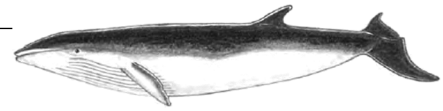
CHAPTER 12

Mutation



Minke whale, Example 12.2

> Mutation is the ultimate source of all the genetic variation necessary for evolution by natural selection; without mutation evolution would soon cease.
>
> **(Michael C. Whitlock & Sarah P. Otto 1999, p. 295)**

> Mutations can critically affect the viability of small populations by causing inbreeding depression, by maintaining potentially adaptive genetic variation in quantitative characters, and through the erosion of fitness by accumulation of mildly deleterious mutations.
>
> **(Russell Lande 1995, p. 782)**

Mutations are errors in the transmission of genetic information from parents to progeny. The process of mutation is the ultimate source of all genetic variation in natural populations. Nevertheless, this variation comes at a cost because most mutations that have phenotypic effects are harmful (deleterious). Mutations occur both at the chromosomal level and the molecular level. As we will see, mutations may or may not have a detectable effect on the phenotype of individuals.

An understanding of the process of mutation is important for conservation for several reasons. The amount of **standing genetic variation** within populations is largely a balance between the gain of genetic variation from mutations and the loss of genetic variation from genetic drift. Thus, an understanding of mutation is needed to interpret patterns of genetic variation observed in natural populations.

Moreover, the increased homozygosity of deleterious mutations is the primary source of **inbreeding depression** (Chapter 17). The frequency of deleterious mutations results from a balance between mutation and natural selection (Section 12.3). We have seen that natural selection is less effective in small populations (Section 8.6). Therefore, deleterious mutations will tend to accumulate more rapidly in small populations; this effect can further threaten the persistence of small populations (Section 18.6). On the other hand, the rate of adaptive response to environmental change is proportional to the amount of standing genetic variation for fitness within populations. Thus, long-term persistence of populations may require large population sizes in order to maintain important adaptive genetic variation.

Unfortunately, there is little empirical data available about the process of mutation because mutations are rare. The data that are available generally come from model organisms (e.g., mice, *Drosophila*, or *Arabidopsis*) that are selected because of their short generation times and suitability for raising a large number of individuals in the laboratory. However, we must be careful in generalizing results from such model species; the very characteristics that make these organisms suitable for these experiments, including short generation lengths, may make them less suitable for generalizing to other species.

How common are mutations? On a per-locus or per-nucleotide level they are rare. For example, the rate of mutation for a single nucleotide in plants and animals ranges from 10^{-7} to 10^{-10} per year, and 10^{-7} to 10^{-9} per generation (Lynch et al. 2016). This

Conservation and the Genomics of Populations, Third Edition. Fred W. Allendorf, *et al*., Oxford University Press.

means that only one in a billion or so gametes, on average, will have a mutation at a specific base pair. Mutation rates vary considerably among species (Figure 12.1; Hanlon et al. 2019). Species with longer generation lengths on average have lower mutation rates per year, but also have higher rates per generation. However, from a genome-wide perspective, mutations are actually very common. The genome of most species consists of billions of base pairs. Therefore, it has been estimated (Lynch et al. 1999) that each individual may possess hundreds of new mutations! Fortunately, almost all of these mutations are in nonessential regions of the genome and have no phenotypic effect.

In this chapter, we consider the processes resulting in mutations and examine the expected relationships between mutation rates and the amount of genetic variation within populations. We examine evidence for both harmful and advantageous mutations in populations. Finally, we examine the effects of mutation rates on the rate of recovery of genetic variation following a population bottleneck.

