A_1A_1, A_2A_2, and A_3A_3 is by picking two of the same kind of gamete and the frequencies are p^2, q^2, and r^2. The heterozygotes can be formed from more than one kind of first gamete and their frequencies are obtained by addition. The total chance of forming an A_1A_3 zygote is $pr + rp = 2pr$; of forming an A_1A_2 zygote is $pq + qp = 2pq$; and of an A_2A_3 zygote is $2qr$. The complete Hardy–Weinberg proportions are:

A_1A_1	A_1A_2	A_1A_3	A_2A_2	A_2A_3	A_3A_3
p^2	$2pq$	$2pr$	q^2	$2qr$	r^2

5.4 We can test, by simple observation, whether genotypes in a population are at the Hardy–Weinberg equilibrium

Natural populations may or may not fit the Hardy–Weinberg equilibrium

The Hardy–Weinberg theorem depends on three main assumptions: no selection, random mating, and large population size. In a natural population, any of these could be false; we cannot assume that natural populations will be at the Hardy–Weinberg equilibrium. In practice, we can find out whether a population is at the Hardy–Weinberg equilibrium for a locus simply by counting the genotype frequencies. From those frequencies, we first calculate the gene frequencies; then, if the observed homozygote frequencies equal the square of their gene frequencies, the population is in Hardy–Weinberg equilibrium. If they do not, it is not.

The MN blood group system in humans is a good example, because the three genotypes are distinct and the genes have reasonably high frequencies in human populations. Three phenotypes, *M*, *MN*, and *N* are produced by three genotypes (*MM*, *MN*, *NN*) and two alleles at one locus. The phenotypes of the MN group, like the better known ABO group, are recognized by reactions with antisera. The antisera are made by injecting blood into a rabbit, which then makes an antiserum to the type of blood that was injected. If the rabbit has been injected with M-type human blood, it produces

Table 5.2
The frequencies of the MM, MN, and NN blood groups in three American populations. The figures for expected proportions and numbers have been rounded.

Population		MM	MN	NN	Total	Frequency *M*	Frequency *N*
African Americans	Observed number	79	138	61	278		
	Expected proportion	0.283	0.499	0.219		0.532	0.468
	Expected number	78.8	138.7	60.8			
European Americans	Observed number	1,787	3,039	1,303	6,129		
	Expected proportion	0.292	0.497	0.211		0.54	0.46
	Expected number	1,787.2	3,044.9	1,296.9			
Native Americans	Observed number	123	72	10	205		
	Expected proportion	0.602	0.348	0.05		0.776	0.224
	Expected number	123.3	71.4	10.3			

Specimen calculation for African Americans:

Frequency of *M* allele = $79 + (\frac{1}{2} \times 138) = 0.532 = p$
Frequency of *N* allele = $61 + (\frac{1}{2} \times 138) = 0.468 = q$

Expected proportion of MM = $p^2 = (0.532)^2 = 0.283$
Expected proportion of MN = $2pq = 2(0.532)\,(0.468) = 0.499$
Expected proportion of NN = $q^2 = (0.468)^2 = 0.219$

Expected numbers = expected proportion × total number (n)

Expected number of MM = $p^2n = 0.283 \times 278 = 78.8$
Expected number of MN = $2pqn = 0.499 \times 278 = 138.7$
Expected number of NN = $q^2n = 0.219 \times 278 = 60.8$

anti-M serum. Anti-M serum agglutinates blood from humans with one or two *M* alleles in their genotypes; likewise anti-N blood agglutinates the blood of humans with one or two *N* alleles. Therefore *MM* individuals are recognized as those whose blood reacts only with anti-M, *NN* individuals react only with anti-N, and *MN* individuals react with both.

The MN human blood group system is close to Hardy–Weinberg equilibrium

Table 5.2 gives some measurements of the frequencies of the MN blood group genotypes for three human populations. Are they at Hardy–Weinberg equilibrium? In European Americans, the frequency of the *M* gene (calculated from the usual $p = P + {}^1/_2Q$ relation) is 0.54. If the population is at the Hardy–Weinberg equilibrium, the frequency of *MM* homozygotes (p^2) will be $2 \times 0.54 = 0.2916$ (1,787 in a sample of 6,129 individuals); and the frequency of *MN* heterozygotes ($2pq$) will be $2 \times 0.54 \times 0.46 = 0.497$ (3,045 in a sample of 6,129). As the table shows, these are close to the observed frequencies. In fact all three populations are at Hardy–Weinberg equilibrium. We shall see in Section 5.6 that the same calculations do not correctly predict the genotype frequencies after selection has operated.

5.5 The Hardy–Weinberg theorem is important conceptually, historically, in practical research, and in the workings of theoretical models

We have just seen how to find out whether a real population is in Hardy–Weinberg equilibrium. The importance of the Hardy–Weinberg theorem, however, is not mainly as an empirical prediction. We have no good reason to think that genotypes in natural populations will generally have Hardy–Weinberg frequencies, because it would require both no selection and random mating, which are rarely found. The interest of the theorem lies elsewhere, in three other areas.

The Hardy–Weinberg theorem matters conceptually, . . .

One is historical and conceptual. We saw in Section 2.9 (p. 37) how with blending inheritance the genetic variation in a population is rapidly blended out of existence and the population becomes genetically uniform. With Mendelian genetics, variation is preserved and the Hardy–Weinberg theorem gives quantitative demonstration of that fact. The theorem was published in the first decade of the twentieth century, as Mendelism was becoming accepted, and it was historically influential in proving to people that Mendelian inheritance did allow variation to be preserved.

. . . in research . . .

A second interest of the theorem is as a kind of springboard, that launches us toward interesting empirical problems. If we compare genotype frequencies in a real population with Hardy–Weinberg ratios, then if they deviate it suggests something interesting (such as selection or non-random mating) may be going on, which would merit further research.

. . . and in theory

A third interest is theoretical. In the general model of population genetics (Section 5.2) there were five stages, joined by four calculations. The Hardy–Weinberg theorem simplifies the model wonderfully. If we assume random mating, we can go directly from the adult frequencies in generation n to the genotype frequencies at birth in generation $n + 1$, collapsing three calculations into one (Figure 5.3). If we know the adult genotype frequencies in generation n (stage 1), we only need to calculate the gene

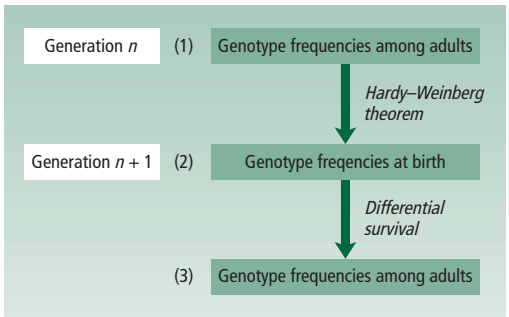

Figure 5.3
The general model of population genetics simplified by the Hardy–Weinberg theorem.

frequencies: the genotype frequencies at birth in the next generation (stage 2) must then have Hardy–Weinberg frequencies, because the gene frequencies do not change between the adults of one generation and the newborn members of the next generation. A simple model of selection can concentrate on how the genotype frequencies are modified between birth and the adult reproductive stage (from stage 2 to stage 3 of Figure 5.3).

5.6 The simplest model of selection is for one favored allele at one locus

We shall start with the simplest case. It is the case of natural selection operating on only one genetic locus, at which there are two alleles, one dominant to the other. Suppose that individuals with the three genotypes have the following relative chances of survival from birth to the adult stage:

Genotype	Chance of survival
AA, *Aa*	1
aa	$1 - s$

Population genetic models specify the fitness of all genotypes

s is a number between 0 and 1, and is called the *selection coefficient*. Selection coefficients are expressed as reductions in fitness relative to the best genotype. If *s* is 0.1 then *aa* individuals have a 90% chance of survival, relative to 100% for *AA* and *Aa* individuals. These are relative values: in a real case the chance of survival from birth to reproduction of an individual with the best genotype might be 50%, much less than 100%. If it was 50%, then an *s* of 0.1 would mean that *aa* individuals really had a 45% chance of survival. (The convention of giving the best genotype a relative 100% chance of survival simplifies the algebra. If you are suspicious, check whether it makes any difference in what follows if the chances of survival are 50%, 50%, and 45% for *AA*, *Aa*, and *aa*, respectively, rather than 100%, 100%, and 90%.) The chance of survival is the *fitness* of the genotype (we are assuming that all surviving individuals produce the same number of offspring). Fitnesses are, like the chances of survival, expressed relative to a figure of 1 for the best genotype. This can be spelled out more by referring to fitnesses as

Table 5.3
(a) Algebraic calculation of genotype frequences after selection, with selection against a recessive genotype. (b) A numerical illustration. See text for further explanation.

(a)

	Genotype		
	AA	***Aa***	***aa***
Birth			
Frequency	p^2	$2pq$	q^2
Fitness	1	1	$1 - s$
Adult			
Relative frequency	p^2	$2pq$	$q^2(1 - s)$
Frequency	$p^2/(1 - sq^2)$	$2pq/(1 - sq^2)$	$q^2(1 - s)/(1 - sq^2)$

(b)

	Genotype			
	AA	***Aa***	***aa***	**Total**
Birth				
Number	1	18	81	100
Frequency	0.01	0.18	0.81	
Fitness	1	1	0.9	
Adult				
Number	1	18	73	92
Frequency	1/92	18/92	73/92	

"relative fitnesses." However, biologists usually just say "fitness." With the fitnesses given above, selection will act to eliminate the *a* allele and *fix* the *A* allele. (To "fix" a gene is genetic jargon for carry its frequency up to 1. When there is only one gene at a locus, it is said to be "fixed" or in a state of "fixation.") If s were 0, the model would lapse back to the Hardy–Weinberg case and the gene fequencies would be stable.

Notice that alleles do not have any tendency to increase in frequency just because they are dominant, or to decrease because they are recessive. Dominance and recessivity only describe how the alleles at a locus interact to produce a phenotype. Changes in gene frequency are set by the fitnesses. If the recessive homozygote has higher fitness, the recessive allele will increase in frequency. If, as here, the recessive homozygote has lower fitness, the recessive allele decreases in frequency.

We construct a model for the change in gene frequency per generation

How rapidly will the population change through time? To find out, we seek an expression for the gene frequency of A (p') in one generation in terms of its frequency in the previous generation (p). The difference between the two, $\Delta p = p' - p$, is the change in gene frequency between two successive generations. The model has the form of Figure 5.3, and we shall work through both the general algebraic version and a numerical example (Table 5.3).

To begin with, at birth the three genotypes have Hardy–Weinberg frequencies as they are produced by random mating among adults of the previous generation. Selection then operates; *aa* individuals have a lower chance of survival and their frequency among the adults is reduced. As the numerical example shows (Table 5.3b), the total number of adults is less than the number at birth and we have to divide the adult numbers of each genotype by the total population size to express the adult numbers as frequencies comparable to the frequencies at birth. In the algebraic case, the relative frequencies after selection do not add up to 1, and we correct them by dividing by the *mean fitness*.

$$\text{Mean fitness} = p^2 + 2pq + q^2(1-s) = 1 - sq^2 \tag{5.3}$$

Dividing by mean fitness in the algebraic case is the same as dividing by the population size after selection in the numerical example. Notice that now the adult genotype frequencies are not in Hardy–Weinberg ratios. If we tried to predict the proportion of *aa* from q^2, as in the MN blood group (Section 5.4), we should fail. The frequency of *aa* is $q^2(1-s)/1-sq^2$, not q^2.

