4

Genetic Drift

In deriving the Hardy–Weinberg Law in Chapter 2, we assumed that the population size was infinite. In some derivations of the Hardy–Weinberg Law, this assumption is not stated explicitly, but it enters implicitly by the act of equating allele and genotype *probabilities* to allele and genotype *frequencies*. The allele frequency is simply the number of alleles of a given type divided by twice the total population size for an autosomal locus in a diploid species. Likewise, the genotype frequency is simply the number of individuals with a specified genotype divided by the total population size. But in deriving the Hardy–Weinberg Law in Chapter 2, we actually calculated allele and genotype *probabilities*. For example, we calculated the *probability* of drawing an A allele from the gene pool as p and then stated that this is also the *frequency* of the A allele in the next generation. This stems from the common definition that the probability of an event is the frequency of the event in an infinite number of trials (Appendix A). But what happens when the population is finite in size so there is not an infinite number of trials? For example, suppose that a population has a gene pool with two alleles, say H and T, such that the probability of drawing either allele is 0.5 (i.e. $p = q = {}^1/_2$). Now, suppose that $2N$ (a finite number) gametes are drawn from this gene pool to form the next generation of N diploid adults. Will the *frequency* of H and T be 0.5 in this finite population?

You can simulate this situation. For example, let $N = 5$ (corresponding to a sample of 10 gametes), and place 10 coins in a box, shake the box, and count the number of heads (i.e. allele "H"). Suppose that after doing this experiment one time, six heads were observed. Hence, the frequency of the H "allele" in this simulation was $6/10 = 0.6$, which is not the same as 0.5, the probability of H. Figure 4.1a shows the results of doing this coin flip simulation 20 times, and you are strongly encouraged to do this experiment yourself. As you can see from Figure 4.1a or from your own simulations, when population size is finite, the *frequency* of an allele in the next generation is often not the same as the *probability* of drawing that allele from the gene pool. Given our definition of evolution as a change in allele frequency, we note that this random sampling error can induce evolution. **Genetic drift** is the random change in allele frequency due to sampling error in a finite population**.** Genetic drift is an evolutionary force that can alter the genetic make-up of a population's gene pool through time and shows that the Hardy–Weinberg "equilibrium" and its predicted stability of allele frequencies do not hold exactly for any finite population. The purpose of this chapter is to investigate the evolutionary properties and significance of genetic drift.

Population Genetics and Microevolutionary Theory, Second Edition. Alan R. Templeton.

Companion website: www.wiley.com/go/templeton/populationgenetics2

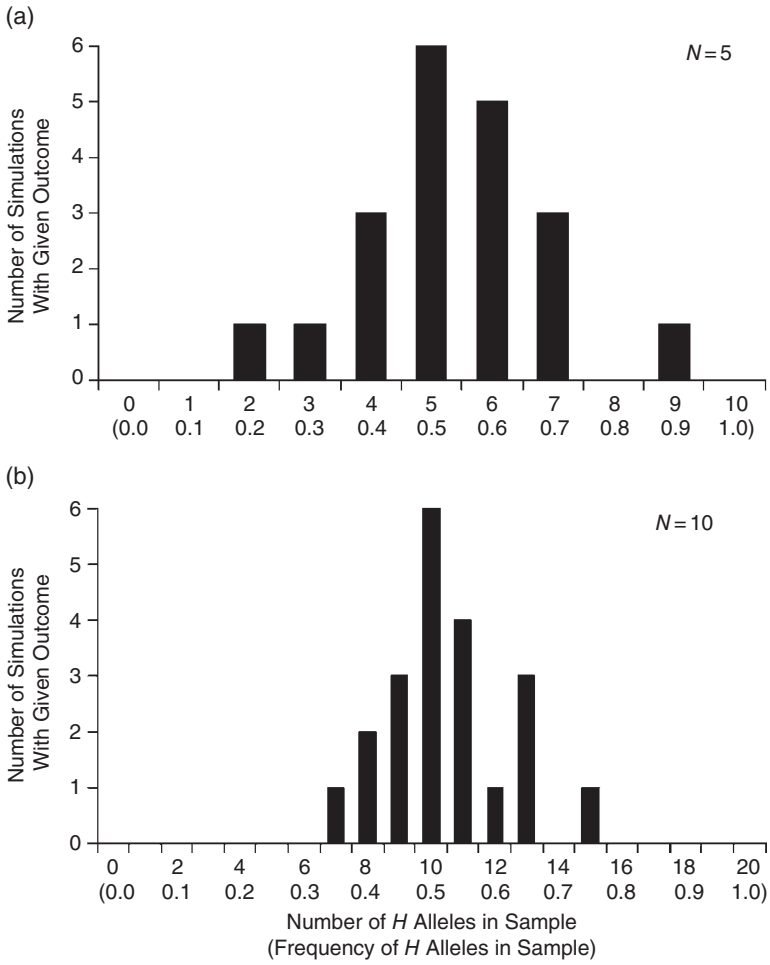

Figure 4.1 (a) The number of heads (the *H* "allele") observed after shaking a box with 10 coins. The experiment was repeated 20 times. (b) The number of heads (the *H* "allele") observed after shaking a box with 20 coins. The experiment was repeated 20 times.