12.1 Process of mutation

Chromosomes and DNA sequences are normally copied exactly during the process of replication and are transmitted to progeny. However, sometimes errors occur that produce new chromosomes or new DNA sequences. Empirical information on the rates of mutation is hard to come by because mutations are so rare. It used to be necessary to study thousands of progeny to detect mutational events. Thus, estimating the rates of mutations or describing the types of changes brought about by mutation was generally very difficult. Most of our direct information about the process of mutation until recently came from model organisms (Example 12.1). However, next-generation sequencing has allowed estimation of mutation rates in pedigrees

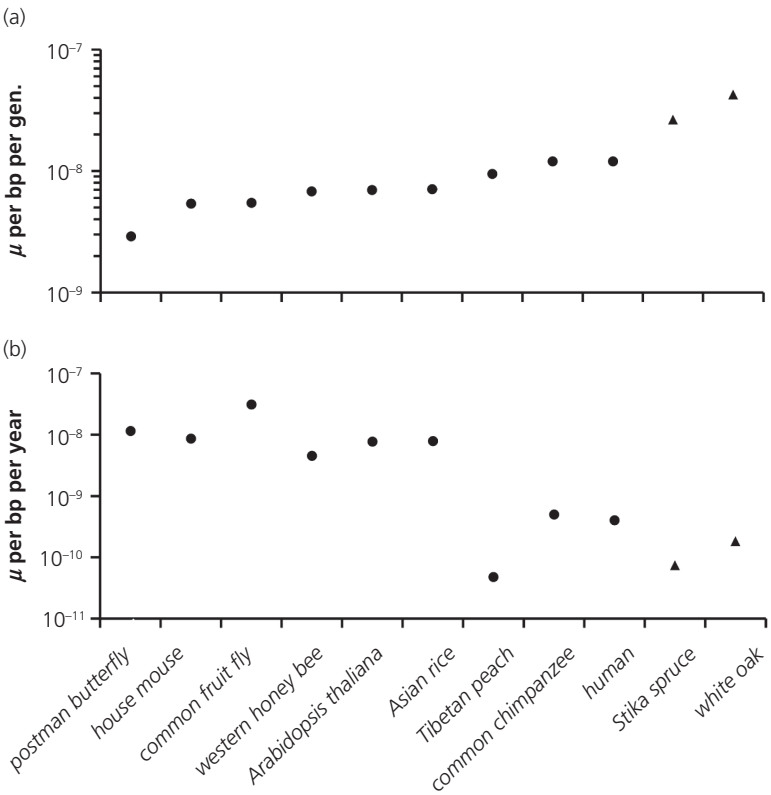


Figure 12.1 Mutation rates (a) per generation and (b) per year for multicellular species. Per-year rates were calculated using a reported age, an estimate of generation time, or an average of two estimates of generation time. For the trees, the somatic mutation rate was estimated in exceptionally long-lived specimens (triangles). Modified from Hanlon et al. (2019).

Example 12.1 Coat color mutation rate in mice

Schlager & Dickie (1971) presented the results of a direct and massive experiment to estimate the rate of mutation to five recessive coat color alleles in mice: albino, brown, nonagouti, dilute, and leaden. They examined more than 7 million mice in 28 inbred strains. Overall, they detected 25 mutations in over 2 million gene transmissions for an average mutation rate of 1.1×10^{-5} per gene transmission. As expected, the **reverse mutation rate** (from the recessive to the dominant allele) was much lower, $\sim 2.5 \times 10^{-6}$ per gene transmission. The reverse mutation rate is expected to be lower because there are more ways to eliminate the function of a gene than to reverse a defect. This assumes that the recessive coat color mutations are caused by mutations that result in a loss of function.

from nonmodel species such as wolves and flycatchers, both of which reported mutation rates of 4.6×10^{-9} mutations per site per generation (Smeds et al. 2016; Koch et al. 2019).

Mutations can occur in germline cells, and these gametic mutations occur primarily during meiosis. In animals, only gametic mutations result in genetic variation passed on to progeny. However, in plants, mutations during mitosis in somatic cells can result in variation that can be passed on because reproductive structures are produced by lineages of vegetative cells; for example, flowers produced at the top of the plant or ends of branches. In long-lived or large organisms such as trees, somatic mutations are a substantive source of variation. Ally et al. (2010) found evidence that somatic mutations in long-lived aspen clones reduce fertility and may eventually lead to senescence.