What is the relation between p' and p? Remember that the frequency of the gene *A* at any time is equal to the frequency of *AA* plus half the frequency of *Aa*. We have just listed those frequencies in the adults after selection:

$$p' = \frac{p^2 + pq}{1 - sq^2} = \frac{p}{1 - sq^2} \tag{5.4}$$

(remember $p + q = 1$, and therefore $p^2 + pq = p(p + q) = p$.) The denominator $1 - sq^2$ is less than 1, because s is positive, so p' is greater than p: selection is increasing the frequency of the *A* gene. We can now derive a result for Δp, the change in gene frequency in one generation. The algebra looks like this.

$$\begin{aligned}\Delta p = p' - p &= \frac{p}{1 - sq^2} - p \\ &= \frac{p - p + spq^2}{1 - sq^2} \\ &= \frac{spq^2}{1 - sq^2}\end{aligned} \tag{5.5}$$

For example, if $p = q = 0.5$ and *aa* individuals have fitness 0.9 compared with *AA* and *Aa* individuals ($s = 0.1$) then the change in gene frequency to the next generation will be $(0.1 \times 0.5 \times (0.5)^2)/(1 - 0.1 \times (0.5)^2) = 0.0128$; the frequency of *A* will therefore increase to 0.5128.

The model predicts the rate of change in gene frequency as the superior gene is fixed

We can use this result to calculate the change in gene frequency between successive generations for any selection coefficient (s) and any gene frequency. The result in this simple case is that the *A* gene will increase in frequency until it is eventually fixed (that is, has a frequency of 1). Table 5.4 illustrates how gene frequencies change when selection acts against a recessive allele, for each of two selection coefficients. There are two points to notice in the table. One is the obvious one that with a higher selection coefficient against the *aa* genotype, the *A* gene increases in frequency more rapidly. The other is the more interesting observation that the increase in the frequency of *A* slows down when it becomes common, and it would take a long time finally to eliminate the *a* gene. This is because the *a* gene is recessive. When *a* is rare it is almost always found in *Aa* individuals, who are selectively equivalent to *AA* individuals: selection can no longer "see" the *a* gene, and it becomes more and more difficult to eliminate them. Logically, selection cannot eliminate the one final *a* gene from the population, because if there is only one copy of the gene it must be in a heterozygote.

Table 5.4
A simulation of changes in gene frequency for selection against the recessive gene *a*, using two selection coefficients: $s = 0.05$ (i.e., *aa* individuals have a relative chance of survival of 95%, against 100% for *AA* and *Aa*) and $s = 0.01$ (i.e., *aa* individuals have a relative chance of survival of 99%, against 100% for *AA* and *Aa*). The change between generation 0 and 100 is found by applying the equation in the text 100 times successively.

	Gene frequency, $s = 0.05$		Gene frequency, $s = 0.01$	
Generation	***A***	***a***	***A***	***a***
0	0.01	0.99	0.01	0.99
100	0.44	0.56	0.026	0.974
200	0.81	0.19	0.067	0.933
300	0.89	0.11	0.15	0.85
400	0.93	0.07	0.28	0.72
500	0.95	0.05	0.43	0.57
600	0.96	0.04	0.55	0.45
700	0.96	0.04	0.65	0.35
800	0.97	0.03	0.72	0.28
900	0.97	0.03	0.77	0.23
1,000	0.98	0.02	0.80	0.20

Just as equation 5.4 can be used to calculate a gene frequency change given the fitnesses, so it can be used to calculate the fitnesses given the frequency changes. If we know the gene frequency in two successive generations then equations 5.4 and 5.5 can be rearranged to:

$$s = \frac{\Delta p}{p' q^2} \tag{5.6}$$

to find *s*.

We need to know more to understand completely the rate of evolution of whole organs

Haldane (1924) first produced this particular model of selection. One important feature of the model is that it shows how rapidly, in evolutionary time, natural selection can produce change. When we look at the complex organs and behavior patterns of living creatures, including ourselves, it is easy to wonder whether there has really been enough time for them to have evolved in the manner suggested by Darwin's theory. To find out, for any particular organ, such as the heart, liver, or brain, we need answers to two questions: (i) how many genetic changes did its evolution require; and (ii) how long did each change take.

A model like the one in this section gives us an idea of the answer to the second question. (We shall look more at the first question in Section 10.5, p. 266.) The fitness differences of 1–5% in Table 5.4 are small, relative to many of the risks we take though our lives; but they are enough to carry a gene up from being negligibly rare to being the

majority form in the population in 1,000 to 10,000 generations. On the evolutionary timescale, 10,000 generations are an eye-blink: too short a period to be resolved in the fossil record. A quantitative model such as Haldane's was needed to answer the quantitative question of how rapidly selection can drive evolution.

The model can be extended

The model can be extended in various ways. The modifications for different degrees of dominance, and separate selection on heterozygotes and homozygotes, are conceptually straightforward, though they make the algebra more complex. Other modifications can be made to analyze the other stages in the general picture of Figure 5.1: to analyze non-random mating, non-Mendelian inheritance, or fitnesses that vary according to fertility rather than survival. However, for our purposes it is mainly important to see how an exact model of selection can be built and exact predictions made from it. The model is simplified, but it can help us to understand a number of real cases — as we shall now see.

5.7 The model of selection can be applied to the peppered moth

5.7.1 *Industrial melanism in moths evolved by natural selection*

The peppered moth *Biston betularia* provides one of the best known stories in evolutionary biology (Figure 5.4). In collections made in Britain in the eighteenth century, the form of the moth was always a light, peppered color. A dark (melanic) form was first recorded in 1848 near Manchester. That melanic form then increased in frequency until it made up more than 90% of the populations in polluted areas in the mid-twentieth century. In unpolluted areas, the light form remained common. Clean air laws were passed in the mid-twentieth century, and the frequency of the melanic form decreased in formerly polluted areas.

We can estimate the fitness differences during peppered moth evolution

The peppered moth can be used to illustrate the simple model of the previous section. A controversy has grown up about the peppered moth concerning the reason why the melanic and light-colored moths differed in fitness, although this does not matter while we are simply estimating fitnesses. The increase in frequency of the melanic form in polluted areas has classically been explained by bird predation. Some doubts have been raised about the evidence for this view. Section 5.7.4 looks at the controversy, but we begin by looking at estimates of fitness. All we need to know for these estimates is that natural selection is acting — just how it is acting, whether by bird predators or other factors, is another question.

Before we can apply the theory of population genetics to a character, we need to know its genetics. Breeding experiments initially suggested that the difference in color was controlled by one main locus. The original, peppered form was one homozygote (*cc*) and the melanic form was another homozygote (*CC*), and the *C* allele is dominant. However, in other experiments the melanic allele was less dominant and the heterozygotes were intermediate; there seem to be a number of different melanic alleles. It may be that selection initially favored a melanic allele with no or weak dominance, and subsequently some other melanic alleles with stronger dominance. In any case, the degree of dominance of the melanic allele that was originally favored in the nineteenth century

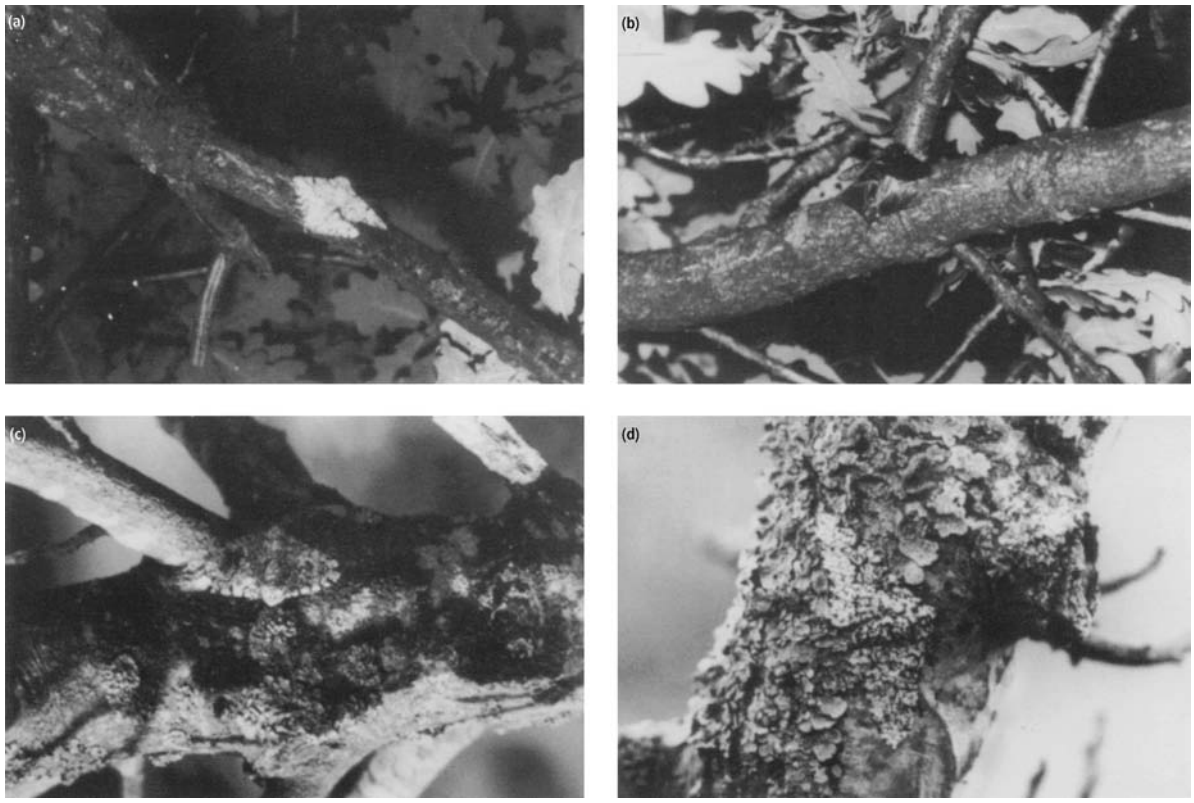

Figure 5.4
Peppered moths naturally settle on the undersides of twigs in higher branches of trees (and not on tree trunks, as is sometimes said). Melanic forms are better camouflaged in polluted areas: compare (a) the peppered form and (b) the melanic form, both photographed in a polluted area. (c) and (d) show that peppered forms are well camouflaged in unpolluted areas. Reprinted, by permission of the publisher, from Brakefield (1987).

is uncertain, and it may have differed from the dominance shown by the melanic alleles that exist in modern populations.

The first estimates of fitnesses were made by Haldane (1924), and he dealt with the problem of varying degrees of dominance by making two estimates of fitness, one assuming that the *C* allele is dominant and the other assuming that the heterozygote is intermediate. The real average degree of dominance was probably between the two. Here we shall look only at the estimate for a dominant *C* gene.

5.7.2 *One estimate of the fitnesses is made using the rate of change in gene frequencies*

What were the relative fitnesses of the genes controlling the melanic and light coloration during the phase from the early nineteenth to the mid-twentieth centuries, while the melanic form increased in frequency in polluted areas? For the first method we need

measurements of the frequencies of the different color forms for at least two times. We can then estimate the gene frequencies from the genotype frequencies, and substitute them in equation 5.6 to solve s, the selection coefficient.

The observed gene frequency changes suggest $s \approx 0.33$

The melanic form was first seen in 1848; but it was probably not a new mutation then. It probably existed at a low frequency in the population, in what is called "mutation–selection balance." Mutation–selection balance means that the gene is disadvantageous, and exists at a low frequency determined by a balance between being formed by mutation and being lost by selection (Section 5.11). We shall see that the frequency of a gene can be calculated from its mutation rate m and its selective disadvantage s. The values of m and s are unknown for the gene in the early nineteenth century. However, typical mutation rates for genes are about 10^{-6} and a selective disadvantage of about 10% for the melanic mutants in preindustrial times may be approximately correct. With these figures, and using equation 5.9 below, the melanic C gene would have had a frequency of 10^{-5} up to the year 1848. By 1898, the frequency of the light-colored genotype was 1–10% in polluted areas (it was not more than 5% near the industrial city of Manchester, for example, implying a gene frequency of about 0.2). There would have been about 50 generations between 1848 and 1898.

