

## Genetic Drift: Evolution at Random

# 10

One of the first and most important lessons a student of science learns is that many words have very different meanings in a scientific context than in everyday speech. The word “chance” is a good example. Many nonscientists think that evolution occurs “by chance.” What they mean is that evolution occurs without purpose or goal. But by this token, everything in the natural world—chemical reactions, weather, planetary movements, earthquakes—happens by chance, for none of these phenomena have purposes. In fact, scientists consider purposes or goals to be unique to human thought, and they do not view any natural phenomena as purposeful. But scientists don’t view chemical reactions or planetary movements as chance events, either—because in science, “chance” has a very different meaning.

Although the meaning of “chance” is a complex philosophical issue, scientists use chance, or **randomness**, to mean that when physical causes can result in any of several outcomes, we cannot predict what the outcome will be in any particular case. Nonetheless, we may be able to specify



### **Polymorphism in snails.**

Populations of the European land snail *Cepaea nemoralis* are genetically polymorphic for background color and for the number and width of the dark bands on their shells. Extensive research has shown that both genetic drift and natural selection affect the allele frequencies for these traits. (Photo by D. McIntyre.)

the *probability*, and thus the *frequency*, of one or another outcome. Although we cannot predict the sex of someone's next child, we can say with considerable certainty that there is a probability of 0.5 that it will be a daughter.

Almost all phenomena are affected simultaneously by both chance (unpredictable) and nonrandom, or DETERMINISTIC (predictable), factors. Any of us may experience a car accident due to the unpredictable behavior of other drivers, but we are predictably more likely to do so if we drive after drinking. So it is with evolution. As we will see in the next chapter, natural selection is a deterministic, nonrandom process. But at the same time, there are important random processes in evolution, including mutation (as discussed in Chapter 8) and random fluctuations in the frequencies of alleles or haplotypes: the process of **random genetic drift**.

Genetic drift and natural selection are the two most important causes of allele substitution—that is, of evolutionary change—in populations. Genetic drift occurs in all natural populations because, unlike ideal populations at Hardy-Weinberg equilibrium, natural populations are finite in size. Random fluctuations in allele frequencies can result in the replacement of old alleles by new ones, resulting in **nonadaptive evolution**. That is, while natural selection results in adaptation, genetic drift does not—so this process is not responsible for those anatomical, physiological, and behavioral features of organisms that equip them for reproduction and survival. Genetic drift nevertheless has many important consequences, especially at the molecular genetic level: it appears to account for much of the difference in DNA sequences among species.

Because all populations are finite, alleles at all loci are potentially subject to random genetic drift—but all are not necessarily subject to natural selection. For this reason, and because the expected effects of genetic drift can be mathematically described with some precision, some evolutionary geneticists hold the opinion that genetic drift should be the “null hypothesis” used to explain an evolutionary observation unless there is positive evidence of natural selection or some other factor. This perspective is analogous to the “null hypothesis” in statistics: the hypothesis that the data do not depart from those expected on the basis of chance alone.\* According to this view, we should not assume that a characteristic, or a difference between populations or species, is adaptive or has evolved by natural selection unless there is evidence for this conclusion.

The theory of genetic drift, much of which was developed by the American geneticist Sewall Wright starting in the 1930s, and by the Japanese geneticist Motoo Kimura starting in the 1950s, includes some of the most highly refined mathematical models in biology. (But fear not! We shall skirt around almost all the math.) We will first explore the theory and then see how it explains data from real organisms. In our discussion of the theory of genetic drift, we will describe random fluctuations in the frequencies (proportions) of two or more kinds of self-reproducing entities that do not differ *on average* (or differ very little) in reproductive success (fitness). For the purposes of this chapter, those entities are alleles. But the theory applies to any other self-replicating entities, such as chromosomes, asexually reproducing genotypes, or even species.

## The Theory of Genetic Drift

### Genetic drift as sampling error

That chance should affect allele frequencies is readily understandable. Imagine, for example, that a single mutation,  $A_2$ , appears in a large population that is otherwise  $A_1$ . If the population size is stable, each mating pair leaves an average of two progeny that survive to reproductive age. From the single mating  $A_1A_1 \times A_1A_2$  (for there is only one copy of  $A_2$ ), the probability that one surviving offspring will be  $A_1A_1$  is  $1/2$ ; therefore, the probability

\*For example, if we measure height in several samples of people, the null hypothesis is that the observed means differ from one another only because of random sampling, and that the parametric means of the populations from which the samples were drawn do not differ. A statistical test, such as a *t*-test or analysis of variance, is designed to show whether or not the null hypothesis can be rejected. It will be rejected if the sample means differ more than would be expected if samples had been randomly drawn from a single population.

that two surviving progeny will both be  $A_1A_1$  is  $\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$ —which is the probability that the  $A_2$  allele will be immediately lost from the population. We may assume that mating pairs vary at random, around the mean, in the number of surviving offspring they leave (0, 1, 2, 3 ...). In that case, as the pioneering population geneticist Ronald Fisher calculated, the probability that  $A_2$  will be lost, averaged over the population, is 0.368. He went on to calculate that after the passage of 127 generations, the cumulative probability that the allele will be lost is 0.985. This probability, he found, is not greatly different if the new mutation confers a slight advantage: as long as it is rare, it is likely to be lost, just by chance.