Plomion et al. (2018) compared whole genome sequences of three different branch tips from a single mature white oak tree, and found 46 somatic mutations. They were also able to demonstrate the transmission of nine of these mutations to progeny by sequencing acorns. Hanlon et al. (2019) used targeted sequence capture to sequence about 10 megabases of the exome from the bottom and the top of 20 exceptionally tall, old-growth Sitka spruce trees (see back cover). While they found few mutations, the resulting estimated rate of somatic mutations (ignoring mutations during meiosis) translated into one of the highest mutation rates found for a eukaryote (Figure 12.1). One old-growth Sitka spruce could have on the order of 100,000 different somatic mutations scattered throughout its crown, although only a very small subset could be passed on through pollen or seed. Despite the high number of somatic mutations detected, the rate per year and per meter of height growth is lower than expected given their size, suggesting that the stem cells in tree apical meristems may have evolved mechanisms to reduce mutations, or do

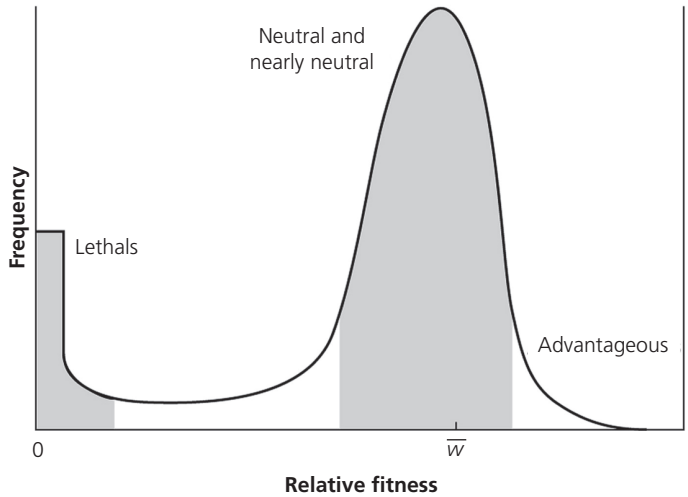


Figure 12.2 Hypothetical frequency of the fitness of new mutations relative to the mean fitness of the population ($\bar{w}$). Redrawn from Hedrick (2011).

not themselves divide as frequently as previously thought (Orr et al. 2020).

Most mutations with phenotypic effects tend to reduce fitness (Figure 12.2). Thus, as we will see in Chapter 18, the accumulation of mutations can decrease the probability of survival of small populations. Nevertheless, rare beneficial mutations are important for adaptive evolutionary change (Elena et al. 1996). In addition, work with the plant *Arabidopsis thaliana* has suggested that nearly half of new spontaneous mutations increase fitness (Shaw et al. 2002); however, this result has been questioned in view of data from other experiments (Bataillon 2003; Keightley & Lynch 2003). Work in *Escherichia coli* shows that intermediate amounts of mutation were beneficial for adaptation to novel environments, but this effect was not present at high mutation rates (Sprouffske et al. 2018).

Mutations occur randomly, but there is evidence that some aspects of the process of mutation may be an adaptive response to environmental conditions. There has been an ongoing controversy that mutations in prokaryotes may be directed toward particular environmental conditions (Lenski & Sniegowski 1995). There is some evidence that the rate of mutations in eukaryotes may increase under stressful conditions and thus create new genetic variability that may be important in adaptation to changing environmental conditions (Capy et al. 2000). However, the relationship between stress and mutation is complex, and stress does not necessarily increase mutation rate (Ferenci 2019). In addition, there is little evidence for an influence of the environment on the effects of new mutations (Halligan & Keightley 2009).