We now know all we need. What selective coefficient would generate an increase in its frequency from 10^{-5} to 0.8 in 50 generations? Equation 5.6 gives the selection coefficient in terms of gene frequencies in two successive generations, but between 1848 and 1898 there would have been 50 generations. The formula therefore has to be applied 50 times over, which is most easily done by computer. A change from 10^{-5} to 0.8 in 50 generations, it turns out, requires $s \approx 0.33$: the peppered moths had two-thirds the survival rate of melanic moths (Table 5.5). The calculations are rough, but they show how fitness can be inferred from the observed rate of change in gene frequency.

Table 5.5
Theoretical changes in gene frequencies in the evolution of melanism in the peppered moth, starting with an initial frequency of C of 0.00001 (rounded to 0 in the table). C is dominant and c is recessive: genotypes CC and Cc are melanic and cc is peppered in color. 1848 is generation zero in the simulation. Selection coefficient $s = 0.33$.

	Gene frequency	
Generation date	*C*	*c*
1848	0.00	1.00
1858	0.00	1.00
1868	0.03	0.97
1878	0.45	0.55
1888	0.76	0.24
1898	0.86	0.14
1908	0.90	0.10
1918	0.92	0.08
1928	0.94	0.06
1938	0.96	0.04
1948	0.96	0.04

Table 5.6
Frequencies of melanic and light peppered moths in samples recaptured at two sites in the UK: Birmingham (polluted) and Deanend Wood, Dorset (unpolluted). The observed numbers are the actual numbers recaught; the expected numbers are the numbers that would have been recaught if all morphs survived equally (equals proportion in released moths times the number of moths recaptured). The recaptured moths at Birmingham were taken over a period of about 1 week, at Deanend Wood over about 3 weeks. Data from Kettlewell (1973).

	Light moths	Melanic moths
Birmingham (polluted)		
Numbers recaptured		
Observed	18	140
Expected	36	122
Relative survival rate	0.5	1.15
Relative fitness	5/1.15 = 0.43	1.15/1.15 = 1
Deanend wood (unpolluted)		
Numbers recaptured		
Observed	67	32
Expected	53	46
Relative survival rate	1.26	0.69
Relative fitness	1.26/1.26 = 1	0.69/1.26 = 0.55

5.7.3 *A second estimate of the fitnesses is made from the survivorship of the different genotypes in mark–recapture experiments*

Mark–recapture experiments suggest $s \approx 0.57$

The estimate of fitness can be checked against other estimates. The gene frequency change was (and still is) thought to be produced by survival differences between the two forms of moth in nature, rather than differential fertility. We can measure the rate of survival of the two forms in nature, and see how they differ. Kettlewell (1973) measured survival rates by mark–recapture experiments in the field. He released melanic and light-colored peppered moths in known proportions in polluted and unpolluted regions, and then later recaught some of the moths (which are attracted to mercury vapor lamps). He counted the proportions of melanic and light-colored moths in the moths recaptured from the two areas.

Table 5.6 gives some results for two sites, Birmingham (polluted) and Deanend Wood, an unpolluted forest in Dorset, UK. The proportions in the recaptured moths are as we would expect: more light-colored moths in the Deanend Wood samples and more melanic moths in the Birmingham samples. In Birmingham, melanic moths were recaptured at about twice the rate of light-colored ones, implying $s = 0.57$. This is a higher fitness difference than the $s = 0.33$ implied by the change in gene frequency.

The discrepancy is unsurprising because both estimates are uncertain; it could have a number of causes. Possible causes include sampling error in the mark–recapture experiments (the numbers in Table 5.6 are small) and errors in the assumptions of the estimate from gene frequency changes. For instance, the initial gene frequency may have been less than 10^{-5}. Also, the relative fitness of the two moth forms probably changed over time and moths may have migrated between polluted and unpolluted areas. Whatever the cause of the discrepancy, the two calculations do illustrate two important methods of estimating fitness.

5.7.4 *The selective factor at work is controversial, but bird predation was probably influential*

So far we have concentrated on estimating fitnesses, and have ignored the factors that cause the fitness difference between the melanic and light-colored forms of the moth. The material thus far is uncontroversial. The gene frequency changes have undoubtedly occurred, and provide an excellent example of evolution by natural selection. Now we can move on to ask what the agent, or agents, of natural selection were in this example.

Peppered moth evolution is classically explained by bird predation

The classic answer, due to the research of Kettlewell (1973), has been bird predation. The light-colored form is better camouflaged in unpolluted woods and therefore less likely to be eaten by visually hunting birds. But smoke pollution killed the lichens that covered the trees, after which the melanic form was better camouflaged (Figure 5.4). Several lines of evidence support Kettlewell's explanation. Birds do eat the moths, and have been photographed in the act. Birds also have been shown to take more of the poorly camouflaged form, in various experimental set-ups. Also, the gene frequency changes closely match the rise and fall of air pollution. The melanic form increased in frequency following the industrial revolution, and then decreased in frequency after air pollution decreased in the late twentieth century. Indeed, the case for Kettlewell's explanation is arguably now stronger than when he worked. The decrease in frequency of the melanic form has become particularly clear from 1970 to 2000, adding a new line of evidence that was unavailable to Kettlewell (whose main work was in the 1950s).

Some of the classic experiments have been criticized

However, not everyone accepts that bird predation is the selective agent. Some of Kettlewell's research has itself been criticized. We looked above at fitness estimates from gene frequency changes and from mark–recapture experiments. Kettlewell and others also estimated fitnesses by pinning out dead moths of the two forms on tree trunks in polluted and unpolluted areas. He then measured how many moths of each form disappeared over time. These experiments were particularly criticized after it was discovered in the 1980s that peppered moths do not naturally settle on tree trunks, but on the higher branches and twigs of trees (Figure 5.4). Other criticisms were also made. However, Kettlewell's case does not depend on these pin-out experiments. As we saw, he also did mark–recapture experiments in which he released live moths. Those moths presumably settled, and behaved, in a natural manner. The results of all the experiments — pin-outs and mark–recapture — were similar, so the fact that the moths were pinned out in the wrong place did not bias the fitness estimates.

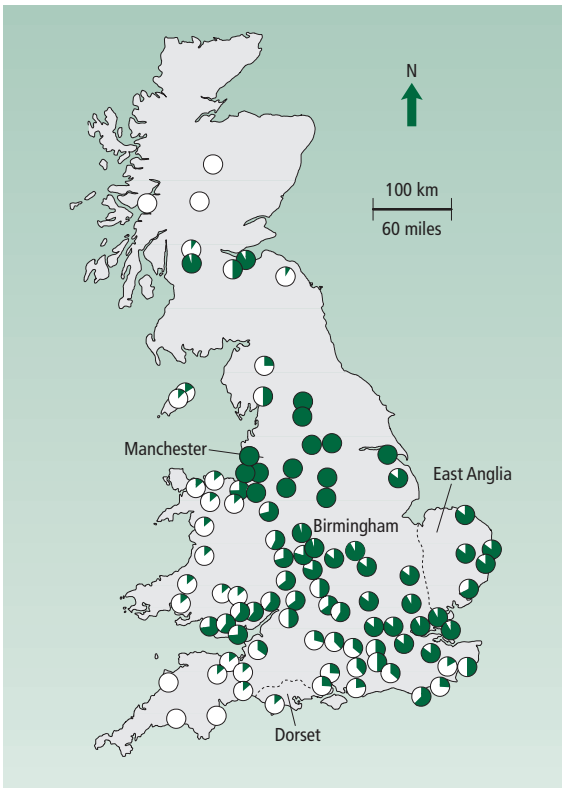

Figure 5.5
Frequency of melanic and light-colored forms of the peppered moth in different parts of Britain when the frequency of the melanic form was near its peak. The green part of each pie diagram is the frequency of the melanic form in that area. Melanic moths are generally higher in industrial areas, such as central England; but note the high proportion in East Anglia. Melanic frequencies have subsequently decreased (see Figure 5.6, for instance). Redrawn, by permission of the publisher, from Lees (1971).

The fitness estimates have been repeated many times

Cook (2000) reviewed about 30 experimental fitness estimates, done by several teams of biologists,[3] and they all gave similar results. The fitness estimates for the two forms of the peppered moth are about the most repeated result in evolutionary biology, and do not depend on the details of any particular experiment. The repeated results amount to an almost overwhelming case that the rise and fall of the melanic form of the peppered moth depended on air pollution. The evidence that air pollution exerted its effect via bird predation is also strong, if not overwhelming.

Other factors have been suggested

Evidence has also been put forward for other factors, in addition to bird predation. Migration is one extra factor. The geographic distribution of the two forms does not exactly fit Kettlewell's theory. The melanic form, for example, had a frequency of up to 80% in East Anglia, where pollution is low (Figure 5.5). And in some polluted areas, the dark form did not seem to have a high enough frequency. It never exceeded about 95% even though it was clearly better camouflaged and ought for that reason to have had a frequency of 100%. However, male moths can fly long distances to find females, and a male peppered moth mates on average 1.5 miles (2.5 km) away from where it is born.

[3] It has even been suggested that Kettlewell faked his results. The charge has only been supported by indirect evidence that is open to innocent interpretations. But however that may be, Kettlewell's explanation for evolution in the peppered moth — bird predation — does not depend on Kettlewell's own research. His results have been independently repeated.

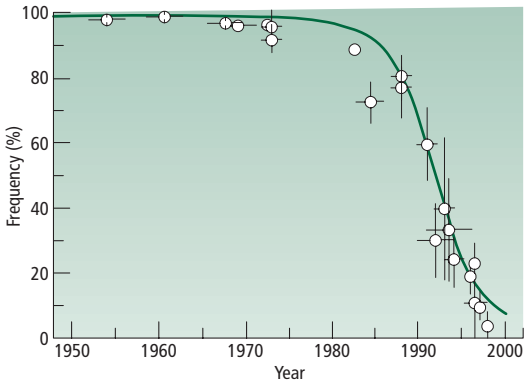

Figure 5.6
Decrease in frequency of the melanic form of the peppered moth in the region around Manchester. The decrease did not become really noticeable until about 1990. Redrawn, by permission of the publisher, from Cook *et al.* (1999).

Migration may explain why melanic moths are found in some unpolluted areas such as East Anglia and why light-colored moths persisted in polluted areas where they were less well camouflaged.

A second additional factor is that the two forms may differ in fitness independently of bird predation. Creed *et al.* (1980) collected all the measurements that had been made on survival to adulthood in the laboratory. They analyzed the results of 83 broods, containing 12,569 offspring; the original measurements had been made by many different geneticists in the previous 115 years. The viability of light-colored homozygotes, it turned out, was about 30% less on average than that of the melanic homozygote in the laboratory, where there is no bird predation — the reason is not known, but the fact alone implies there is some "inherent" advantage to the melanic genotype. The fitness advantage detected in the lab implies that melanic moths would replace light ones even without bird predation in polluted areas. In unpolluted areas, light-colored moths may remain only because birds eat more of the conspicuous melanic moths.

The melanic form may have an "inherent" advantage

Some biologists have suggested that three factors — bird predation, inherent advantage to melanic genotypes, and migration — are needed to explain peppered moth evolution. The importance of migration in addition to bird predation is generally accepted, but the inherent advantage to the melanic form is controversial. Since the measurements compiled in Creed *et al.* (1980) were made, the decrease in the melanic form's frequency has been more and more widely documented. The decrease did not happen around the formerly industrial Manchester region until the 1990s (Figure 5.6). The decrease makes sense if the advantage to the melanic form depends on air pollution, but not if it has an inherent advantage. Therefore, other biologists explain the observations in terms of bird predation (supplemented by migration) alone, and rule out the inherent advantage.

But the decrease in melanic frequency since the air became cleaner supports the classic explanation

In conclusion, the industrial melanism of the peppered moth is a classic example of natural selection. It can be used to illustrate the one-locus, two-allele model of selection. The model can be used to make a rough estimate of the difference in fitness between the two forms of moth using their frequencies at different times; the fitnesses can also be estimated from mark–recapture experiments. Good evidence exists that bird predation is at least partly the agent of selection, but some biologists suggest other factors are at work too.