Basic Evolutionary Properties of Genetic Drift

The coin flip simulation shown in Figure 4.1a illustrates that finite population size can induce random changes in allele frequencies due to sampling error. But what exactly is the relationship between genetic drift and finite population size? To answer this question, repeat the coin flip experiment but now use 20 coins for the simulation of a diploid population of $N = 10$ and $2N = 20$ gametes (double our previous population size). Figure 4.1b shows the results of such a simulation repeated 20 times. As can be seen from Figure 4.1b, there are random deviations from 0.5, so genetic drift is still operating in this larger but still finite population. By comparing Figures 4.1a and b, an important property of genetic drift is revealed: the simulated frequencies are more tightly clustered around 0.5 when $2N = 20$ (Figure 4.1b) than when $2N = 10$ (Figure 4.1b). This means that, on the average, the observed allele frequencies deviate less from the expected allele probability when the population size is larger. Thus, the amount of evolutionary change associated with random

sampling error is *inversely* related to population size. The larger the population, the less the allele frequency will change on the average. Hence, genetic drift is most powerful as an evolutionary force when N is small.

In the coin box experiments, the outcome was about equally likely to deviate above and below 0.5 (Figures 4.1a and b). Hence, for a large number of identical populations, the overall allele frequency remains 0.5, although in any individual population, it is quite likely that the allele frequency will change from 0.5. The fact that deviations are equally likely above and below 0.5 simply means that there is *no direction* to genetic drift. Although we can see that finite population size is likely to alter allele frequencies due to sampling error, we cannot predict the precise outcome or even the direction of the change in any specific population.

The coin box simulations given in Figure 4.1 only simulate one generation of genetic drift starting with an initial allele frequency of 0.5. The coin box simulations do not simulate the impact of drift over multiple generations because the *probability* of a coin flip producing an H allele remains unchanged at 0.5. However, suppose drift caused the allele frequency to change from 0.5 to 0.6 in one particular population. How about the next generation? Is it equally likely to be above or below 0.5, as it was in the first generation and will always be in our coin flip simulations? The answer is no, drift at one generation is always centered around the allele frequency of the previous generation, and allele frequencies in more ancient generations are irrelevant. Thus, after the allele frequency drifts to 0.6 from 0.5, the probability of drawing an H allele is now 0.6 and sampling error in the second generation is centered around 0.6 and not 0.5. This in turn means that after two generations of drift and given that the first generation experienced a deviation above 0.5, it is no longer true that deviations will be equally likely above and below 0.5. Once the population drifted to a frequency of 0.6, the next generation's allele frequency is more likely to stay above 0.5. Under genetic drift, there is *no tendency to return to ancestral allele frequencies*. With each passing generation, it becomes more and more likely to deviate from the initial conditions.

The action of drift over several generations can be simulated using a computer in which each generation drifts around the allele frequency of the previous generation. Figure 4.2 shows the results of 20 replicates of simulated drift in diploid populations of size 10 ($2N = 20$) over multiple generations, and Figure 4.3 shows the results in populations of size 25 ($2N = 50$). In both cases, the initial allele frequency starts at 0.5, but, with increasing generation number, more and more of the populations deviate from 0.5, and by larger amounts. As can be seen by contrasting Figure 4.2 ($N = 10$) with Figure 4.3 ($N = 25$), the smaller population size tends to have more radical changes in allele frequency in a given amount of time, as was shown in Figure 4.1 for one generation. However, Figure 4.3 shows that even with the larger population size of 25, substantial changes have occurred by generation 10. In general, we expect to obtain larger and larger deviations from the initial conditions with increasing generation time. Figures 4.2 and 4.3 show that N determines the rate of change caused by drift and that even large populations can be affected by drift if given enough time. The evolutionary changes in allele frequencies caused by genetic drift *accumulate with time*.