12.1.1 Chromosomal mutations

We saw in Chapter 3 that rates of chromosomal evolution vary tremendously among different taxonomic groups. There are two primary factors that may be responsible for differences in the rate of chromosomal change: (1) the rate of chromosomal mutation, and (2) the rate of incorporation of such mutations into populations (Rieseberg 2001). Differences among taxa in rates of chromosomal change may result from differences in either of these two effects.

White (1978) estimated a general mutation rate for chromosomal rearrangements of the order of one per 1,000 gametes in a wide variety of species from lilies to grasshoppers to humans. Lande (1979) considered different forms of chromosomal rearrangements in animals and produced a range of estimates between 10^{-4} and 10^{-3} per gamete per generation. There is evidence in some groups that chromosomal mutation rates may be substantially higher than this. For example, Porter & Sites (1987) detected spontaneous chromosomal mutations in five of 31 male lizards that were examined.

The apparent tremendous variation in chromosomal mutation rates suggests that some of the differences among taxa could result from differences in mutation rates. In addition, there is some evidence that chromosomal polymorphisms may contribute to increased chromosomal mutation rates. That is, chromosomal mutation rates may be greater in chromosomal heterozygotes than homozygotes (King 1993). We will see in Section 12.1.4 that genomes with more **transposable elements** may also have higher chromosomal mutation rates.

12.1.2 Molecular mutations

There are several types of **molecular mutations** in DNA sequences: (1) substitutions, the replacement of one nucleotide with another; (2) recombinations, the exchange of a sequence from one homologous chromosome to the other; (3) deletions, the loss of one or more nucleotides; (4) insertions, the addition of one or more nucleotides; and (5) inversions, the rotation by 180° of a double-stranded DNA segment of two or more base pairs (Graur & Li 2000).

Mutation rates are sometimes estimated indirectly by an examination of rates of substitutions over evolutionary time in regions of the genome that are not affected by natural selection. The expected rate of substitution per generation will be equal to the mutation rate for selectively neutral mutations (Kimura 1983). Using this approach, the average rate of mutation in mammalian nuclear DNA has

been estimated to be 3–5 $\times$ 10^{-9} nucleotide substitutions per base pair per year (Graur & Li 2000), which is not far off estimates from short-term experiments (Figure 12.1). However, the substitution rate varies enormously among different regions of the nuclear genome.

The rate of mutation also differs between genomes. In mammalian mtDNA, mutation rate has been estimated to be at least 10 times higher than the average nuclear rate (Brown et al. 1982), while the rate of mutation in the plant mitochondrial genome is estimated to be 40–100 times slower in plants than in animals (Palmer 1992). In angiosperm plants, the mutation rate of the chloroplast genome is estimated to be five times slower than the nuclear rate and the mitochondrial rate is 16 times slower than that in the nuclear genome (Drouin et al 2008). The chloroplast genome has a slower mutation rate than the mitochondrial genome in protists (Smith 2015). However, these are general overall rates and different regions have different rates. For example, genes and spacer regions in the chloroplast genome show quite variable rates of variation making them useful for estimation of relationships at different taxonomic levels (Shaw et al. 2014).

Whole genome sequencing has made estimating mutation rate much more straightforward. Rates can be estimated through mutation accumulation experiments with model organisms (Lynch et al. 2016). In these experiments, isogenic (genetically identical) individuals are used to start multiple lineages over multiple generations with extreme bottlenecks produced by either clonal reproduction of one individual or mating of two individuals within lines. Mutations accumulate in each lineage. At the end of the experiment, individuals from each lineage are whole genome sequenced, and mutations identified. From this, the mutation rate can be calculated. **Single nucleotide polymorphisms (SNPs)** are generally bi-allelic and follow the infinite-sites mutation model (Morin et al. 2004). The per base pair per generation rate of all organisms is $<10^{-7}$, but it is as low as 10^{-9} or 10^{-10} in some unicellular eukaryotes and bacteria. Lynch et al. (2016) hypothesize that lower mutation rates have been able to evolve in these unicellular organisms due to their very large effective population sizes.