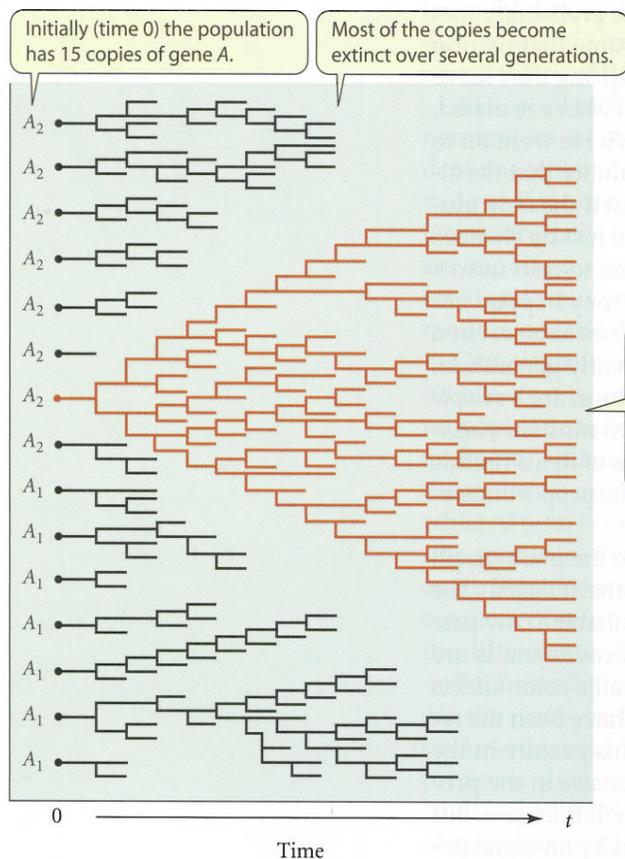
In this example, the frequency of an allele can change (in this instance, to zero from a frequency very near zero) because the one or few copies of the  $A_2$  allele may happen not to be included in those gametes that unite into zygotes, or may happen not to be carried by the offspring that survive to reproductive age. The genes included in any generation, whether in newly formed zygotes or in offspring that survive to reproduce, are a *sample* of the genes carried by the previous generation. Any sample is subject to random variation, or **sampling error**. In other words, the proportions of different kinds of items (in this case,  $A_1$  and  $A_2$  alleles) in a sample are likely to differ, by chance, from the proportions in the set of items from which the sample is drawn.

Imagine, for example, a population of land snails (*Cepaea nemoralis*; see the photograph that opens this chapter) in which (for the sake of argument) offspring inherit exactly the brown or yellow color of their mothers. Suppose 50 snails of each color inhabit a cow pasture. (The proportion of yellow snails is  $p = 0.50$ .) If 2 yellow and 4 brown snails are stepped on by cows,  $p$  will change to 0.511. Since it is unlikely that a snail's color affects the chance of its being squashed by cows, the change might just as well have been the reverse, and indeed, it may well be the reverse in another pasture, or in this pasture in the next generation. In this random process, the chances of increase or decrease in the proportion of yellow snails are equal in each generation, so the proportion will fluctuate. But an increase of, say, 1 percent in one generation need not be compensated by an equal decrease in a later generation—in fact, since this process is random, it is very unlikely that it will be. Therefore the proportion of yellow snails will wander over time, eventually ending up near, and finally at, one of the two possible limits: 0 and 1.0. It seems reasonable, too, that if the population should start out with, say, 80 percent brown and 20 percent yellow snails, it is more likely that the proportion of yellow will wander to zero than to 100 percent. In fact, the probability of yellow being lost from the population is exactly 0.20. Conversely, the probability that brown will reach 100 percent—that is, that it will be fixed—is 0.80.

### Coalescence

The concept of random genetic drift is so important that we will take two tacks in developing the idea. Figure 10.1 shows a hypothetical, but realistic, history of gene lineages. First, imagine the figure as depicting lineages of individual asexual organisms, such as bacteria, rather than genes. We know from our own experience that not all members of our parents' or grandparents' generations had equal numbers of descendants; some had none. Figure 10.1 diagrams this familiar fact. We note that the individuals in generation  $t$  (at the right of the figure) are the progeny of only some of those that existed in the previous generation ( $t - 1$ ): purely by chance, some individuals in generation  $t - 1$  failed to leave descendants. Likewise, the population at generation  $t - 1$  stems from only some of those individuals that existed in generation  $t - 2$ , and similarly back to the original population at time 0.

Now think of the objects in Figure 10.1 as copies of genes at a locus, in either a sexual or an asexual population. Figure 10.1 shows that as time goes on, more and more of the original gene lineages become extinct, so that the population consists of descendants of fewer and fewer of the original gene copies. In fact, if we look backward rather than forward in time, *all the gene copies* in the population ultimately *are descended from a single ancestral gene copy*, because given long enough, all other original gene lineages become extinct. The genealogy of the genes in the present population is said to **coalesce** back to a single common ancestor. Because that ancestor represents one of the several original alleles, the population's genes, descended entirely from that ancestral gene copy, must even-



**Figure 10.1** A possible history of descent of gene copies in a population that begins (at time 0, at left) with 15 copies, representing two alleles. Each gene copy has 0, 1, or 2 descendants in the next generation. The gene copies present at time  $t$  (at right) are all descended from (coalesce to) a single ancestral copy, which happens to be an  $A_2$  allele (the lineage shown in red). Gene lineages descended from all other gene copies have become extinct. If the failure of gene copies to leave descendants is random, then the gene copies at time  $t$  could equally likely have descended from any of the original gene copies present at time 0. (After Hartl and Clark 1989.)