Also, note in these simulations (particularly for $N = 10$) that eventually populations tend to go to allele frequencies of 0 (loss of the allele) or 1 (fixation of the allele). Genetic drift, like any other evolutionary force, can only operate as an evolutionary force when there is genetic variability. Hence, as long as p is not equal to 0 or 1, drift will cause changes in allele frequency. However, once an allele is lost or fixed, genetic drift can no longer cause allele frequency changes (all evolution requires genetic variation). Once lost or fixed, the allele stays lost or fixed, barring new mutations or the reintroduction of allelic variation by genetic interchange with an outside population. Genetic drift is like a room with flypaper on all the walls. The walls represent loss and fixation, and

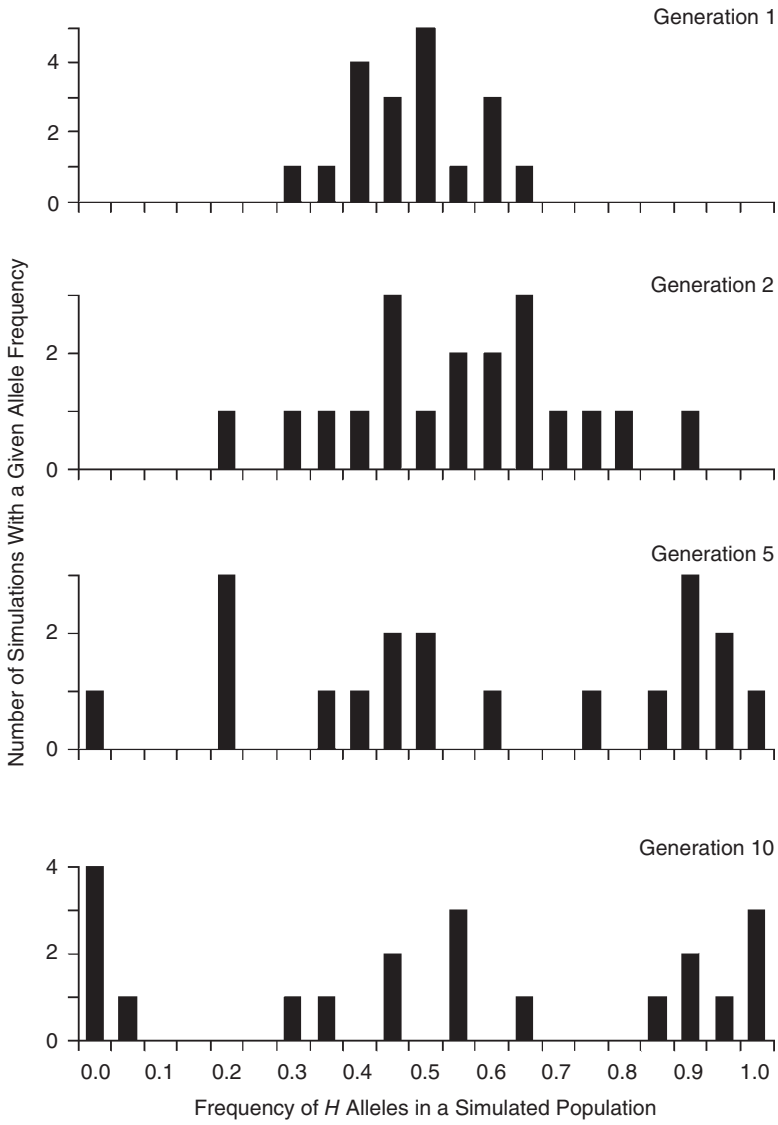

Figure 4.2 Results of simulating 20 replicates of a finite population of size 10 ($2N = 20$) for 10 generations starting from an initial gene pool of $p = {}^{1}/_{2}$. The distribution of allele frequencies is shown after 1, 2, 5, and 10 generations of genetic drift.

sooner or later (depending upon population size, which in this analogy is directly related to the size of the room), the fly (allele frequency) will hit a wall and be "stuck." Genetic drift causes *a loss of genetic variation* within a finite population.

In Figures 4.1–4.3, we simulated several replicates of the initial population. Now suppose that several subpopulations are established from a common ancestral population such that they are all genetically isolated from one another (that is, no gametes are exchanged between the subpopulations). Population subdivision into isolated demes is called **fragmentation**. Figures 4.1–4.3 can