Table 12.1 Mutations at the *OGO1c* microsatellite locus in pink salmon (Steinberg et al. 2002). Approximately 1,300 parent–progeny transmissions were observed in 50 experimental matings. Mutations were found only in the four matings shown. The mutant allele is indicated by bold-face type and the most likely progenitor of the mutant allele is underlined. All of the putative mutations differ by one repeat unit from their most likely progenitor. The overall mutation rate estimated from these data is 3.9 $\times$ 10^{-3} (5/1,300).

Dam	Sire	Progeny genotypes				
a/b	*c/d*	*a/c*	*a/d*	*b/c*	*b/d*	Mutant genotypes
342/350	408/474	1	1	3	3	342/**478**
295/366	303/362	1	2	4	2	295/**366**
269/420	346/450	8	16	10	8	420/**446** (2)
348/348	309/448	5	4	0	0	348/**444**

The rate of mutation at microsatellite loci is much greater than other regions of the genome because of the presence of simple sequence repeats (Li et al. 2002). Two mechanisms are thought to be responsible for mutations at microsatellite loci: (1) mispairing of DNA strands during replication, and (2) recombination. Estimates of mutation rates at microsatellite loci have generally been approximately 10^{-3} mutants per generation (Ellegren 2000a; Table 12.1). Microsatellite mutations appear largely to follow the stepwise mutation model (SMM) where single repeat units are added or deleted with near equal frequency (Valdes et al. 1993; Figure 12.3). However, the actual mechanisms of microsatellite mutation are more complicated than this simple model (Estoup & Angers 1998; Li et al. 2002; Anmarkrud et al. 2008).

The mutation rate for protein coding loci (e.g., allozymes) is relatively low. In addition, not all DNA mutations will result in a change in the amino acid sequence because of the inherent redundancy of the genetic code. Nei (1987, p. 30) reviewed the literature on direct and indirect estimates of mutation rates for allozyme loci. Most direct estimates of mutation rates in allozymes have failed to detect any mutant alleles; for example, Kahler et al. (1984) examined a total of 841,260 gene transmissions from parents to progeny at five loci and failed to detect any mutant alleles. General estimates of mutation rates for allozyme loci are on the order of 10^{-6} to 10^{-7} mutants per gene transmission (Nei 1987).

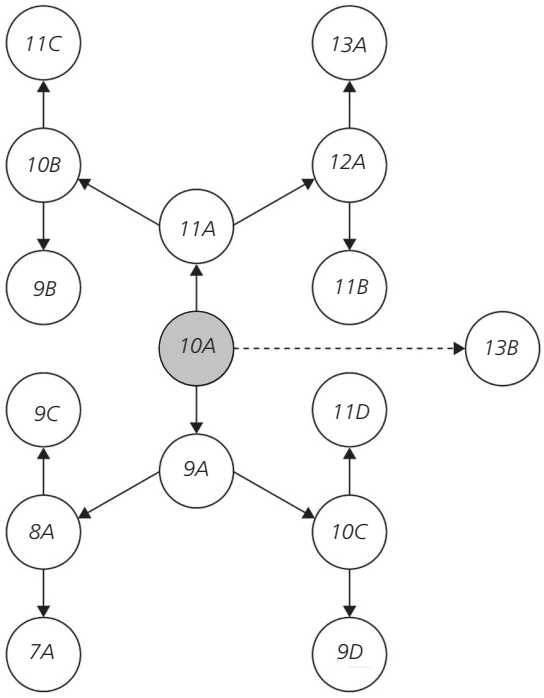


Figure 12.3 Pattern of mutation for microsatellites beginning with a single ancestral allele (shaded circle in the middle) with 10 repeats. Most mutations are a gain or loss of a single repeat (stepwise mutation model, SMM). The dashed arrow shows a multiple-step mutation from 10 to 13 repeats. Alleles are designated by the number of repeats and a letter that distinguishes homoplastic alleles, which are alike in state (number of repeats) but differ in origin.