5.8 Pesticide resistance in insects is an example of natural selection

Malaria is caused by a protozoan blood parasite (Section 5.12.2), and humans are infected with it by mosquitoes (family Culicidae — genera include *Aedes*, *Anopheles*, *Culex*). It can therefore be prevented by killing the local mosquito population, and health workers have recurrently responded to malarial outbreaks by spraying insecticides such as DDT in affected areas. DDT, sprayed on a normal insect, is a lethal nerve poison. When it is first sprayed on a local mosquito population, the population goes into abrupt decline. What happens then depends on whether DDT has been sprayed before.

Pests, such as mosquitoes, evolve resistance to pesticides, such as DDT

On its first use, DDT is effective for several years; in India, for example, it remained effective for 10–11 years after its first widespread use in the late 1940s. DDT, on a global scale, was one reason why the number of cases of malaria reduced to 75 million or so per year by the early 1960s. But by then, DDT-resistant mosquitoes had already begun to appear. DDT-resistant mosquitoes were first detected in India in 1959, and they have increased so rapidly that when a local spray program is begun now, most mosquitoes become resistant in a matter of months rather than years (Figure 5.7). The malarial statistics reveal the consequence. The global incidence of the disease almost exploded, up to somewhere between 300 and 500 million people at present. Malaria currently kills over 1 million people per year, mainly children aged 1–4 years. Pesticide resistance was not the only reason for the increase, but it was important.

The fitnesses can be estimated, . . .

DDT becomes ineffective so quickly now because DDT-resistant mosquitoes exist at a low frequency in the global mosquito population and, when a local population is sprayed, a strong force of selection in favor of the resistant mosquitoes is immediately created. It is only a matter of time before the resistant mosquitoes take over. A graph such as Figure 5.7 allows a rough estimate of the strength of selection. As for the peppered moth, we need to understand the genetics of the character, and to measure the genotype frequencies at two or more times. We can then use the formula for gene frequency change to estimate the fitness.

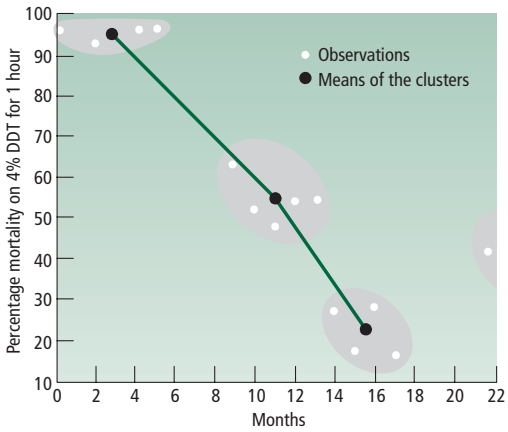

Figure 5.7
Increase in frequency of pesticide resistance in mosquitoes (*Anopheles culicifacies*) after spraying with DDT. A sample of mosquitoes was captured at each time indicated and the number that were killed by a standard dose of DDT (4% DDT for 1 hour) in the laboratory was measured. Redrawn, by permission of the publisher, from Curtis *et al.* (1978).

. . . given certain assumptions

We have to make a number of assumptions. One is that resistance is controlled by a single allele (we shall return to this below). Another concerns the degree of dominance: the allele conferring resistance might be dominant, recessive, or intermediate, relative to the natural susceptibility allele. The case of dominant resistance is easiest to understand. (If resistance is recessive we follow the same general method, but the exact result differs.) Let us call the resistance allele R and the susceptibility allele r. All the mosquitoes that die, in the mortality tests used in Figure 5.7, would then have been homozygous (rr) for susceptibility. Assuming (for simplicity rather than exact accuracy) Hardy–Weinberg ratios, we can estimate the frequency of the susceptibility gene as the square root of the proportion of mosquitoes that die in the tests. The selection coefficients are defined as follows, where fitness is measured as the chance of survival in the presence of DDT:

Genotype	RR	Rr	rr
Fitness	1	1	$1-s$

If we define p as the frequency of R and q as the frequency of r, equation 5.5 again gives the change in gene frequency: selection is working against a recessive gene. Figure 5.7 shows the decline in frequency of the susceptible mosquitoes, which are the recessive homozygotes. We therefore need a formula for the change in q in one generation (Δq), rather than Δp (as on p. 106). The decrease in q is the mirror image of the increase in p, and we just need to put a minus sign in front of equation 5.5:

$$\Delta q = \frac{-spq^2}{1 - sq^2} \tag{5.7}$$

The selection coefficient $s \approx 0.5$

The generation time is about 1 month. (The generations of mosquitoes overlap, rather than being discrete as the model assumes; but the exact procedure is similar in either case, and we can ignore the detailed correction for overlapping generations.) Table 5.7 shows how the genotype frequencies were read off Figure 5.7 in two stages, giving two estimates of fitness. Again, the formula for one generation has to be applied

Table 5.7
Estimated selection coefficients against DDT-susceptible *Anopheles culicifacies*, from Figure 5.7, where the relative fitness of the susceptible type is $(1-s)$. The estimate assumes the resistance allele is dominant. Simplified from Curtis *et al.* (1978).

Frequency of susceptible type			
Before	**After**	**Time (months)**	**Selection coefficient**
0.96	0.56	8.25	0.4
0.56	0.24	4.5	0.55

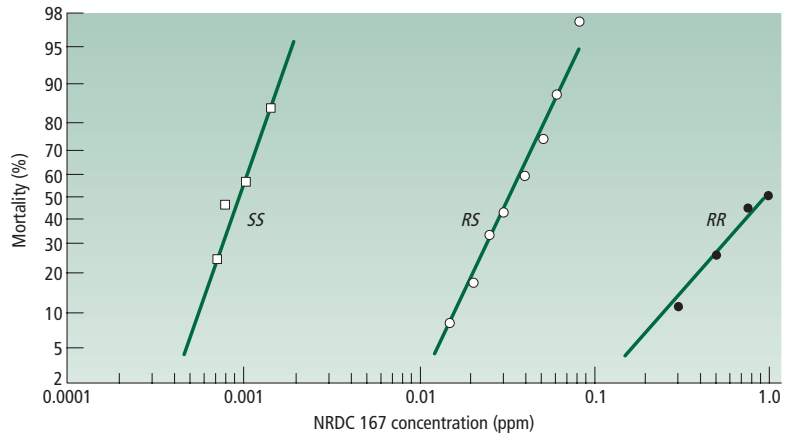

Figure 5.8
The mortality of mosquitos (*Culex quinquifasciatus*) of three genotypes at a locus when exposed to various concentrations of permethrin. The susceptible homozygote (*SS*) dies at lower concentrations of the poison than the resistant homozygote (*RR*). The heterozygote (*RS*) has intermediate resistance. Redrawn, by permission of the publisher, from Taylor (1986).

recurrently, for 8.25 and 4.5 generations in this case, to give an average fitness for the genotypes through the period. It appears that in Figure 5.7 the resistant mosquitoes had about twice the fitness of the susceptible ones — which is very strong selection.

The real genetics of resistance is known in some cases

The genetics of resistance in this case are not known, and the one-locus, two-allele model is an assumption only; but they are understood in some other cases. Resistance is often controlled by a single resistance allele. For example, Figure 5.8 shows that the resistance of the mosquito *Culex quinquifasciatus* to permethrin is due to a resistance (*R*) allele, which acts in a semidominant way, with heterozygotes intermediate between the two homozygotes. In houseflies, resistance to DDT is due to an allele called *kdr*. *kdr* flies are resistant because they have fewer binding sites for DDT on their neurons. In other cases, resistance may be due not to a new point mutation, but to gene amplification. *Culex pipiens*, for instance, in one experiment became resistant to an organophosphate insecticide called temephos because individuals arose with increased numbers of copies of a gene for an esterase enzyme that detoxified the poison. In the absence of temephos, the resistance disappeared, which suggests that the amplified genotype has to be maintained by selection. A number of mechanisms of resistance are known, and Table 5.8 summarizes the main ones that have been identified.

When an insect pest has become resistant to one insecticide, the authorities often respond by spraying it with another insecticide. The evolutionary pattern we have seen here then usually repeats itself, and on a shorter timescale. On Long Island, New York, for example, the Colorado potato beetle (*Leptinotarsa septemlineata*) was first attacked with DDT. It evolved resistance to it in 7 years. The beetles were then sprayed with azinphosmethyl, and evolved resistance in 5 years; next came carbofuran (2 years), pyrethroids (another 2 years), and finally pyrethroids with synergist (1 year). The decreasing time to evolve resistance is probably partly due to detoxification mechanisms that work against more than one pesticide. Pesticides cost money to develop, and the evolution of resistance reduces the economic lifetime of a pesticide. Box 5.2 looks at how the lifetime of a pesticide may be lengthened by slowing the evolution of resistance.

The theory has practical applications

Insecticide resistance matters not only in the prevention of disease, but also in farming. Insect pests at present destroy about 20% of world crop production, and it has been estimated that in the absence of pesticides as much as 50% would be lost. Insect pests

Table 5.8
The main mechanisms of resistance to insecticides. Reprinted, by permission of the publisher, from Taylor (1986).

Mechanism	Insecticides affected
Behavioral	
Increased sensitivity to insecticide	DDT
Avoid treated microhabitats	Many
Increased detoxification	
Dehydrochlorinase	DDT
Microsome oxidase	Carbamates
	Pyrethroids
	Phosphorothioates
Glutathione transferase	Organophosphates (*O*-dimethyl)
Hydrolases, esterases	Organophosphates
Descreased sensitivity of target site	
Acetylcholinesterase	Organophosphates
	Carbamates
Nerve sensitivity	DDT
	Pyrethroids
Cyclodiene-resistance genes	Cyclodienes (organochlorines)
Decreased cuticular penetration	Most insecticides

are a major economic and health problem. The evolution of resistance to pesticides causes misery to millions of people, whether through disease or reduced food supply. The fact that insects can rapidly evolve resistance is not the only problem with using pesticides against pests — the pesticides themselves (as is well known) can cause ecological side effects that range from the irritating to the dangerous. But however that may be, pesticides did not exist during the hundreds of millions of years that insects lived for before they were introduced in the 1940s, and the rapid evolution since then of resistance to pesticides provides a marvellously clear example of evolution by natural selection (Section 10.7.3, p. 276, extends the story, and Box 8.1, p. 213, looks at drug resistance in the malaria organism itself).

5.9 Fitnesses are important numbers in evolutionary theory and can be estimated by three main methods

Fitness can be measured . . .

The fitness of a genotype, in the theory and examples we have met, is its relative probability of survival from birth to adulthood. The fitness also determines the change in gene frequencies between generations. These two properties of fitness allow two methods of measuring it.

Box 5.2
Resistance Management

The evolution of resistance to each new pesticide, and antibiotic, is probably ultimately inevitable. However, we may be able to prolong the economically useful lives of these defensive chemicals by slowing down the evolution of resistance. The time it takes for resistance to evolve will be influenced by several factors. Two such factors can be seen in the simple models of selection we have been considering.

1. The degree of genetic dominance. The frequency of an advantageous dominant gene increases much more rapidly by natural selection than does the frequency of an advantageous recessive gene. An advantageous gene, such as one producing resistance to a pesticide, will initially be present only in one copy, in a heterozygote. If the gene is recessive, it is not expressed in that heterozygote. Natural selection cannot "see" the gene until it is found in a homozygote. If the gene is dominant, it is immediately expressed and natural selection immediately favors it. A recessive resistance gene will increases in frequency much more slowly than a dominant resistance gene.
2. The relative fitness of the resistant and non-resistant genotypes. A genotype with a large fitness advantage will increase in frequency more rapidly than one with a low fitness advantage. For instance, in Table 5.4 we can see that a genotype with a 1% advantage takes five times as long to reach a frequency of 80% than does a genotype with a 5% advantage.

Thus the evolution of resistance could be slowed down if we could make the resistance gene more recessive (or less dominant), and if we could reduce its fitness advantage relative to the non-resistant types.