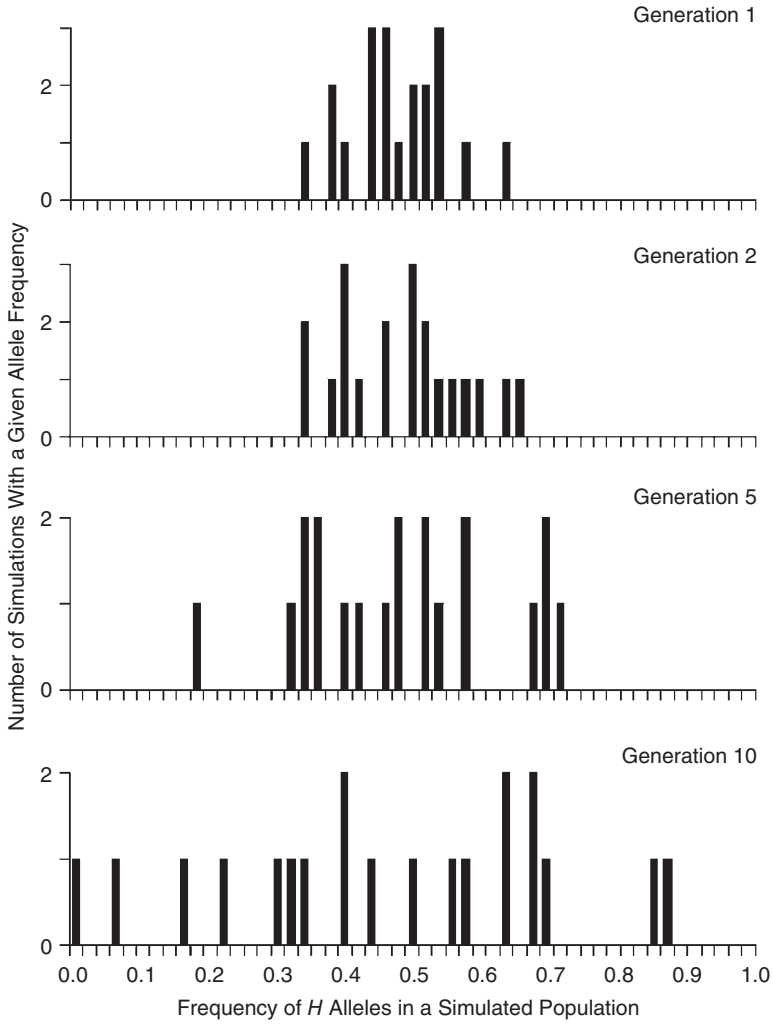

Figure 4.3 Results of simulating 20 replicates of a finite population of size 25 ($2N = 50$) for 10 generations starting from an initial gene pool of $p = {}^1/_2$. The distribution of allele frequencies is shown after 1, 2, 5, and 10 generations of genetic drift.

therefore also be regarded as simulations of population fragmentation of a common ancestral population such that the fragmented subpopulations are all of equal size. Note that the ancestral gene pool is the same ($p = 0.5$) in all the populations simulated. Therefore, these same figures allow us to examine the role of genetic drift upon fragmented populations. Now, we shift our focus from the evolution within each fragmented deme to the evolution of changes between subpopulations. Because of the genetic isolation under fragmentation, drift will operate independently in each subpopulation. Because of the randomness of the evolutionary direction of drift, it is unlikely that all the independent subpopulations will evolve in the same direction. This is shown in Figures 4.1–4.3 by regarding the replicate elements of the histograms as isolated subpopulations. The spread of these histograms around the initial allele frequency shows that different subpopulations evolve

away from 0.5 in different directions and magnitudes. Thus, although each subpopulation began with the same allele frequency, they now have many different allele frequencies. Genetic drift causes an *increase of allele frequency differences* among finite subpopulations.

All of these properties of genetic drift have been demonstrated empirically by Buri (1956), as shown in Figure 4.4. He initiated 107 populations of 8 males and 8 females of the fruit fly *Drosophila melanogaster*, all with two eye color alleles (*bw* and bw^{75}) at equal frequency. He then followed the