By time  $t$ , all copies of the gene present in the population are descended from (coalesce to) a single ancestral gene copy.

tually become monomorphic: one or the other of the original alleles becomes fixed (reaches a frequency of 1.00). The smaller the population, the more rapidly all gene copies in the current population coalesce back to a single ancestral copy, since it takes longer for many than for few

gene lineages to become extinct by chance.

In our example, all gene copies have descended from a copy of an  $A_2$  allele, but because this is a random process,  $A_1$  might well have been the “lucky” allele if the sequence of random events had been different. If, in the generation that included the single common ancestor of all of today’s gene copies,  $A_1$  and  $A_2$  had been equally frequent ( $p = q = 0.5$ ), then it is equally likely that the ancestral gene copy would have been  $A_1$  or  $A_2$ ; but if  $A_1$  had had a frequency of 0.9 in that generation, then the probability is 0.9 that the ancestral gene would have been an  $A_1$  allele. *Our analysis therefore shows that by chance, a population will eventually become monomorphic for one allele or the other, and that the probability that allele  $A_1$  will be fixed, rather than another allele, equals the initial frequency of  $A_1$ .*

According to this analysis, for example, all the mitochondria of the entire human population are descended from the mitochondria carried by a single woman, who has been called “mitochondrial Eve,” at some time in the past. (Mitochondria are transmitted only through eggs.) This does *not* mean, however, that the population had only one woman at that time: “mitochondrial Eve” happened to be the one among many women to whom all mitochondria trace their ancestry (in a pattern like that seen in Figure 10.1). Various nuclear genes likewise are descended from single gene copies in the past that were carried by many different members of the ancestral human population.

If this process occurs in a large number of independent, non-interbreeding populations, each with the same initial number of copies of each of two alleles at, say, locus  $A$ , then we would expect a fraction  $p$  of the populations to become fixed for  $A_1$  and a fraction  $1 - p$  to become fixed for  $A_2$ . Thus the genetic composition of the populations would diverge by chance. If the original populations had each contained three (or more) different alleles, rather than two, each of those alleles would become fixed in some of the populations, with a probability equal to its initial frequency (say,  $p_i$ ).

As allele frequencies in a population change by genetic drift, so do the genotype frequencies, which conform to Hardy-Weinberg equilibrium among the new zygotes in each generation. If, for example, the frequencies  $p$  and  $q$  (that is,  $p$  and  $1 - p$ ) of alleles  $A_1$  and  $A_2$  change from 0.5:0.5 to 0.45:0.55, then the frequencies of genotypes  $A_1A_1$ ,  $A_1A_2$ , and  $A_2A_2$  change from 0.25:0.50:0.25 to 0.2025:0.4950:0.3025. As was described in Chapter 9, the frequency of heterozygotes,  $H$ , declines as one of the allele frequencies shifts closer to 1 (and the other moves toward 0):

$$H = 2p(1 - p)$$

Bear in mind that this model, as developed so far, includes only the effects of random genetic drift. It assumes that other evolutionary processes—namely, mutation, gene flow, and natural selection—do not operate. Thus the model does not describe the evolution of

adaptive traits—those that evolve by natural selection. We will incorporate natural selection in the following chapters.

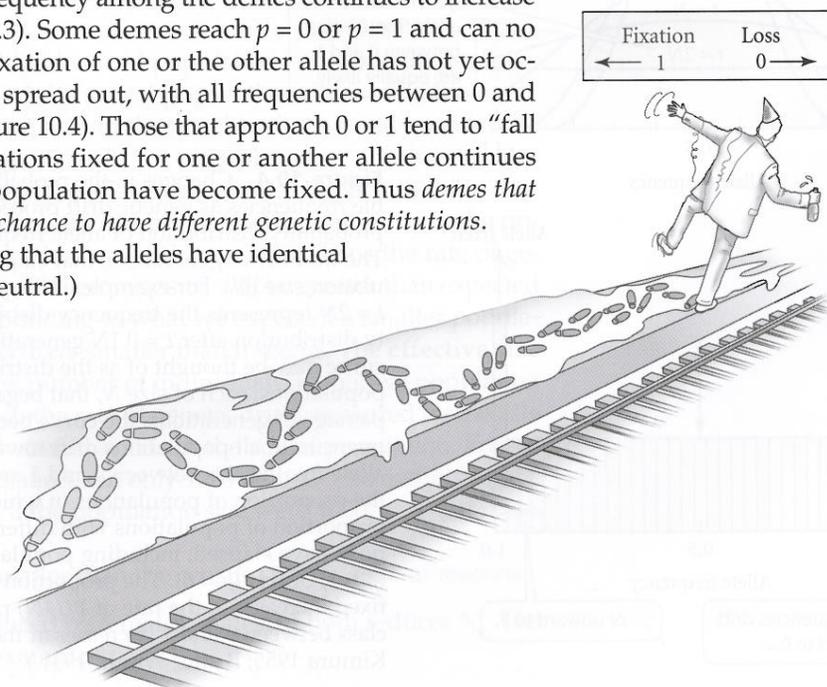
### Random fluctuations in allele frequencies

Let us take another, more traditional, approach to the concept of random genetic drift. Assume that the frequencies of alleles  $A_1$  and  $A_2$  are  $p$  and  $q$  in each of many independent populations, each with  $N$  breeding individuals (representing  $2N$  gene copies in a diploid species). Small independent populations are sometimes called **demes**, and an ensemble of such populations may be termed a **metapopulation**. As before, we assume that the phenotypes do not differ, *on average*, in survival or reproductive success—that is, the alleles are **neutral** with respect to fitness.