One way to make the resistance gene recessive might be to apply the pesticide in large doses. The resistance gene may code for a protein that somehow neutralizes the pesticide. If there are small quantities of pesticide, a single copy of the resistance gene (in a heterozygote) may produce enough of the protein to cope with the pesticide. The gene is then effectively dominant, because it produces resistance in heterozygotes. The gene will spread fast. But if large amounts of the pesticide are used, the single gene may be overwhelmed. Two resistance genes (in a homozygote) may be needed to cope. The large amount of pesticide makes the resistance gene effectively recessive.

The relative fitnesses of the resistant and non-resistant genotypes may be influenced by the way the pesticide is applied in space. If pesticides are applied in some places but not others, the non-resistant genotypes will have a selective advantage in the localities where there is no pesticide. The average fitness of the resistant genotype will then not be so high, relative to the non-resistant genotype, as it would be if the pesticide were applied indiscriminately in the whole region.

Rausher (2001) refers to the combination of these two policies as the "high dose/refuge strategy." However, the strategy requires certain conditions to succeed in slowing the evolution of resistance, even in theory, and very little practical work has been done to test it. Currently, it is a research problem for the future. However, the idea does illustrate how the evolutionary models of this chapter can have practical applications. The economic value of these models could even turn out to be huge.

Further reading: Rausher (2001).

. . . by relative survival within a generation . . .

The first method is to measure the relative survival of the genotypes within a generation. Kettlewell's mark–recapture experiment with the peppered moth is an example. If we assume that the relative rate of recapture of the genotypes is equal to their relative chance of survival from egg to adulthood, we have an estimate of fitness. The assumption may be invalid. The genotypes may, for instance, differ in their chances of survival at some stage of life other than the time of the mark–recapture experiment. If the survival of adult moths is measured by mark–recapture, any differences among genotypes in survival at the egg and caterpillar stages will not be detected. Also, the genotypes may differ in fertility: fitnesses estimated by differences in survival are only accurate if all the

genotypes have the same fertility. These assumptions can all be tested by further work. For instance, survival can be measured at the other life stages too, and fertility can also be assessed. In a few cases, lifetime fitnesses have been measured comprehensively, by tracing survival and reproduction from birth to death.

. . . or rate of gene frequency change between generations . . .

The second method is to measure changes in gene frequencies between generations. We then substitute the measurements into the formula that expresses fitness in terms of gene frequencies in successive generations (equation 5.6). Both methods have been used in many cases; the main problems are the obvious difficulties of accurately measuring survival and gene frequencies, respectively. Apart from them, in the examples we considered there were also difficulties in understanding the genetics of the characters: we need to know which phenotypes correspond to which genotypes in order to estimate genotype fitnesses.

. . . or other methods

We shall meet a third method of estimating fitness below, in the case of sickle cell anemia (see Table 5.9, p. 126). It uses deviations from the Hardy–Weinberg ratios. It can be used only when the gene frequencies in the population are constant between the stages of birth and adulthood, but the genotypes have different survival. It therefore cannot be used in the examples of directional selection against a disadvantageous gene that we have been concerned with so far, because in them the gene frequency in the population changes between birth and adult stages.

We have discussed the inference of fitness in detail because the fitnesses of different genotypes are among the most important variables — perhaps the most important variables — in the theory of evolution. They determine, to a large extent, which genotypes we can expect to see in the world today. The examples we have looked at, however, illustrate that fitnesses are not easy to measure. We require long time series and large sample sizes, and even then the estimates may be subject to "other things being equal" assumptions. Therefore, despite their importance, they have been measured in only a small number of the systems that biologists are interested in. (That does not mean that the absolute number of such studies is small. A review of research on natural selection in the wild by Endler in 1986 contains a table (24 pages long) listing all the work he had located. Fitnesses have only been measured in a minority — an unknown minority — of those 24 pages' worth of studies of natural selection, but the number could still be non-trivial.) Many unsolved controversies in evolutionary biology implicitly concern values of fitnesses, but in systems in which it has not been possible to measure fitnesses directly with sufficient accuracy or in a sufficiently large number of cases. The controversy about the causes of molecular evolution in Chapter 7 is an example. When we come to discuss controversies of this sort it is worth bearing in mind what would have to be done to solve them by direct measurements of fitness.

5.10 Natural selection operating on a favored allele at a single locus is not meant to be a general model of evolution

Evolutionary change in which natural selection favors a rare mutation at a single locus, and carries it up to fixation, is one of the simplest forms of evolution. Sometimes

Other factors will be at work in real examples

evolution may happen that way. But things can be more complicated in nature. We have considered selection in terms of different chances of survival from birth to adulthood; but selection can also take place by differences in fertility, if individuals of different genotypes — after they have survived to adulthood — produce different numbers of offspring. The model had random mating among the genotypes: but mating may be non-random. Moreover, the fitness of a genotype may vary in time and space, and depend on what genotypes are present at other loci (a subject we shall deal with in Chapter 8). Much of evolutionary change probably consists of adjustments in the frequencies of alleles at polymorphic loci, as fitnesses fluctuate through time, rather than the fixation of new favorable mutations.

These complexities in the real world are important, but they do not invalidate — or trivialize — the one-locus model. For the model is intended as a model. It should be used as an aid to understanding, not as a general theory of nature. In science, it is a good strategy to build up an understanding of nature's complexities by considering simple cases first and then building on them to understand the complex whole. Simple ideas rarely provide accurate, general theories; but they often provide powerful paradigms. The one-locus model is concrete and easy to understand and it is a good starting point for the science of population genetics. Indeed, population geneticists have constructed models of all the complications listed in the previous paragraph, and those models are all developments within the general method we have been studying.

5.11 A recurrent disadvantageous mutation will evolve to a calculable equilibrial frequency

A disadvantageous mutation may arise recurrently

The model of selection at one locus revealed how a favorable mutation will spread through a population. But what about unfavorable mutations? Natural selection will act to eliminate any allele that decreases the fitness of its bearers, and the allele's frequency will decrease at a rate specified by the equations of Section 5.6; but what about a recurrent disadvantageous mutation that keeps arising at a certain rate? Selection can never finally eliminate the gene, because it will keep on reappearing by mutation. In this case, we can work out the equilibrial frequency of the mutation: the equilibrium is between the mutant gene's creation, by recurrent mutation, and its elimination by natural selection.

To be specific, we can consider a single locus, at which there is initially one allele, a. The gene has a tendency to mutate to a dominant allele, A. We must specify the mutation rate and the selection coefficient (fitness) of the genotypes: define m as the mutation rate from a to A per generation. We will ignore back mutation (though actually this assumption does not matter). The frequency of a is q, and of A is p. Finally, we define the fitnesses as follows:

Genotype	aa	Aa	AA
Fitness	1	$1-s$	$1-s$

Evolution in this case will proceed to an equilibrial frequency of the gene A (we can write the stable equilibrium frequency as p^*). If the frequency of A is higher than the

equilibrium, natural selection removes more A genes than mutation creates and the frequency decreases; vice versa if the frequency is lower than the equilibrium. At the equilibrium, the rate of loss of A genes by selection equals their rate of gain by mutation.

We construct a model of the gene frequency of a recurrent disadvantageous mutation

We can use that statement to calculate the equilibrial gene frequency $p^\star$. What is the rate per generation of creation of A genes by mutation? Each new A gene originates by mutation from an a gene and the chance that any one a gene mutates to an A gene is the mutation rate m. A proportion $(1 - p)$ of the genes in the population are a genes. Therefore:

Total rate of creation of A genes by mutation $= m(1-p)$

And what is the rate at which A genes are eliminated? Each A gene has a $(1-s)$ chance of surviving, or an s chance of dying. A proportion p of the genes in the population are A. Therefore:

Total rate of loss of A genes by selection $= ps$

At the equilibrium gene frequency ($p^\star$):

Rate of gain of A gene = rate of loss of A gene

$$m(1-p^\star) = p^\star s \tag{5.8}$$

Which can be multiplied out:

$$m - mp^\star = p^\star s$$
$$p^\star = m/(s+m)$$

Of the two terms in the denominator, the mutation rate (maybe 10^{-6}, Section 2.6, p. 32) will usually be much less than the selection coefficient (perhaps 10^{-1} or 10^{-2}). With these values $s + m \approx s$ and the expression is therefore usually given in the approximate form:

$$p^\star = m/s \tag{5.9}$$

The disadvantageous mutation has a low equilibrium frequency . . .

The simple result is that the equilibrium gene frequency of the mutation is equal to the ratio of its mutation rate to its selective disadvantage. The result is intuitive: the equilibrium is the balance between the rates of creation and elimination of the gene. To obtain the result, we used an argument about an equilibrium. We noticed that at the equilibrium the rate of loss of the gene equals the rate of gain and used that to work out the exact result. This is a powerful method for deriving equilibria, and we shall use an analogous argument in the next section.

The expression $p = m/s$ allows a rough estimate of the mutation rate of a harmful mutation just from a measurement of the mutant gene's frequency. If the mutation is rare, it will be present mainly in heterozygotes, which at birth will have frequency $2pq$. If p is small, $q \approx 1$ and $2pq \approx 2p$. N is defined as the frequency of mutant bearers, which equals the frequency of heterozygotes: i.e., $N = 2p$. As $p = m/s$, $m = sp$; if we substitute

. . . which can sometimes be used to estimate the mutation rate

$p = N/2$, $m = sN/2$. If the mutation is highly deleterious, $s \approx 1$ and $m = N/2$. The mutation rate can be estimated as half the birth rate of the mutant type. The estimate is clearly approximate, because it relies on a number of assumptions. In addition to the assumptions of high s and low p, mating is supposed to be random. We usually have no means of checking whether it is.

Chondrosdystrophic dwarfism is a dominant deleterious mutation in humans. In one study, 10 births out of 94,075 had the gene, a frequency of 10.6×10^{-5}. The estimate of the mutation rate by the above method is then $m = 5.3 \times 10^{-5}$. However, it is possible to estimate the selection coefficient, enabling a more accurate estimate of the mutation rate. In another study, 108 chondrodystrophic dwarves produced 27 children; their 457 normal siblings produced 582 children. The relative fitness of the dwarves was (27/108)/(582/457) = 0.196; the selection coefficient s = 0.804. Instead of assuming $s = 1$, we can use $s = 0.804$. Then the mutation rate is $sN/2 = 4.3 \times 10^{-5}$, a rather lower figure because with lower selection the same gene frequency can be maintained by a lower mutation rate.

For many genes, we do not know the dominance relations of the alleles at the locus. A similar calculation can be done for a recessive gene, but the formula is different, and it differs again if the mutation has intermediate dominance. We can only estimate the mutation rate from $p = m/s$ if we know the mutation is dominant. The method is therefore unreliable unless its assumptions have been independently verified. However, the general idea of this section — that a balance between selection and mutation can exist and explain genetic variation — will be used in later chapters.

5.12 Heterozygous advantage

5.12.1 *Selection can maintain a polymorphism when the heterozygote is fitter than either homozygote*

In some cases, heterozygotes have higher fitness than homozygotes

We come now to an influential theory. We are going to consider the case in which the heterozygote is fitter than both homozygotes. The fitnesses can be written:

Genotype	*AA*	*Aa*	*aa*
Fitness	$1 - s$	1	$1 - t$

t, like s, is a selection coefficient and has a value between 0 and 1. What happens here? There are three possible equilibria, but two of them are trivial. $p = 1$ and $p = 0$ are stable equilibria, but only because there is no mutation in the model. The third equilibrium is the interesting one; it has both genes present, and we can calculate the equilibrial gene frequencies by a similar argument to the one outlined in the previous section. The condition in which a population contains more than one gene is called *polymorphism*.

A genes and *a* genes are both removed by selection. The *A* genes are removed because they appear in the inferior *AA* homozygotes and the *a* genes because they appear in *aa* homozygotes. At the equilibrium, both genes must have the same chance of being removed by selection. If an *A* gene has a higher chance of being removed than an *a* gene,

the frequency of *a* is increasing, and vice versa. Only when the chance is the same for both will the gene frequencies be stable.