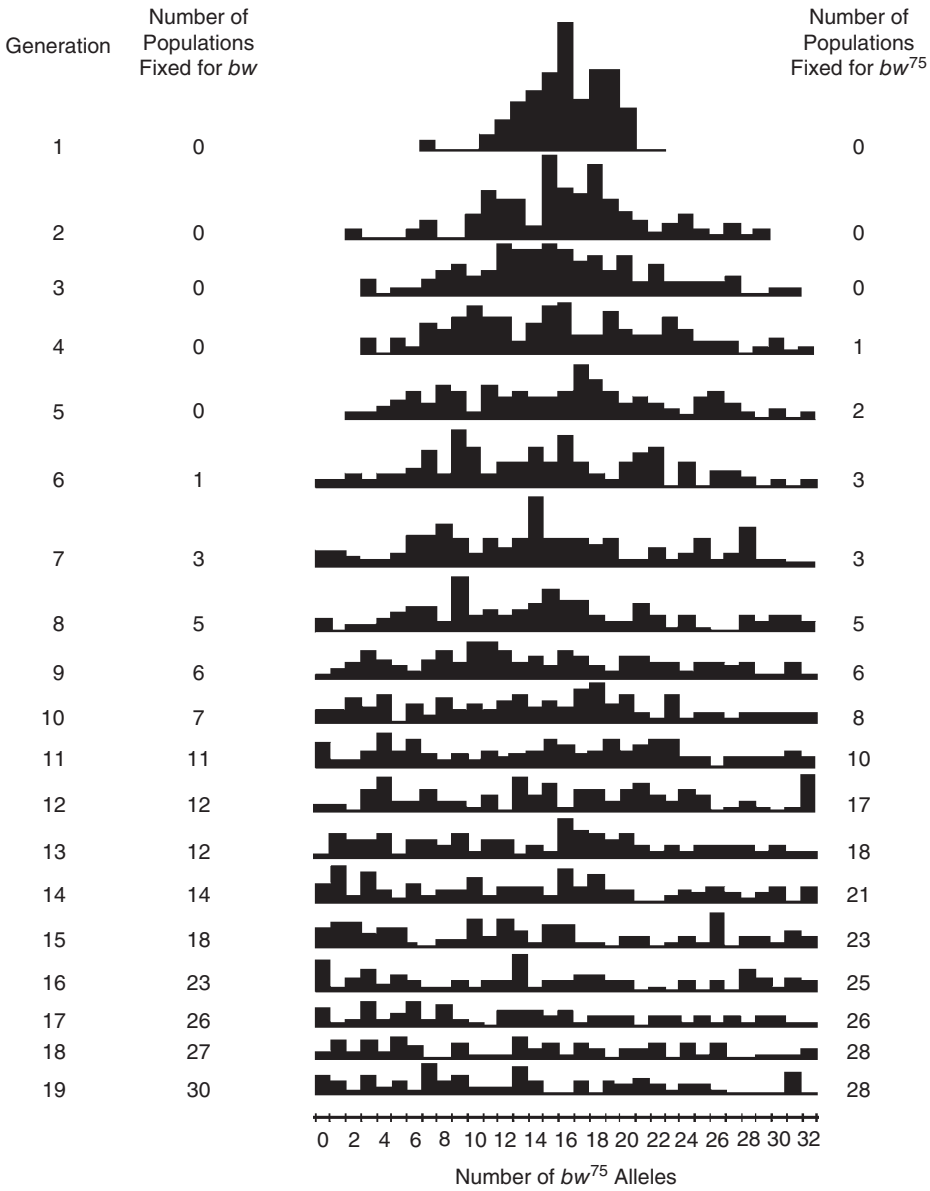

Figure 4.4 Allele frequency distributions in 107 replicate populations of *Drosophila melanogaster*, each of size 16 and with discrete generations. *Source:* Buri (1956). © 1956, The Society for the Study of Evolution.

evolutionary fate of these replicate populations for 19 generations. Note the following from his experimental results shown in Figure 4.4:

- When allele frequencies are averaged over all 107 populations, there is almost no change from the initial allele frequencies of 0.5. *Drift has no direction*.
- The chances of any particular population deviating from 0.5 and the magnitude of that deviation increase with each generation. *Evolutionary change* via *drift accumulates with time*.
- With increasing time, more and more populations become fixed for one allele or the other. By generation 19, over half of the populations had lost their genetic variation at this locus. Ultimately, all populations are expected to become fixed. *Drift causes the loss of genetic variability within a population*.
- As alleles are lost by drift, it is obvious that many copies of the remaining allele have to be identical-by-descent. For example, the original gene pool had 16 *bw* alleles in it, but those populations that are fixed for the *bw* allele have 32 copies of that allele. Moreover, some of these copies of the *bw* allele found in a fixed subpopulation are descended via DNA replication from just one of the original *bw* copies found in the initial population. When two or more copies of a gene are of the same allelic state and descended via DNA replication from a single common gene in some initial reference population, the genes are said to be identical-by-descent, as noted in Chapter 3. In the subpopulations fixed for the *bw* allele in Buri's experiments (and similarly for those fixed for bw^{75}), many individuals will be homozygous for alleles that are identical-by-descent. This means that the copy of the gene the individual received from its mother is identical-by-descent to the copy it received from its father. As fixation proceeds, homozygosity from identity-by-descent tends to increase with each succeeding generation subject to drift. *Drift causes the average probability of identity-by-descent to increase within a population*.
- All populations started out with identical gene pools, but with time, the populations deviate not only from the ancestral condition but from each other as well. For example, at generation 19, 30 populations are fixed for *bw* and 28 for bw^{75}. These populations no longer share any alleles at this locus, even though they were derived from genetically identical ancestral populations. *Drift causes an increase of genetic differences between populations*.

Founder and Bottleneck Effects

As shown in the previous section, genetic drift causes its most dramatic and rapid changes in small populations. However, even a population that is large most of the time but has an occasional generation of very small size can experience pronounced evolutionary changes due to drift in the generation(s) of small size. If the population size grows rapidly after a generation of small size, the increased population size tends to decrease the force of subsequent drift, thereby freezing in the drift effects that occurred when the population was small. These features are illustrated via computer simulation in Figure 4.5. Figure 4.5a shows four replicate simulations of genetic drift in populations of size 1000, over 100 generations, with an initial allele frequency of 0.5. Figure 4.5b shows parallel simulations, but with just one difference: at generation 20, the population size was reduced to 4 individuals and then immediately restored to 1000 at generation 21. In contrasting Figure 4.5a with 4.5b, the striking difference is the radical change in allele frequency that occurs in each population during the transition from generation 20 to 21, reflecting drift during the generation of small size. However, there is relatively little subsequent change from the allele frequencies that existed at