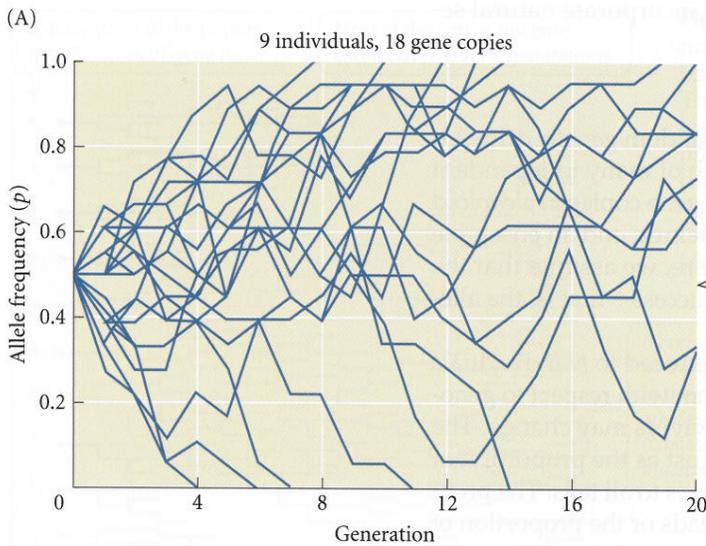
In each generation, the large number of newborn zygotes is reduced to  $N$  individuals by the time the next generation breeds, by mortality that is random with respect to genotype. By sampling error, the proportion of  $A_1$  ( $p$ ) among the survivors may change. The new  $p$  (call it  $p'$ ) could take on any possible value from 0 to 1.0, just as the proportion of heads among  $N$  tossed coins could, in principle, range from all heads to all tails. The probability of each possible value—whether it be the proportion of heads or the proportion of  $A_1$  allele copies—can be calculated from the binomial theorem, generating a **PROBABILITY DISTRIBUTION**. Among a large number of demes, the new allele frequency ( $p'$ ) will vary, by chance, around a mean—namely, the original frequency,  $p$ .

Now if we trace one of the demes, in which  $p$  has changed from 0.5 to, say, 0.47, we see that in the following generation, it will change again from 0.47 to some other value, *either higher or lower with equal probability*. This process of random fluctuation continues over time. Since no stabilizing force returns the allele frequency toward 0.5,  $p$  will eventually wander (drift) either to 0 or to 1: *the allele is either lost or fixed*. (Once the frequency of an allele has reached either 0 or 1, it cannot change unless another allele is introduced into the population, either by mutation or by gene flow from another population.) The allele frequency describes a **random walk**, analogous to a New Year's Eve reveler staggering along a very long train platform with a railroad track on either side: if he is so drunk that he doesn't compensate steps toward one side with steps toward the other, he will eventually fall off the edge of the platform onto one of the two tracks, if the platform is long enough (Figure 10.2).

Just as an allele's frequency may increase by chance in some demes from one generation to the next, it may decrease in other demes. As a result, allele frequencies may vary among the demes. The *variance* in allele frequency among the demes continues to increase from generation to generation (Figure 10.3). Some demes reach  $p = 0$  or  $p = 1$  and can no longer change. Among those in which fixation of one or the other allele has not yet occurred, the allele frequencies continue to spread out, with all frequencies between 0 and 1 eventually becoming equally likely (Figure 10.4). Those that approach 0 or 1 tend to “fall over the edge,” so the number of populations fixed for one or another allele continues to increase, until all demes in the metapopulation have become fixed. Thus *demes that initially are genetically identical evolve by chance to have different genetic constitutions*. (Remember, though, that we are assuming that the alleles have identical effects on fitness—that is, that they are neutral.)

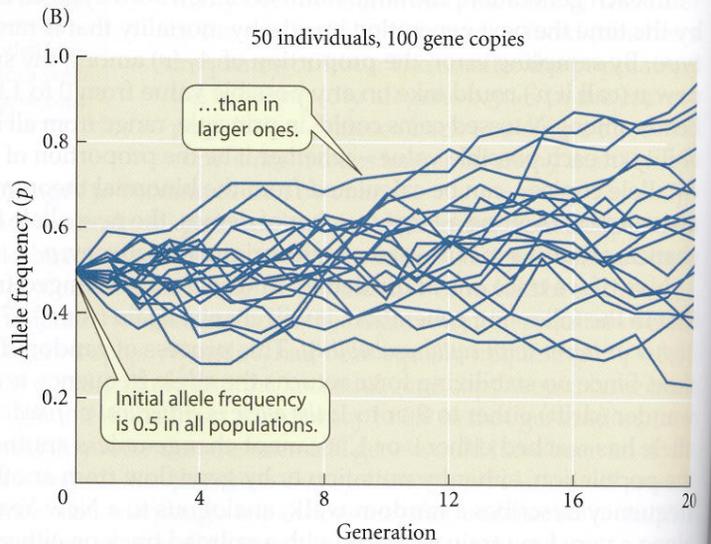


**Figure 10.2** A “random walk” (or “drunkard’s walk”). The reveler eventually falls off the platform if he is too far gone to steer a course toward the middle. The edges of the platform (“0” and “1”) represent loss and fixation of an allele, respectively.