New mutations sometimes occur in clusters and have been associated with hotspots for crossing over during meiosis (Chan & Gordenin 2015). Woodruff & Thompson (1992) found as many as 20% of new mutations in *Drosophila* represented clusters of identical mutant alleles sharing a common pre-meiotic origin. Cluster mutations at microsatellite loci have been found in several other species (Jones et al. 1999; Steinberg et al. 2002). The occurrence of clustered mutations results in nonuniform distributions of novel alleles in a population that could influence interpretations of mutation rates and patterns as well as estimates of genetic population structure. For example, Woodruff et al. (1996) have shown that mutant alleles that are part of clusters are more likely to persist and be fixed in a population than mutant alleles entering the population independently.

12.1.3 Quantitative characters

As we saw in Chapter 11, the amount of genetic variation in quantitative characters for morphology, physiology, and behavior that can respond to natural selection is measured by the additive genetic variance (V_A). The rate of loss of additive genetic variance due to genetic drift in the absence of selection is the same for the loss of heterozygosity (i.e., $1/2N_e$). The effective mutation rate for quantitative traits is much higher than the rate for single gene traits because a mutation at any one of the many loci underlying a polygenic trait can affect a quantitative trait. The input of additive genetic variance per generation by mutation is V_m. The expected genetic variance at equilibrium between these two factors is $V_A = 2N_eV_m$ (Lande 1995, 1996).

Estimates of mutation rates for quantitative characters are rare and imprecise. It is thought that V_m is roughly on the order of 10^{-3} V_A (Lande 1995). However, some experiments suggest that the great majority of these mutations are highly detrimental and therefore are not likely to contribute to the amount of standing genetic variation within a population. Thus, the effective V_m responsible for much of the standing variation in quantitative traits in natural populations may be an order of magnitude lower, 10^{-4} V_A (Lande 1996; Barton & Keightley 2002; Mackay 2010).

12.1.4 Transposable elements, stress, and mutation rates

Much of the genome of eukaryotes consists of sequences associated with transposable elements that possess an intrinsic capability to make multiple copies and insert themselves throughout the genome (Graur & Li 2000). Transposable elements were first discovered by Nobel Laureate Barbara McClintock, who studied these "jumping genes" in maize (McClintock 1950; Ravindran 2012). Approximately half of the human genome consists of DNA sequences associated with transposable elements (Lynch 2001). This activity is analogous to the "cut and paste" mechanism of a word processor. Transposable elements are potent agents of **mutagenesis** (Kidwell 2002). For example, Clegg & Durbin (2000) found that nine of 10 mutations affecting flower

color in the morning glory were the result of the insertion of transposable elements into genes. A consideration of the molecular basis of transposable elements is beyond our consideration (see Chapter 7 of Graur & Li 2000). Nevertheless, the mutagenic activity of these elements is of potential significance for conservation.

Transposable elements can cause a wide variety of mutations. They can induce chromosomal rearrangements such as deletions, duplications, inversions, and reciprocal translocations. Kidwell (2002, p. 2219) has suggested that "transposable elements are undoubtedly responsible for a significant proportion of the observed karyotypic variation among many groups." In addition, transposable elements are responsible for a wide variety of substitutions in DNA sequences, ranging from insertion of the transposable element sequence to substitutions, deletions, and insertions of a single nucleotide (Kidwell 2002).

Stress has been defined as any environmental change that drastically reduces the fitness of an organism (Hoffmann & Parson 1997). McClintock (1984) first suggested that transposable element activity could be induced by stress. A number of transposable elements in plants have been shown to be activated by stress (Grandbastien 1998; Capy et al. 2000). Some transposable elements in *Drosophila* have been shown to be activated by heat stress, but other studies have not found an effect of heat shock (Capy et al. 2000). In addition, hybridization has also been found to activate transposable elements and cause mutations (Kidwell & Lisch 1998).