We construct a model of gene frequencies with heterozygous advantage

What is the chance that an *A* gene will be carried by an individual who will die without reproducing? An *A* gene is either (with chance q) in a heterozygote and survives or (with chance p) in an *AA* homozygote and has a chance s of dying. Its total chance of dying is therefore ps. An *a* gene similarly is either (with chance p) in a heterozygote and survives or (with chance q) in an *aa* homozygote and has chance t of dying: its chance of death is qt. At the equilibrium,

Chance of death of an *A* gene = chance of death of an *a* gene

$$p^*s = q^*t \tag{5.10}$$

Substitute $p^*s = (1 - p^*)t$

and rearrange $$p^* = t/(s + t) \tag{5.11}$$

Similarly if we substitute $q = (1 - p)$, $q^* = s/(s + t)$. Now we have derived the equilibrial gene frequencies when both homozygotes have lower fitness than the heterozygote. The equilibrium has all three genotypes present, even though the homozygotes are inferior and are selected against. They continue to exist because it is impossible to eliminate them. Matings among heterozygotes generate homozygotes. The exact gene frequency at equilibrium depends on the relative selection against the two homozygotes. If, for instance, *AA* and *aa* have equal fitness, then $s = t$ and $p = 1/2$ at equilibrium. If *AA* is relatively more unfit than *aa* then $s > t$ and $p < 1/2$; there are fewer of the more strongly selected against genotypes.

When heterozygotes are fitter than the homozygotes, therefore, natural selection will maintain a polymorphism. The result was first proved by Fisher in 1922 and independently by Haldane. We shall come later to consider in more detail why genetic variability exists in natural populations, and *heterozygous advantage* will be one of several controversial explanations to be tested.

5.12.2 *Sickle cell anemia is a polymorphism with heterozygous advantage*

Sickle cell anemia illustrates the theory

Sickle cell anemia is the classic example of a polymorphism maintained by heterozygous advantage. It is a nearly lethal condition in humans, responsible for about 100,000 deaths a year. It is caused by a genetic variant of α-hemoglobin. If we symbolize the normal hemoglobin allele by *A* and the sickle cell hemoglobin by *S*, then people who suffer from sickle cell anemia are *SS*. Hemoglobin *S* causes the red blood cells to become curved and distorted (sickle shaped); they can then block capillaries and cause severe anemia if the blocked capillary is in the brain. About 80% of *SS* individuals die before reproducing. With such apparently strong selection against hemoglobin *S* it was a puzzle why it persisted at quite high frequencies (10% or even more) in some human populations.

If we compare a map of the incidence of malaria with a map of the gene frequency (Figure 5.9), we see that they are strikingly similar. Perhaps hemoglobin *S* provides

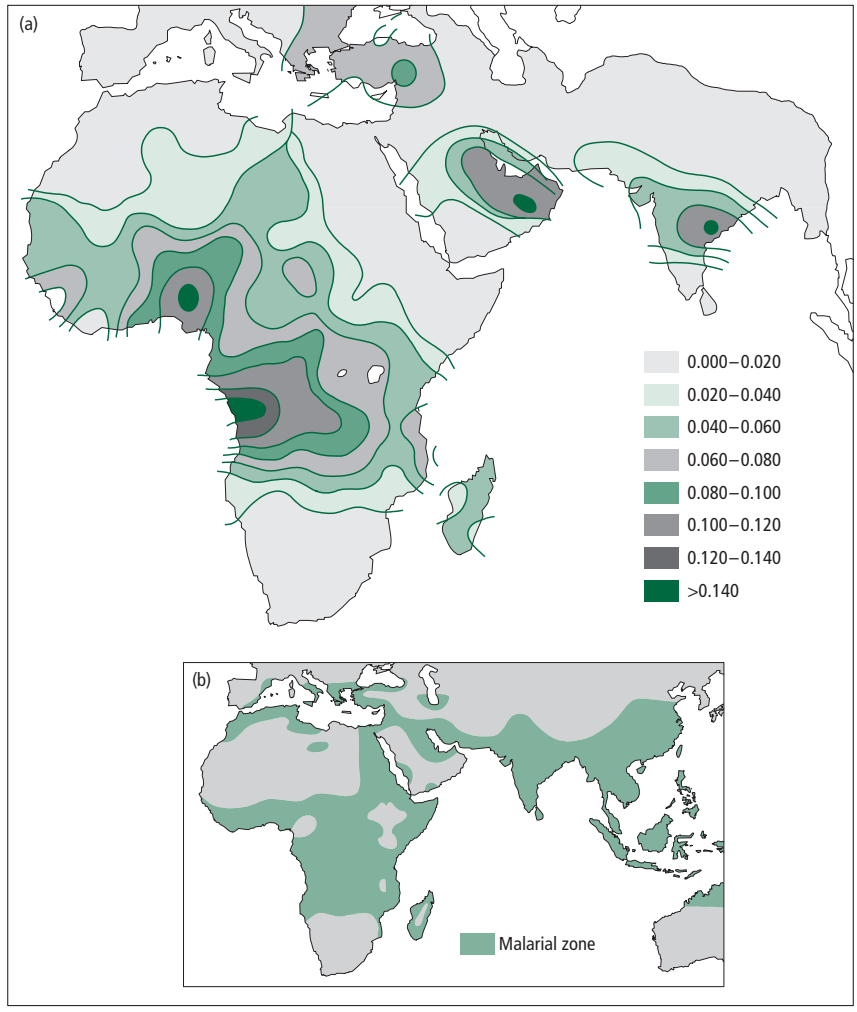

Figure 5.9
The global incidence of malaria coincides with that of the sickle cell form of hemoglobin. (a) A map of the frequency of the *S* allele of hemoglobin. (b) A map of malarial incidence. Redrawn, by permission of the publisher, from Bodmer & Cavalli-Sforza (1976).

Sickle cell hemoglobin confers resistance to malaria

some advantage in malarial zones. Allison (1954) showed that, although *SS* is almost lethal, the heterozygote *AS* is more resistant to malaria than the homozygote *AA*. (Allison's was the first demonstration of natural selection at work in a human population.) The full reason was discovered later — *AS* red blood cells do not normally sickle, but they do if the oxygen concentration falls. When the malarial parasite *Plasmodium falciparum* enters a red blood cell it destroys (probably eats) the hemoglobin, which causes the oxygen concentration in the cell to go down. The cell sickles and is destroyed, along with the parasite. The human survives because most of the red blood cells are uninfected and carry oxygen normally. Therefore, where the malarial parasite is common, *AS* humans survive better than *AA*, who suffer from malaria.

Once the heterozygote had been shown physiologically to be at an advantage, the adult genotype frequencies can be used to estimate the relative fitnesses of the three genotypes. The fitnesses are:

Table 5.9
Estimates of selection coefficients for sickle cell anemia, using genotype frequencies in adults. The sickle cell hemoglobin allele is *S*, and the normal hemoglobin (which actually consists of more than one allele) is *A*. The genotype frequencies are for the Yorubas of Ibadan, Nigeria. One small detail is not explained in the text. The observed : expected ratio for the heterozygote may not be equal to 1. Here it turned out to be 1.12. All the observed : expected ratios are therefore divided by 1.12 to make them fit the standard fitness regime for heterozygote advantage. From Bodmer & Cavalli-Sforza (1976).

Genotype	Observed adult frequency (*O*)	Expected Hardy–Weinberg frequency (*E*)	Ratio *O* : *E*	Fitness
SS	29	187.4	0.155	$0.155/1.12 = 0.14 = 1 - t$
SA	2,993	2,672.4	1.12	$1.12/1.12 = 1.00$
AA	9,365	9,527.2	0.983	$0.983/1.12 = 0.88 = 1 - s$
Total	12,387	12,387		

Calculation of expected frequencies: gene frequency of *S* = frequency of *SS* + $^1/_2$ (frequency of *SA*) = (29 + 2,993/2)/12,387 = 0.123. Therefore the frequency of *A* allele = 1 − 0.123 = 0.877. From the Hardy–Weinberg theorem, the expected genotype frequencies are $(0.123)^2 \times 12{,}387$, $2(0.877)(0.123) \times 12{,}387$, and $(0.877)^2 \times 12{,}387$, for *AA*, *AS*, and *SS*, respectively.

Genotype	*AA*	*AS*	*SS*
Fitness	$1 - s$	1	$1 - t$

If the frequency of gene $A = p$ and of gene $S = q$, then the relative genotype frequencies among adults will be $p^2(1 - s) : 2pq : q^2(1 - t)$. If there were no selection ($s = t = 0$), the three genotypes would have Hardy–Weinberg frequencies of $p^2 : 2pq : q^2$.

We deduce selection coefficients of 0.12 and 0.86

Selection causes deviations from the Hardy–Weinberg frequencies. Take the genotype *AA* as an example. The ratio of the observed frequency in adults to that predicted from the Hardy–Weinberg ratio will be $(1 - s)/1$. The frequency expected from the Hardy–Weinberg principle is found by the usual method: the expected frequency is p^2, where p is the observed proportion of *AA* plus half the observed proportion of *AS*. Table 5.9 illustrates the method for a Nigerian population, where $s = 0.12$ $(1 - s = 0.88)$ and $t = 0.86$ $(1 - t = 0.14)$.

The method is only valid if the deviation from Hardy–Weinberg proportions is caused by heterozygous advantage and the genotypes differ only in their chance of survival (not their fertility). If heterozygotes are found to be in excess frequency in a natural population, it may indeed be because the heterozygote has a higher fitness. However, it could also be for other reasons. Disassortative mating, for instance, can produce the same result (in this case, disassortative mating would mean that *aa* individuals preferentially mate with *AA* individuals). But for sickle cell anemia, the physiological observations showed that the heterozygote is fitter and the procedure is well justified. Indeed, in this case, although it has not been checked whether mating is

random, the near lethality of *SS* means that disassortative mating will be unimportant; however, the assumption that the genotypes have equal fertility may well be false.

5.13 The fitness of a genotype may depend on its frequency

The next interesting complication is to consider selection when the fitness of a genotype depends on its frequency. In the models we have considered so far, the fitness of a genotype (1, 1 − *s*, or whatever) was constant, regardless of whether the genotype was rare or common. Now we consider the possibility that the fitness of a genotype goes up or down as the genotype frequency increases in the population (Figure 5.10). *Frequency-dependent selection* means that natural selection is acting and the fitnesses of the genotypes vary with the frequency of the genotypes. The two main kinds are *negative frequency dependence*, in which the fitness of a genotype goes down as its frequency goes up, and *positive frequency dependence*, in which the fitness of a genotype goes up as its frequency goes up.

In host–parasite relations, the fitness of a genotype may depend on frequency

Negative frequency dependence can arise in host–parasite interactions. For instance, two genotypes of a host may differ in their ability to keep out two genotypes of a parasite. This kind of set-up is like a lock and key. It is as if the two host genotypes are like two different locks, and the two parasite genotypes are like two different keys. One of the parasite keys fits one of the host locks and the other parasite key fits the other host lock. Then, if one of the host genotypes is in high frequency, natural selection will favor the parasite genotype that can penetrate that common kind of host. The result is that a high frequency automatically brings a disadvantage to a host genotype, because it creates an advantage for the kind of parasite than can exploit it. As the frequency of a host genotypes increases, its fitness soon decreases.