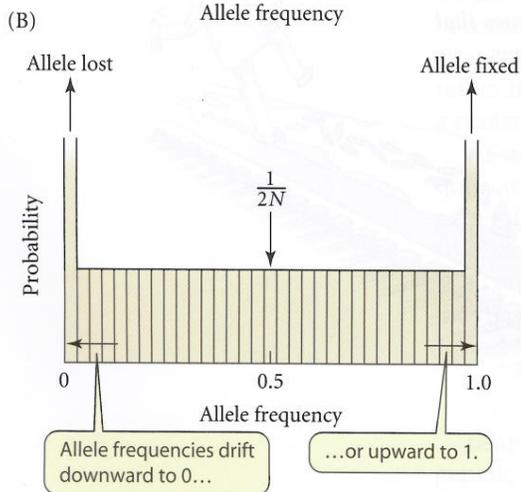
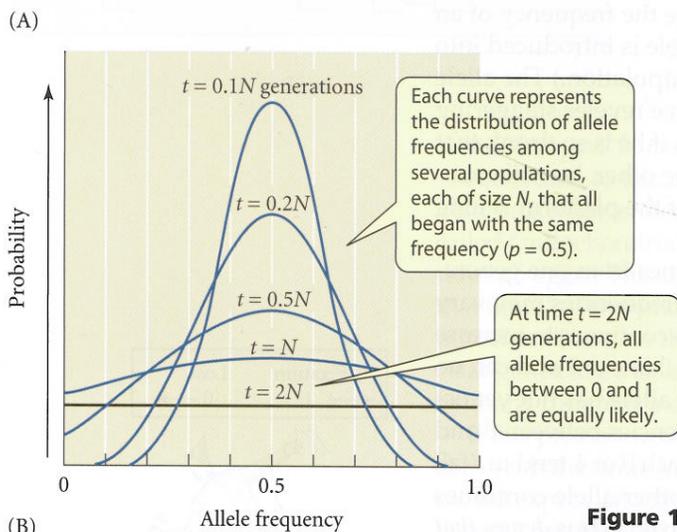
**Figure 10.3** Computer simulations of random genetic drift in populations of (A) 9 diploid individuals ( $2N = 18$  gene copies) and (B) 50 diploid individuals ( $2N = 100$  gene copies). Each line traces the frequency ( $p$ ) of one allele for 20 generations. Each panel shows allele frequency changes in 20 replicate populations, all of which begin at  $p = 0.5$  (i.e., half the gene copies are  $A_1$  and half  $A_2$ ). (After Hartl and Clark 1989.)

Oscillations are larger, and alleles are more rapidly fixed or lost, in small populations...



...than in larger ones.

Initial allele frequency is 0.5 in all populations.



**Figure 10.4** Changes in the probability that an allele will have various possible frequencies as genetic drift proceeds over time. (A) Each curve shows the probability distribution of allele frequencies between 0 and 1 at different times. The number of generations that elapse ( $t$ ) is measured in units of the initial population size ( $N$ ). For example, if the population begins with  $N = 50$  individuals,  $t = 2N$  represents the frequency distribution after 100 generations. The probability distribution after  $t = 0.1N$  generations is shown by the uppermost curve. This curve may be thought of as the distribution of allele frequencies among several populations, each of size  $N$ , that began with the same allele frequency. With the passage of generations, the curve becomes lower and broader as the allele frequencies in all populations drift toward either 0 or 1. At  $t = 2N$  generations, all allele frequencies between 0 and 1 are equally likely. (This panel does not show the proportion of populations in which the allele has been fixed or lost.) (B) The proportion of populations with different allele frequencies after  $t = 2N$  generations have elapsed, including populations in which the allele has been fixed ( $p = 1$ ) or lost ( $p = 0$ ). The proportion of populations in which the allele is lost or fixed increases at the rate of  $1/(4N)$  per generation, and each allele frequency class between 0 and 1 decreases at the rate of  $1/(2N)$  per generation. (A after Kimura 1955; B after Wright 1931.)

## Evolution by Genetic Drift

The following points, which follow from the previous discussion, are some of the most important aspects of evolution by genetic drift:

- 1. Allele (or haplotype) frequencies fluctuate at random within a population, and eventually one or another allele becomes fixed.
- 2. Therefore, the genetic variation at a locus declines and is eventually lost. As the frequency of one of the alleles approaches 1.0, the frequency of heterozygotes,  $H = 2p(1 - p)$ , declines. The rate of decline in heterozygosity is often used as a measure of the rate of genetic drift within a population.
- 3. At any time, an allele's probability of fixation equals its frequency at that time, and is not affected or predicted by its previous history of change in frequency.
- 4. Therefore, populations with the same initial allele frequency ( $p$ ) diverge, and a proportion  $p$  of the populations is expected to become fixed for that allele. A proportion  $1 - p$  of the populations becomes fixed for alternative alleles.
- 5. If an allele has just arisen by mutation, and is represented by only one among the  $2N$  gene copies in the population, its frequency is

$$p_t = \frac{1}{2N}$$

and this is its likelihood of reaching  $p = 1$ . Clearly, it is more likely to be fixed in a small than in a large population. Moreover, if the same mutation arises in each of many demes, each of size  $N$ , the mutation should eventually be fixed in a proportion  $1/(2N)$  of the demes. Similarly, of all the new mutations (at all loci) that arise in a population, a proportion  $1/(2N)$  should eventually be fixed.