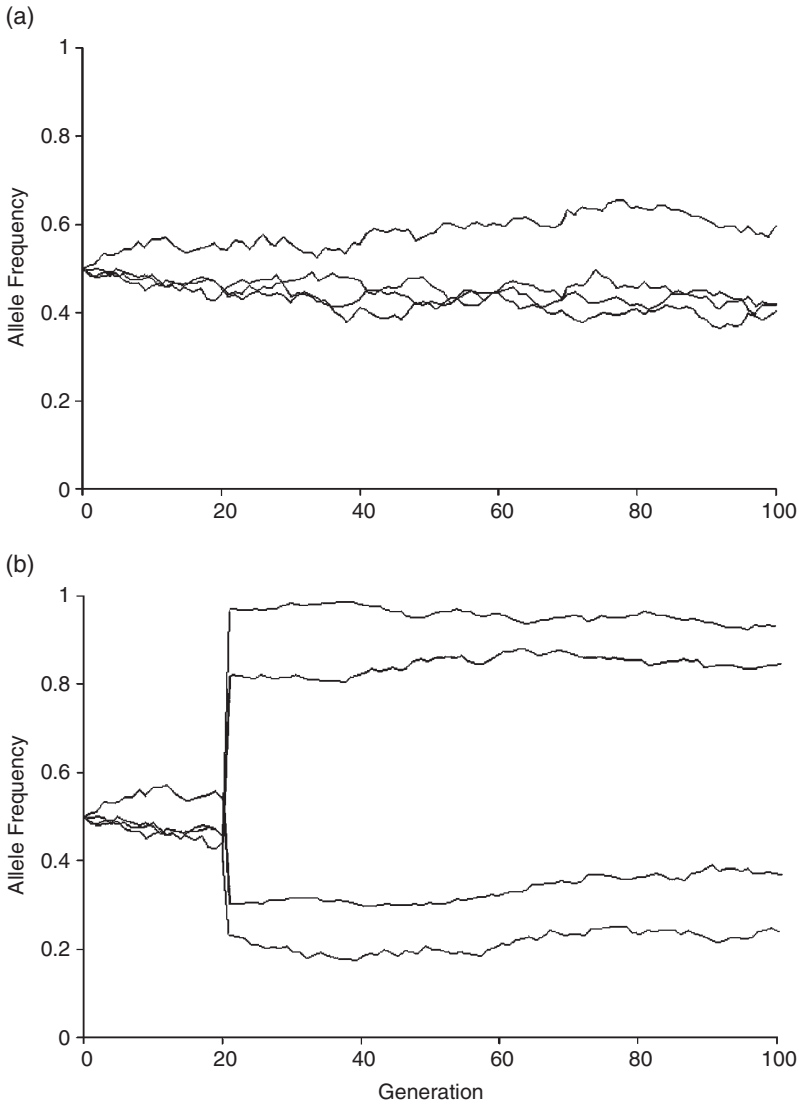

Figure 4.5 A computer simulation of genetic drift in four replicate populations starting with an initial allele frequency of 0.5 over a period of 100 generations. In panel (a), the population size is kept constant at 1000 individuals every generation. In panel (b), the same replicates are repeated until generation 20, at which point the population size is reduced to 4 individuals. The population size then rebounds to 1000 individuals at generation 21 and remains at 1000 for the remainder of the simulation to simulate a bottleneck effect.

generation 21. Thus, the pronounced evolutionary changes induced by the single generation of small population size are "frozen in" by subsequent population growth and have a profound and continuing impact on the gene pool long after the population has grown large. These computer simulations show that genetic drift can cause major evolutionary change in a population that normally has a large population size as long as either:

- the population was derived from a small number of founding individuals drawn from a large ancestral population (**founder effect**), or

- the population went through one or more generations of small size followed by subsequent population growth (**bottleneck effect**).

We will now consider some examples of founder and bottleneck effects.

There are many biological contexts in which a founder event can arise. For example, there is much evidence that individuals of Hawaiian *Drosophila* (fruit flies) are on rare occasions blown to a new island on which the species was previously absent (Carson and Templeton 1984). Because this is such a rare event, it would usually involve only a single female. Most *Drosophila* females typically have had multiple matings and can store sperm for long periods of time. A single female being blown from one island to another would often therefore carry over the genetic material from two or three males. Hence, a founder size of four or less is realistic in such cases. (Single males could also be blown to a new island, but no population could be established in such circumstances.) If the inseminated female found herself on an island for which the ecological niche to which she was adapted was unoccupied, the population size could easily rebound by one or two orders of magnitude in a single generation, resulting in a situation not unlike that shown in Figure 4.5b.