Snails and their parasite provide an example

Lively & Dybdahl (2000) recently described an example where the host is a snail, *Potamopyrgus antipodarum*, which (as its name hints at) lives in New Zealand, in freshwater habitats. The snail suffers from various parasites, of which a trematode called *Microphallus* is the most important (it is a parasitic castrator). The authors distinguished several strains (or clones) of the snail host and measured the frequency of each clone. They then measured, in an experiment, the ability of *Microphallus* to infect each clone. Figure 5.11 shows the infection rates achieved by parasites collected from two lakes, when experimentally exposed to snails taken from one of the two lakes. The local parasites infected the common clones better than the rare clones. It was the high

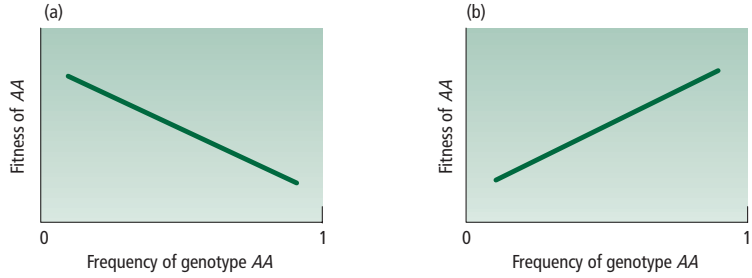

Figure 5.10
Frequency-dependent selection. (a) Negative frequency-dependent fitness means that the fitness of a genotype decreases as the frequency of the genotype increases. (b) Positive frequency-dependent fitness means that the fitness of a genotype goes up as its frequency increases. In general, frequency dependence refers to any case in which the graph is anything other than flat. A flat line, with fitness constant for all genotype frequencies, means that selection is not frequency dependent.

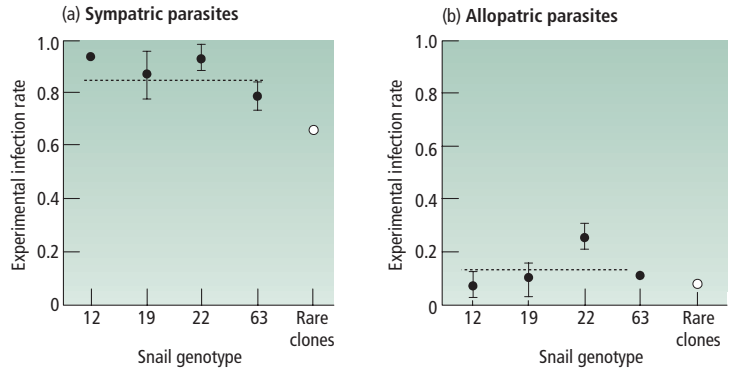

Figure 5.11
Parasites penetrate host genotypes more efficiently when they are locally abundant. Parasites from two lakes (Poerua and Ianthe) were experimentally put with snails of several genetic types (clones) from Lake Poerua. The four clones called 12, 19, 22, and 63 were common in the lake; several other clones were rare and they are all lumped together in the figure. The infection rates achieved by parasites taken from the two lakes were measured for each clone. (a) Infection rates achieved by parasites from Lake Poerua (sympatric parasites). (b) Infection rates achieved by parasites from Lake Ianthe (allopatric parasites). Note the higher infection rates achieved by the parasites on their local snails: the points are higher in (a) than in (b). But mainly note that the Poerua parasites in (a) infected the common snail clones more effectively than the rare clones; whereas the Ianthe parasites in (b) are no more effective with the common than the rare clones. From Lively & Dybdahl (2000). © 2000 Macmillan Magazines Ltd.

frequency of a clone that made it vulnerable to parasites. A clone that was common in one lake but rare in another was vulnerable to parasitism where it was common but not where it was rare.

Parasite–host relations are one important source of negative frequency-dependent selection (we return to this in Section 12.2.3, p. 323). Another important source is *multiple niche polymorphism*, a topic first discussed by Levene (1953). Suppose that a species contains several genotypes, and each genotype is adapted to a different set of environmental conditions. Genotypes *AA* and *Aa* might be adapted to the shade, and *aa* to sunny places (shady, and sunny, places then correspond to two "niches"). Then when the *A* gene is rare, *AA* and *Aa* experience less competition in their preferred areas, because there are fewer of them. As the frequency of *A* goes up, the shady areas become more crowded, competition increases, and fitness will tend to go down.

Frequency dependence is often generated by biological interactions. Competition and parasite–host relations are both biological interactions, and can generate negative frequency dependence. We shall meet some other examples, such as sex ratios (Section 12.5, p. 337) later in the book. Negative frequency-dependent fitnesses are important because they can produce stable polymorphisms within a species. As the frequency of each genotype goes up, its fitness goes down. Natural selection favors a gene when it is rare, but works against it when it is common. The result is that genotypes equilibrate at some intermediate frequency.

Frequency dependence can also arise in other circumstances

Positive frequency-dependent selection does not produce stable polymorphisms. Indeed it actively eliminates polymorphism, producing a genetically uniform population. For example, some species of insects have "warning coloration." They are brightly colored, and poisonous to eat. The bright coloration may reduce the chance of predation. When a bird eats the warningly colored insect, the bird is made sick and will remember not to eat an insect that looks like that again. However, the bird's lesson is not advantageous for the insect that made the bird sick; that insect is probably killed. When warningly colored insects are rare in a population mainly consisting of dull and cryptic individuals, the warningly colored genotypes are likely to have a low fitness. Few other insects exist to "educate" the local birds. This can create a problem in the evolution of warning coloration, because rare new mutants maybe selected against. The problem is not the point here, however. We are only considering it as an example of positive frequency dependence. The fitness of warningly colored genotypes will be

higher at high frequencies, where the local birds are well educated about the dangers of eating the warningly colored forms.

The purpose of Sections 5.11–5.13 has been to illustrate the different mechanisms by which natural selection can maintain polymorphism. In Chapter 6 we look at another mechanism that can maintain polymorphism — genetic drift. Then, in Chapter 7, we tackle the question of how important the mechanisms are in nature.

5.14 Subdivided populations require special population genetic principles

5.14.1 *A subdivided set of populations have a higher proportion of homozygotes than an equivalent fused population: this is the Wahlund effect*

Populations may be subdivided

So far we have considered population genetics within a single, uniform population. In practice, a species may consist of a number of separate populations, each more or less isolated from the others. The members of a species might, for example, inhabit a number of islands, with each island population being separated by the sea from the others. Individuals might migrate between islands from time to time, but each island population would evolve to some extent independently. A species with a number of more or less independent subpopulations is said to have *population subdivision*.

Let us see first what effect population subdivision has on the Hardy–Weinberg principle. Consider a simple case in which there are two populations (we can call them population 1 and population 2), and we concentrate on one genetic locus with two alleles, *A* and *a*. Suppose allele *A* has frequency 0.3 in population 1 and 0.7 in population 2. If the genotypes have Hardy–Weinberg ratios they will have the frequencies, and average frequencies, in the two populations shown in Table 5.10. The average genotype frequencies are 0.29 for *AA*, 0.42 for *Aa*, and 0.29 for *aa*. Now suppose that the two

Table 5.10
The frequency of genotypes *AA*, *Aa*, and *aa* in two populations when *A* has frequency 0.3 in population 1 and 0.7 in population 2. The average genotypes are calculated assuming the two populations are of equal size.

	Genotype			
	AA	***Aa***	***aa***	
Frequency	$(0.3)^2 = 0.09$ $(0.7)^2 = 0.49$	$2(0.3)(0.7) = 0.42$ $2(0.7)(0.3) = 0.42$	$(0.7)^2 = 0.49$ $(0.3)^2 = 0.09$	population 1 population 2
Average	0.58/2 = 0.29	0.84/2 = 0.42	0.58/2 = 0.29	

populations are fused together. The gene frequencies of *A* and *a* in the combined population are $(0.3 + 0.7)/2 = 0.5$, and the Hardy–Weinberg genotype frequencies are:

Genotype	*AA*	*Aa*	*aa*
Frequency	0.25	0.5	0.25

The Wahlund effect concerns the frequency of homozygotes in subdivided populations

In the large, fused population there are fewer homozygotes than in the average for the set of subdivided populations. This is a general, and mathematically automatic, result. The increased frequency of homozygotes in subdivided populations is called the *Wahlund effect*.

The Wahlund effect has a number of important consequences. One is that we have to know about the structure of a population when applying the Hardy–Weinberg principle to it. Suppose, for example, we had not known that populations 1 and 2 were independent. We might have sampled from both, pooled the samples indiscriminately, and then measured the genotype frequencies. We should find the frequency distribution for the average of the two populations (0.29, 0.42, 0.29); but the gene frequency would apparently be 0.5. There would seem to be more homozygotes than expected from the Hardy–Weinberg principle. We might suspect that selection, or some other factor, was favoring homozygotes. In fact both subpopulations are in perfectly good Hardy–Weinberg equilibrium and the deviation is due to the unwitting pooling of the separate populations. We need to look out for population subdivision when interpreting deviations from Hardy–Weinberg ratios.

Second, when a number of previously subdivided populations merge together, the frequency of homozygotes will decrease. In humans, this can lead to a decrease in the incidence of rare recessive genetic diseases when a previously isolated population comes into contact with a larger population. The recessive disease is only expressed in the homozygous condition, and when the two populations start to interbreed, the frequency of those homozygotes goes down.

5.14.2 *Migration acts to unify gene frequencies between populations*

When an individual migrates from one population to another, it carries genes that are representative of its own ancestral population into the recipient population. If it successfully establishes itself and breeds it will transmit those genes between the populations. The transfer of genes is called *gene flow*. If the two populations originally had different gene frequencies and if selection is not operating, migration (or, to be exact, gene flow) alone will rapidly cause the gene frequencies of the different populations to converge. We can see how rapidly in a simple model.

The spatial movement of genes is called gene flow

Consider again the case of two populations and one locus with two alleles (*A* and *a*). Suppose this time that one of the populations is much larger than the other, say population 2 is much larger than population 1 (2 might be a continent and 1 a small island off it); then practically all the migration is from population 2 to population 1. The frequency of allele *a* in population 1 in generation *t* is written $q_{1(t)}$; we can suppose that the frequency of *a* in the large population 2 is not changing between generations and

We construct a model of gene frequencies with migration

write it as q_m. (We are interested in the effect of migration on the gene frequency in population 1 and can ignore all other effects, such as selection.) Now, if we pick on any one allele in population 1 in generation $(t+1)$, it will either be descended from a native of the population or from an immigrant. Define m as the chance that it is a migrant gene. (Earlier in the chapter, m was used for the mutation rate: now it is the *migration* rate.) If our gene is not a migrant (chance $(1-m)$) it will be an a gene with chance $q_{1(t)}$, whereas if it is a migrant (chance m) it will be an a gene with chance q_m. The total frequency of a in population 1 in generation $(t+1)$ is:

$$q_{1(t+1)} = (1-m)q_{1(t)} + mq_m \tag{5.12}$$

This can be rearranged to show the effect of t generations of migration on the gene frequency in population 1. If $q_{1(0)}$ is the frequency in the 0th generation, the frequency in generation t will be:

$$q_{1(t)} = q_m + (q_{1(0)} - q_m)(1-m)^t \tag{5.13}$$

(From $t = 1$ it is easy to confirm that this is indeed a rearrangement of the previous equation.) The equation says that the difference between the gene frequency in population 1 and population 2 decreases by a factor $(1-m)$ per generation. At equilibrium, $q_1 = q_m$ and the small population will have the same gene frequency as the large population (Figure 5.12). In Figure 5.12, the gene frequencies converge in about 30 generations with a migration rate of 10%. Similar arguments apply if, instead of there being one source and one recipient population, the source is a set of many subpopulations, and p_m is their average gene frequency, or if there are two populations both sending migrants to, and receiving them from, another.

Gene flow binds biological species together

Migration will generally unify gene frequencies among populations rapidly in evolutionary time. In the absence of selection, migration is a strong force for equalizing the gene frequencies of populations within a species. Provided that the migration rate is greater than 0, gene frequencies will eventually equalize. Even if only one successful migrant moves into a population per generation, gene flow inevitably draws that population's gene frequency to the species' average. Gene flow acts, in a sense, to bind the species together.