Evolution by genetic drift proceeds faster in small than in large populations. In a diploid population, the average time to fixation of a newly arisen neutral allele that does become fixed is  $4N$  generations, on average. That is a long time if the population size ( $N$ ) is large.

Among a number of initially identical demes in a metapopulation, the average allele frequency ( $\bar{p}$ ) does not change, but since the allele frequency in each deme does change, eventually becoming 0 or 1, the frequency of heterozygotes ( $H$ ) declines to zero in each deme and in the metapopulation as a whole.

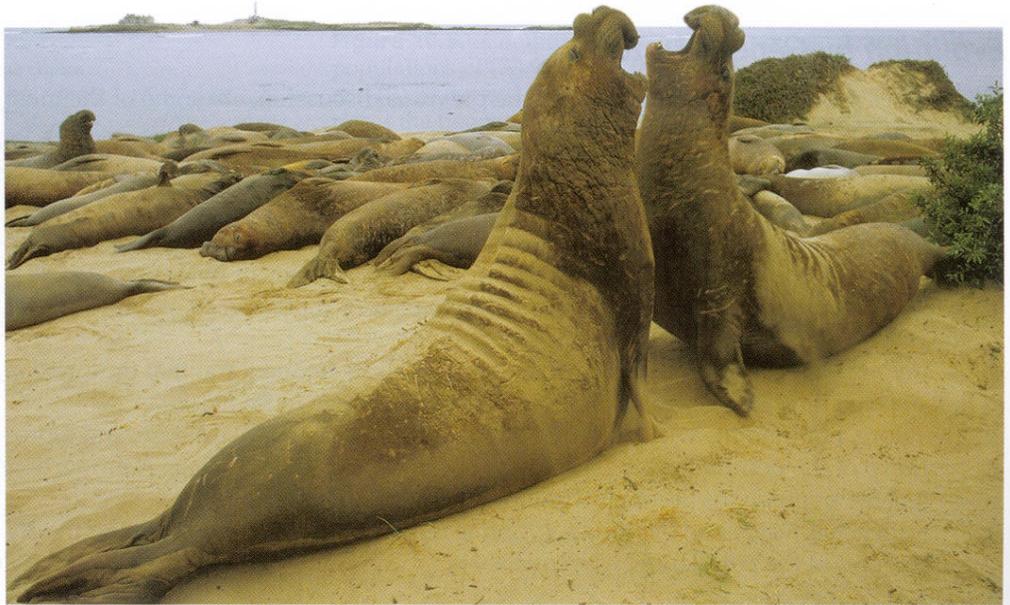
## Effective population size

The theory presented so far assumes highly idealized populations of  $N$  breeding adults. We measure the actual number ( $N$ ) of adults in real populations, however, the number we count (the CENSUS SIZE) may be greater than the number that actually contribute genes to the next generation. Among elephant seals, for example, a few dominant males mate with all the females in a population, so the alleles those males happen to carry contribute disproportionately to following generations; from a genetic point of view, the unsuccessful subdominant males might as well not exist (Figure 10.5). Thus the rate of genetic drift of allele frequencies, and of loss of heterozygosity, will be greater than expected from the population's census size, corresponding to what we expect of a smaller population. In other words, the population is *effectively* smaller than it seems. The **effective size** (denoted  $N_e$ ) of an actual population is the number of individuals in an ideal population (in which every adult reproduces) in which the rate of genetic drift (measured by the rate of decline in heterozygosity) would be the same as it is in the actual population. For instance, if we count 10,000 adults in a population, but only 1000 of them successfully breed, genetic drift proceeds at the same rate as if the population size were 1000, and that is the effective size,  $N_e$ .

The effective population size can be smaller than the census size for several reasons:

1. *Variation in the number of progeny* produced by females, males, or both reduces  $N_e$ . The elephant seal represents an extreme example.

**Figure 10.5** The effective population size among northern elephant seals (*Mirounga angustirostris*) is much lower than the census size because only a few of the large males compete successfully for the smaller females. The winner of the contest here will father the offspring of an entire “harem” of females. (Photo © Richard Hansen/Photo Researchers.)



2. Similarly, a *sex ratio* different from 1:1 lowers the effective population size.
3. *Natural selection* can lower  $N_e$  by increasing variation in progeny number; for instance, if larger individuals have more offspring than smaller ones, the rate of genetic drift may be increased at all neutral loci because small individuals contribute fewer gene copies to subsequent generations.
4. If *generations overlap*, offspring may mate with their parents, and since these pairs carry identical copies of the same genes, the effective number of genes propagated is reduced.
5. Perhaps most importantly, *fluctuations in population size* reduce  $N_e$ , which is more strongly affected by the smaller than by the larger sizes. For example, if the number of breeding adults in five successive generations is 100, 150, 25, 150, and 125,  $N_e$  is approximately 70 (the harmonic mean\*) rather than the arithmetic mean, 110.