Founder events are also common in humans. One example of both a founder effect and a bottleneck effect is given by Roberts (1967, 1968). Tristan da Cunha is an isolated island in the Atlantic Ocean. With the exile of Napoleon on the remote island of St. Helena, the British decided to establish a military garrison in 1816 on the neighboring though still distant island of Tristan da Cunha. In 1817, the British Admiralty decided that Tristan da Cunha was of no importance to Napoleon's security, so the garrison was withdrawn. A Scots corporal, William Glass, asked and received permission to remain on the island with his wife, infant son, and newborn daughter. A few others decided to remain and were joined later by additional men and women, some by choice and some due to shipwrecks. Altogether, there were 20 initial founders. The population size grew to 270 by 1961, mostly due to reproduction but with a few additional immigrants. The growth of this population from 1816 to 1960 is shown in Figure 4.6.

Because there is complete pedigree information over the entire colony history, the gene pool can be reconstructed at any time as the percentage of genes in the total population derived from a particular founding individual (Figure 4.7). This method of portraying the gene pool can be related to our standard method of characterizing the gene pool through allele frequencies by regarding each founder as homozygous for a unique allele at a hypothetical locus. Then, the proportion of the genes derived from a particular founder represents the allele frequency at the hypothetical locus of that founder's unique allele in the total gene pool.

The top histogram in Figure 4.7 shows the gene pool composition in 1855 and 1857. Note from the population size graph in Figure 4.6 that a large drop in population size occurred between those years. This was caused by the death in 1853 of William Glass, the original founder. Following his death, 25 of his descendants left for America in 1856. This bottleneck was also accentuated by the arrival of a missionary minister in 1851. This minister soon disliked the island, preaching that its only fit inhabitants were "the wild birds of the ocean." Under his influence, 45 other islanders left with him, thereby reducing the population size from 103 at the end of 1855 to 33 in March 1857. Note that in going from 1855 to 1857, the gene pool composition changes substantially; the relative contributions of some individuals show sharp decreases (founders 1 and 2) whereas others show sharp increases (founders 3, 4, 9, 10, 11, and 17). Moreover, the genetic contributions of many individuals are completely lost during this bottleneck (founders 6, 7, 12, 13, 14, 15, 16, 19, and 20). Thus, the gene pool is quite different and less diverse after the first bottleneck.

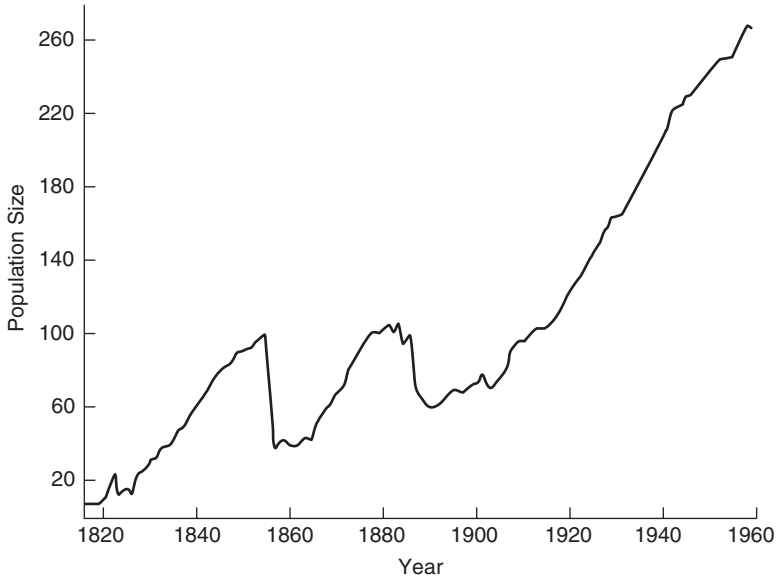

Figure 4.6 Population size of Tristan da Cunha on December 31 of each year from 1816 to 1960. *Source:* Roberts (1968). © 1968, Springer Nature.

Figure 4.6 reveals that the population grew steadily between 1857 and 1884. With the exception of a few new immigrant individuals (founders 21–26), the basic shape of the gene pool histograms changes very little in those 27 years (the second histogram from the top in Figure 4.7). In particular, note that there is much less change in these 27 years than in the 2 years between 1855 and 1857. Hence, the changes induced by the first bottleneck were "frozen in" by subsequent population growth.

Figure 4.6 shows that a second, less drastic bottleneck occurred between 1884 and 1891. The island has no natural harbor, so the islanders had to row out in small boats to trade with passing vessels. In 1884, a boat manned by 15 adult males sank beneath the waves with the resulting death of everyone on board, making Tristan da Cunha the "Island of Widows." Only four adult men were left on the island, two very aged, leading many of the widows and their offspring to leave the island. This reduced the population size from 106 in 1884 to 59 in 1891. The third histogram in Figure 4.7 shows the impact of this second bottleneck on the island's gene pool. As with the first bottleneck, some individual contributions went up substantially (founders 3, 4, and 22), others went down (founders 9 and 10), and many were lost altogether (founders 21, 24, 25, and 26).