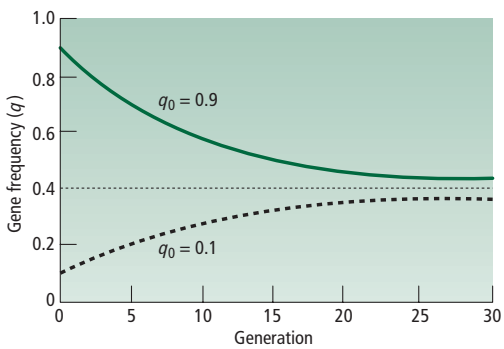

Figure 5.12
Migration causes the rapid convergence of gene frequencies in the populations exchanging migrants. Here a source population with gene frequency $q_m = 0.4$ sends migrants to two subpopulations, with initial gene frequencies of 0.9 and 0.1. They converge, with $m = 0.1$, onto the source population's gene frequency in about 30 generations.

5.14.3 *Convergence of gene frequencies by gene flow is illustrated by the human population of the USA*

The American population illustrates the model . . .

The MN blood group is controlled by one locus with two alleles (Section 5.4). Frequencies of the *M* and *N* alleles have been measured, for example in European and African Americans in Claxton, Georgia, and among West Africans (whom we can assume to be representative of the ancestral gene frequency of the African American population of Claxton). The *M* allele frequency is 0.474 in West Africans, 0.484 in African Americans in Claxton, and 0.507 in the European Americans of Claxton. (The frequency of the *N* allele is equal to 1 minus the frequency of the *M* allele.) The gene frequency among African Americans is intermediate between the frequencies for European Americans and for the West African sample. Individuals of mixed parentage are usually categorized as African American and, if we ignore the possibility of selection favoring the *M* allele in the USA, we can treat the change in gene frequency in the African American population as due to "migration" of genes from the European American population. The measurements can then be used to estimate the rate of gene migration. In equation 5.13, q_m = gene frequency in the European American population (the source of the "migrant" genes), $q_0 = 0.474$ (the original frequency in the African American population), and $q_t = 0.484$. As an approximate figure, we can suppose that the black population has been in the USA for 200–300 years, or about 10 generations. Then:

$$0.484 = 0.507 + (0.474 - 0.507)(1 - m)^{10}$$

. . . with a "migration" rate of about 3.5% per generation

This can be solved to find $m = 0.035$. That is, for every generation on average about 3.5% of the genes at the *MN* locus have migrated from the white population to the black population of Claxton. (Other estimates by the same method but using different gene loci suggest slightly different figures, more like 1%. The important point here is not the particular result; it is to illustrate how the population genetics of gene flow can be analyzed.) Notice again the rapid rate of genetic unification by migration: in only 10 generations, one-third of the gene frequency difference has been removed (after 10 generations the difference is 0.484 − 0.474, against the original difference of 0.507 − 0.474).

5.14.4 *A balance of selection and migration can maintain genetic differences between subpopulations*

If selection is working against an allele within one subpopulation, but the allele is continually being introduced by migration from other populations, it can be maintained by a balance of the two processes. We can analyze the balance between the two processes by much the same arguments as we used above for selection–mutation balance and heterozygous advantage. The simplest case is again for one locus with two alleles. Imagine selection in one subpopulation is working against a dominant *A* allele. The fitnesses of the genotypes are:

We construct a model of selection and migration

AA	Aa	aa
$1-s$	$1-s$	1

The A allele has frequency p in the local population. Suppose that in other subpopulations, natural selection is more favorable to the gene A, and it has a higher frequency in them, p_m on average. p_m will then be the frequency of A among immigrants to our local population. In the local population, A genes are lost at a rate ps per generation. They are gained at a rate $(p_m - p)m$ per generation: m is the proportion of genes that are immigrants in a generation. Immigration increases the frequency in the local population by an amount $p_m - p$ because gene frequency is increased only in so far as the immigrating population has a higher frequency of A than the local population. If the immigrating gene frequency is the same as the local gene frequency, immigration has no effect.

There are three possible outcomes. If migration is powerful relative to selection, the rate of gain of A genes by immigration will exceed the rate of loss by selection. The local population will be swamped by immigrants. The frequency of the A gene will increase until it reaches p_m. If migration is weak relative to selection, the frequency of A will decrease until it is locally eliminated. The third possibility is an exact balance between migration and selection. There will be an equilibrium (with local frequency of $A = p^*$) if:

Rate of gain of A by migration = rate of loss of A by selection

$$(p_m - p^*)m = p^*s \tag{5.14}$$

$$p^* = p_m\left(\frac{m}{s+m}\right) \tag{5.15}$$

In the first case, migration unifies the gene frequencies in both populations, much in the same manner as Section 5.14.2: migration is so strong relative to selection that it is as if selection were not operating. In the second and third cases, migration is not strong enough to unify the gene frequencies and we should observe regional differences in the gene frequency; it would be higher in some places than in others. In the third case there is a polymorphism within the local population; A is maintained by migration even though it is locally disadvantageous.

Polymorphism or genetic unity can result

This section has made two main points. First, a balance of migration and selection is another process to add to the list of processes that can maintain polymorphism. Second, we have seen how migration can be strong enough to unify gene frequencies between subpopulations, or if migration is weaker the gene frequencies of different subpopulations can diverge under selection. This theory is also relevant in the question of the relative importance of gene flow and selection in maintaining biological species (Section 13.7.2, p. 369).

Summary

1 In the absence of natural selection, and with random mating in a large population in which inheritance is Mendelian, the genotype frequencies at a locus move in one generation to the Hardy–Weinberg ratio; the genotype frequencies are then stable.

2 It is easy to observe whether the genotypes at a locus are in the Hardy–Weinberg ratio. In nature they will often not be, because the fitnesses of the genotypes are not equal, mating is non-random, or the population is small.

3 A theoretical equation for natural selection at a single locus can be written by expressing the frequency of a gene in one generation as a function of its frequency in the previous generation. The relation is determined by the fitnesses of the genotypes.

4 The fitnesses of the genotypes can be inferred from the rate of change of gene frequency in real cases of natural selection.

5 From the rate at which the melanic form of the peppered moth replaced the light-colored form, the melanic form must have had a selective advantage of about 50%.

6 The geographic pattern of melanic and light-colored forms of the peppered moth cannot be explained only by the selective advantage of the better camouflaged form. An inherent advantage to the melanic form, and migration, are also needed to explain the observations.

7 The evolution of resistance to pesticides in insects is in some cases due to rapid selection for a gene at a single locus. The fitness of the resistant types can be inferred, from the rate of evolution, to be as much as twice that of the non-resistant insects.

8 If a mutation is selected against but keeps on arising repeatedly, the mutation settles at a low frequency in the population. It is called selection–mutation balance.

9 Selection can maintain a polymorphism when the heterozygote is fitter than the homozygote and when fitnesses of genotypes are negatively frequency dependent.

10 Sickle cell anemia is an example of a polymorphism maintained by heterozygous advantage.

11 Subdivided populations have a higher proportion of homozygotes than an equivalent large, fused population.

12 Migration, in the absence of selection, rapidly unifies gene frequencies in different subpopulations; and it can maintain an allele that is selected against in a local subpopulation.

Further reading

There are a number of textbooks about population genetics. Crow (1986), Gillespie (1998), Hartl (2000), and Maynard Smith (1998) are relatively introductory. More comprehensive works include Hartl & Clark (1997) and Hedrick (2000). Crow & Kimura (1970) is a classic account of the mathematical theory. Dobzhansky (1970) is a standard study; Lewontin *et al.* (1981) contains Dobzhansky's most famous series of papers. Bell (1997a, 1997b) provides a comprehensive and a synoptic guide to selection.

For the peppered moth, Majerus (1998) is a modern, and Kettlewell (1973) a classic, account. Majerus (2002) is a more popular book, and contains a chapter on melanism. Grant (1999) is a review of Majerus (1998) and is also a good minireview of the topic in itself. Grant & Wiseman (2002) discuss the parallel rise and fall of the melanic form of the peppered moth in North America.

On pests and pesticides, see McKenzie (1996) and McKenzie & Batterham (1994). Lenormand *et al.* (1999) add further themes and molecular techniques, demonstrating seasonal cycles. The special issue of *Science* (4 October 2002, pp. 79–183) on the *Anopheles* genome has much background material on insecticide resistance and the various kinds of mosquito. See also Box 8.1 and Section 10.10, and their further reading lists.

See Endler (1986) on measuring fitness in general; Primack & Kang (1989) for plants; and Clutton-Brock (1988) for research on lifetime fitness.

The various selective means of maintaining polymorphisms are explained in the general texts. In addition, see Lederburg (1999) on the classic Haldane (1949a) paper and what it says about heterozygous advantage and sickle cell anemia. A recent possible example of heterozygote advantage in human HLA genes, providing resistance to HIV-1, is described by Carrington *et al.* (1999). Hori (1993) described a marvellous example of frequency dependence in the mouth-handedness of scale-eating cichlid fish. Another example is given by Gigord *et al.* (2001): the habits of naive bumblebees lead to a color polymorphism in an orchid.

Study and review questions

1 The following table gives genotype frequencies for five populations. Which are in Hardy–Weinberg equilibrium? For those that are not, suggest some hypotheses for why they are not.

	Genotype		
Population	*AA*	*Aa*	*aa*
1	25	50	25
2	10	80	10
3	40	20	40
4	0	150	100
5	2	16	32

2 For genotypes with the following fitnesses and frequencies at birth:

Genotype	*AA*	*Aa*	*aa*
Birth frequency	p^2	$2pq$	q^2
Fitness	1	1	$1-s$

(a) What is the frequency of *AA* individuals in the adult population? (b) What is the frequency of the gene *A* in the adult population? (c) What is the mean fitness of the population?

3 What is the mean fitness of this population?

Genotype	*AA*	*Aa*	*aa*
Birth frequency	1/3	1/3	1/3
Fitness	1	$1-s$	1

4 Consider a locus with two alleles, *A* and *a*. *A* is dominant and selection is working against the recessive homozygote. The frequency of *A* in two successive generations is 0.4875 and 0.5. What is the selection coefficient (s) against *aa*? (If you prefer to do it in your head rather than with a calculater, round the frequency of *a* in the first generation to 0.5 rather than 0.5125.)

5 What main assumption(s) is (or are) made in estimating fitnesses by the mark–recapture method?

6 Here are some adult genotype frequencies for a locus with two alleles. The polymorphism is known to be maintained by heterozygous advantage, and the fitneses of the genotypes are known to differ only in survival (and not infertility). What are the fitnesses (or selection coefficients) of the two homozygotes, relative to a fitness of 1 for the heterozygote?

Genotype	*AA*	*Aa*	*aa*
Frequency among adults	1/6	2/3	1/6

7 There are two populations of a species, called population 1 and population 2. Migrants move from population 1 to 2, but not vice versa. For a locus with two alleles *A* and *a*, in generation n, the gene frequency of *A* is 0.5 in population 1 and 0.75 in population 2; in generation 2 it is 0.5 in population 1 and 0.625 in population 2. (a) What is the rate of migration, measured as the chance an individual in population 2 is a first-generation immigrant from population 1? (b) If the rate of migration is the same in the next generation, what will the frequency of *A* be in population 2 in generation 3? [Questions 8–10 are more in the nature of questions for further thought. They are not about things explicitly covered in the chapter, but are slight extensions.]

8 What is the general effect of assortative mating on genotype frequencies, relative to the Hardy–Weinberg equilibrium, for (a) a locus with two alleles, one dominant to the other; and (b) a locus with two alleles, and no dominance (the heterozygote is a distinct phenotype intermediate between the two homozygotes)? And (c) what is the effect on genotype frequencies of a mating preference, in which females preferentially mate with males of (i) the dominant, and (ii) the recessive phenotype?

9 Derive a recurrence relation, giving the frequency of the dominant gene *A* one generation on (p') in terms of the frequency in any generation (p) and of the selection coefficient (s) for selection against the dominant allele.

10 Derive the expression for the equilibrium gene frequency (p^*) for the mutation–selection balance when the disadvantageous mutation is recessive.