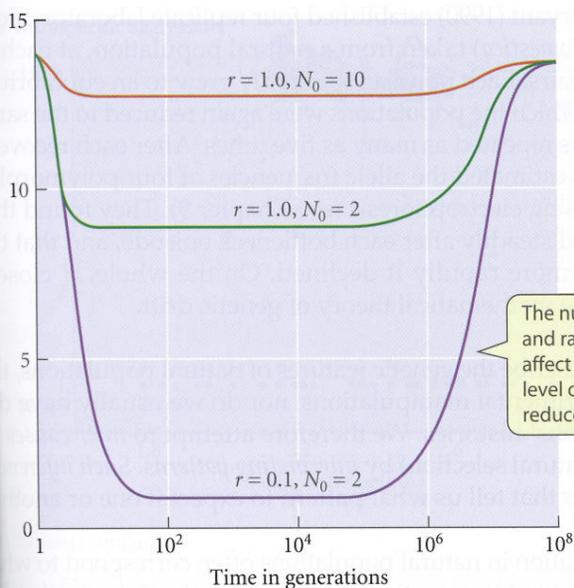
### Founder effects

Restrictions in size through which populations may pass are called **bottlenecks**. A particularly interesting bottleneck occurs when a new population is established by a small number of colonists, or founders—sometimes as few as a single mating pair (or a single inseminated female, as in insects in which females store sperm). The random genetic drift that ensues is often called a **founder effect**. If the new population rapidly grows to a large size, allele frequencies (and therefore heterozygosity) will probably not be greatly altered from those in the source population, although some rare alleles will not have been carried by the founders. If the colony remains small, however, genetic drift will alter allele frequencies and erode genetic variation. If the colony persists and grows, new mutations eventually restore heterozygosity to higher levels (Figure 10.6).

### Genetic drift in real populations

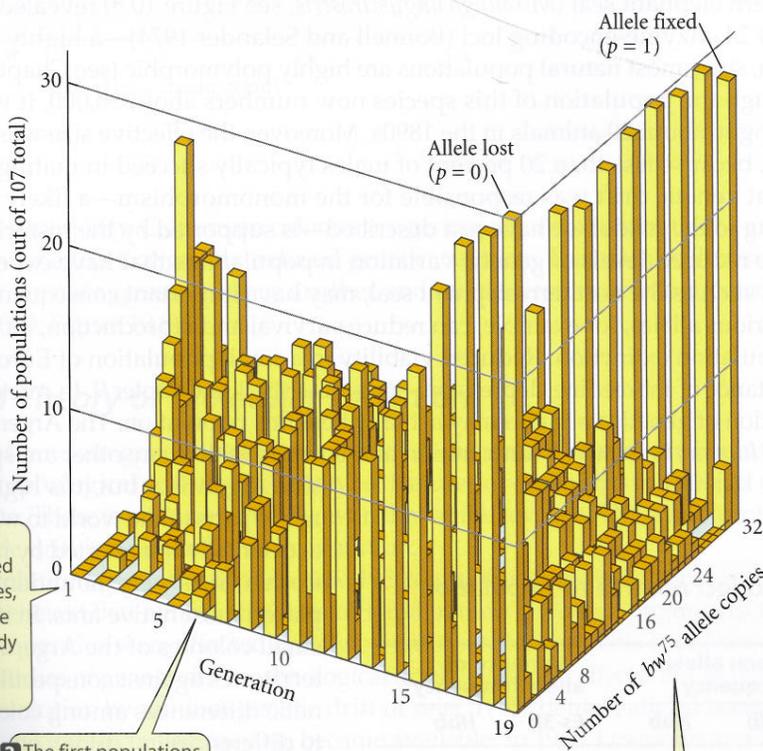
**LABORATORY POPULATIONS.** Peter Buri (1956) described genetic drift in an experiment with *Drosophila melanogaster*. He initiated 107 experimental populations of flies, each with 8 males and 8 females, all heterozygous for two alleles ( $bw$  and  $bw^{75}$ ) that affect eye color (by which all three genotypes are recognizable). Thus the initial frequency of  $bw^{75}$  was 0.5 in all populations. He propagated each population for 19 generations by drawing 8 flies of each sex at random and transferring them to a vial of fresh food. (Thus each generation

\*The HARMONIC MEAN is the reciprocal of the average of a set of reciprocals. If the number of breeding individuals in a series of  $t$  generations is  $N_0, N_1, \dots, N_t$ ,  $N_e$  is calculated from  $1/N_e = (1/t)(1/N_0 + 1/N_1 + \dots + 1/N_t)$ .



**Figure 10.6** Effects of a bottleneck in population size on genetic variation, as measured by heterozygosity. Heterozygosity is reduced more if the number of founders is lower ( $N_0 = 2$ ) than if it is higher ( $N_0 = 10$ ; uppermost curve), and if the rate of population increase is lower ( $r = 0.1$ , lowest curve) than if it is higher ( $r = 1.0$ ). Eventually, mutation supplies new genetic variation, and heterozygosity increases. (After Nei et al. 1975.)

as initiated with 16 flies  $\times$  2 gene copies = 32 gene copies.) The frequency of  $bw^{75}$  rapidly spread out among the populations (Figure 10.7); after one generation, the number of  $bw^{75}$  copies ranged from 7 ( $q = 7/32 = 0.22$ ) to 22 ( $q = 0.69$ ). By generation 19, 30 populations had lost the  $bw^{75}$  allele, and 28 had become fixed for it; among the unfixed populations, intermediate allele frequencies were quite evenly distributed. The results nicely matched those expected from genetic drift theory (see Figure 10.4).