After this second bottleneck, there was another phase of steady population growth (Figure 4.6). The shapes of the gene pool histograms change little from 1891 to 1961 during this phase of increased population growth (the bottom histogram in Figure 4.7, which excludes the impact of a few additional immigrants). Once again, this shows how subsequent population growth freezes in the changes induced by drift during the bottleneck.

As discussed in Chapter 3, the founder and bottleneck effects on Tristan da Cunha also led to pedigree inbreeding, despite a system of mating of avoidance of inbreeding (see Figure 3.5). This is yet another effect of genetic drift: finite population size leads to an increase in the mean inbreeding coefficient (the average probability of uniting gametes bearing alleles identical-by-descent) with time. Each bottleneck accentuates this accumulation of F because the number of founders

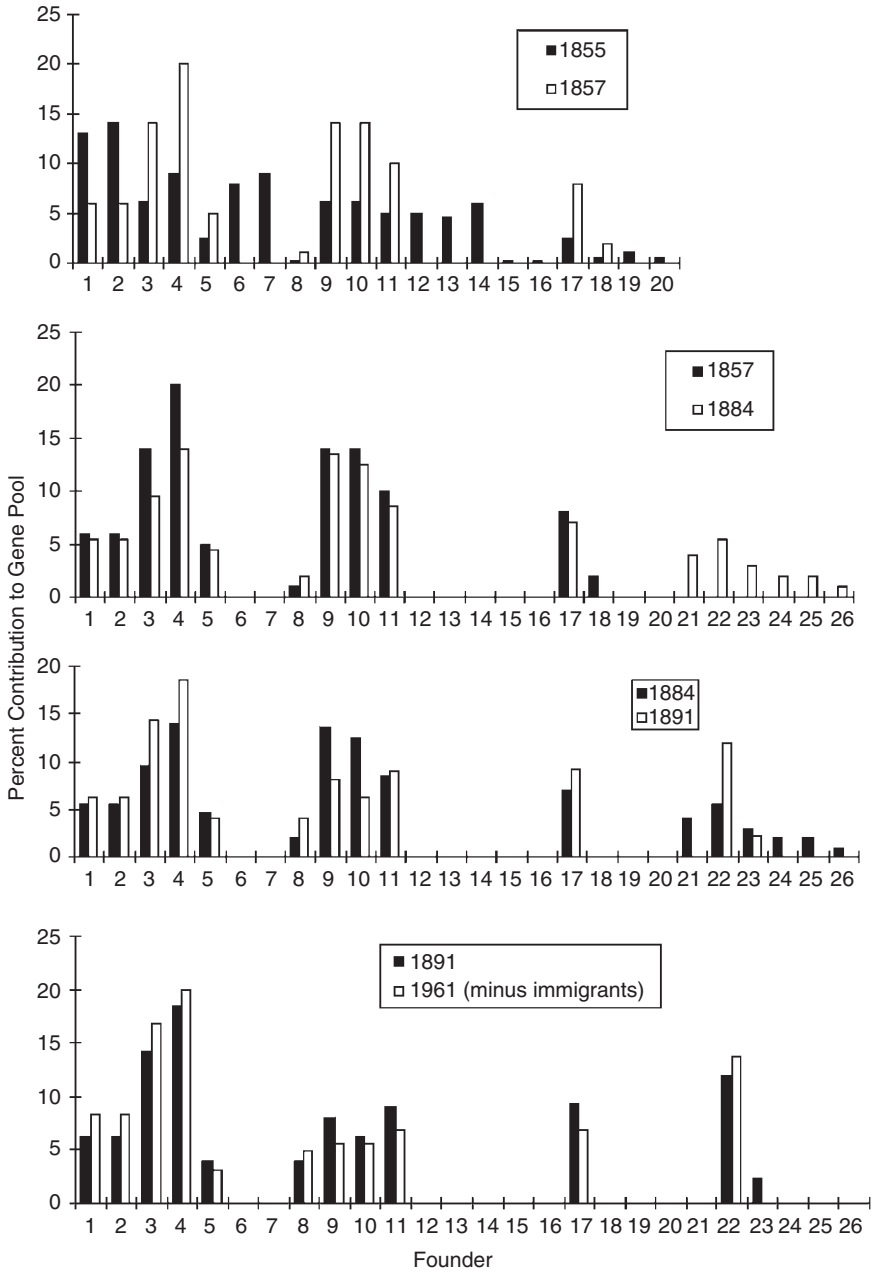

Figure 4.7 Gene pool changes over time in the Tristan da Cunha population. The gene pool is estimated from pedigree data as the proportion of the total gene pool that is derived from a particular founder (who are indicated by numbers on the x-axis). Each histogram contrasts the gene pool at two times, as indicated by the legends.

contributing to the gene pool goes down after each bottleneck event, making it more likely that the surviving individuals must share a common ancestor. Thus, founder and bottleneck effects usually *increase pedigree inbreeding*.