**Figure 10.7** Random genetic drift in 107 experimental populations of *Drosophila melanogaster*, each founded with 16  $bw^{75}/bw$  heterozygotes, and each propagated by 16 flies (8 males and 8 females) per generation. The frequency distribution of the number of  $bw^{75}$  copies is read from front to back, and the generations of offspring proceed from left to right. The number of  $bw^{75}$  alleles, which began at 16 copies in the parental populations (i.e., a frequency of 0.5) became more evenly distributed between 0 and 32 copies with the passage of generations, and the  $bw^{75}$  allele was lost (0 copies) or fixed (32 copies) in an increasing number of populations. (After Hartl and Clark 1989.)

1 Although all 107 populations started with 16  $bw^{75}$  alleles, in generation 1 the populations already varied in allele frequency.

2 The first populations to lose the  $bw^{75}$  allele appeared in generation 6.

3 After 19 generations, allele frequencies (number of  $bw^{75}$  alleles) had become more evenly distributed between 0 and 1.0 (32 copies), and the  $bw^{75}$  allele was lost (0) or fixed (32) in an increasing number of populations.

More recently, McCommas and Bryant (1990) established four replicate laboratory populations, using houseflies (*Musca domestica*) taken from a natural population, at each of three bottleneck sizes: 1, 4, and 16 pairs. Each population rapidly grew to an equilibrium size of about a thousand flies, after which the populations were again reduced to the same bottleneck sizes. This procedure was repeated as many as five times. After each recovery from a bottleneck, the investigators estimated the allele frequencies at four polymorphic enzyme loci for each population, using electrophoresis (see Chapter 9). They found that average heterozygosity ( $\bar{H}$ ) declined steadily after each bottleneck episode, and that the smaller the bottlenecks were, the more rapidly it declined. On the whole,  $\bar{H}$  closely matched the values predicted by the mathematical theory of genetic drift.

**NATURAL POPULATIONS.** When we describe the genetic features of natural populations, the data usually are not based on experimental manipulations, nor do we usually have detailed information on the populations' histories. We therefore attempt to *infer causes* of evolution (such as genetic drift or natural selection) by *interpreting patterns*. Such inferences are possible only on the basis of theories that tell us what pattern to expect if one or another cause has been most important.

Patterns of molecular genetic variation in natural populations often correspond to what we would expect if the loci were affected by genetic drift. For example, Robert Selander (1970) studied allozyme variation at two loci in house mice (*Mus musculus*) from widely scattered barns in central Texas. Selander considered each barn to harbor an independent population because mice rather seldom disperse to new barns, and those that do are often excluded by the residents. Having estimated the population size in each barn, Selander found that although small and large populations had much the same *mean* allele frequencies, the *variation* (variance) in allele frequency was much greater among the small populations, as we would expect from random genetic drift (Table 10.1).

Occasionally, we can check the validity of our inferences using independent information, such as historical data. For example, a survey of electrophoretic variation in the northern elephant seal (*Mirounga angustirostris*; see Figure 10.5) revealed *no* variation at any of 24 enzyme-encoding loci (Bonnell and Selander 1974)—a highly unusual observation, since most natural populations are highly polymorphic (see Chapter 9). However, although the population of this species now numbers about 30,000, it was reduced by hunting to about 20 animals in the 1890s. Moreover, the effective size was probably even lower, because less than 20 percent of males typically succeed in mating. The hypothesis that genetic drift was responsible for the monomorphism—a likely hypothesis according to the model we have just described—is supported by the historical data.

The reduced levels of genetic variation in populations that have experienced bottlenecks, such as the northern elephant seal, may have important consequences. Fixation of deleterious alleles, for example, can reduce survival and reproduction, increasing the risk of population extinction. Reduced viability in a small population of European adders—an instance of inbreeding depression—was described in Chapter 9. In rare cases, however, reduction of genetic variation may actually benefit a population. The Argentine ant (*Linepithema humile*) is relatively uncommon and coexists with many other ant species in its native Argentina, but it is highly invasive in many parts of the world to which it has been

accidentally transported by humans. In California, it is very abundant and has displaced most native ants. In its native range, small colonies of the Argentine ant defend territories against conspecific colonies. Genetic differences among colonies gives rise to differences in "colony odor," which elicit aggression. In California, however, colonies merge with one another to form very large, widely distributed "supercolonies" that competitively exclude other ant species by

**TABLE 10.1** Frequency of alleles at two loci relative to population size of house mice

Estimated population size	Number of populations sampled	Mean allele frequency		Variance of allele frequency <sup>a</sup>	
		<i>Es-3b</i>	<i>Hbb</i>	<i>Es-3b</i>	<i>Hbb</i>
Small (median = 10)	29	0.418	0.849	0.0506	0.1883
Large (median = 200)	13	0.372	0.843	0.0125	0.0083

Source: After Selander 1970.

<sup>a</sup>Note that the variance of allele frequency is greater among small than among large populations.