12.2 Selectively neutral mutations

Many mutations in DNA sequence have no phenotypic effect so that they are neutral with regard to natural selection (e.g., mutations in noncoding regions). In this case, the amount of genetic variation within a population will be a balance between the gain of variation by mutation and the loss by genetic drift. The distribution of neutral genetic variation among populations is primarily a balance between these two forces. Gene flow among subpopulations retards the process of differentiation until eventually a steady state may be reached between the opposing effects of gene flow and genetic drift. However, the process of mutation may also contribute to allele frequency divergence among populations in cases where the mutation rate approaches the gene flow rate.

12.2.1 Genetic variation within populations

The amount of genetic variation within a population at equilibrium will be a balance between the gain of variation as a function of the neutral mutation rate (μ) and the loss of genetic variation by genetic drift as a function of effective population size (N_e).

We will first consider the so-called infinite allele model (IAM) in which we assume that every mutation creates a new allele that has never been present in the population. This model is appropriate if we consider variation in DNA sequences. Even for a relatively short DNA sequence, there are a very large number of possible allelic states, as each nucleotide site can be occupied by one of four bases (A, T, C, or G). For example, there are over 1 million possible alleles if we just consider 10 base pairs (bp) (4^{10} = 1,048,576).

The average expected heterozygosity (H) at a locus (or over many loci with the same mutation rate) is:

$$H = \frac{4N_e\mu}{(4N_e\mu + 1)} = \frac{\theta}{\theta + 1} \qquad (12.1)$$

where μ is the neutral mutation rate and $\theta = 4N_e\mu$ (Kimura & Crow 1964). The multiple parameter $4N_e\mu$ is an important expression in population genetics theory and is called the **population-scaled mutation rate** (θ). This relationship can be used to estimate effective population size if we know the mutation rate (Example 12.2).

The much greater variation at microsatellite loci compared with SNPs and allozymes results from the differences in mutation rates that we discussed in Section 12.1.2. Figure 12.4 shows the equilibrium heterozygosity for microsatellites, allozymes, and 100 bp of DNA using mutation rates of 10^{-4} and 10^{-6}, respectively, and Equation 12.1. Thus, we expect a heterozygosity of 0.038 at allozyme loci (or for a 100-bp sequence) and 0.80 at microsatellite

Example 12.2 How many whales are there in the ocean?

Equation 12.1 can also be used to estimate the effective population size of natural populations if we know the mutation rate (μ). For example, Roman & Palumbi (2003) estimated the historical (pre-whaling) number of humpback, fin, and minke whales in the North Atlantic Ocean by estimating θ for the control region of **mitochondrial DNA (mtDNA)**. In the case of mtDNA, $\theta = 2N_{e(f)}\mu$ because of maternal inheritance and haploidy. Roman & Palumbi (2003) used a range of mutation rates based on observed rates of divergence between mtDNA of different whale species.

Their genetic estimates of historical population sizes for humpback, fin, and minke whales were far greater than those previously calculated, and are 6–20 times higher than the current population estimates for these species. This discrepancy is crucial for conservation because the International Whaling Commission management plan uses the estimated historical population sizes as guidelines for setting allowable harvest rates. We should be careful using estimates of N_e with this approach because there are a host of pitfalls (e.g., how reliable are our estimates of mutation rate?). Roman & Palumbi (2003) provide a useful discussion of the limitations of this method for estimating N_e. Palsbøll et al. (2013) have provided an insightful critique of using this method to estimate historical population sizes.

loci with an effective population size of 10,000. However, we also expect a substantial amount of variation in heterozygosity among loci, especially loci with lower mutation rates (Figure 12.5).

The heterozygosity values for microsatellite loci in Figures 12.4 and 12.5 are likely to be overestimates because of several important assumptions in this expectation. Microsatellite mutations tend to occur in steps of the number of repeat units. Therefore, each mutation will not be unique, but rather will be to an allelic state (say 11 copies of a repeat) that already occurs in the population. This is called homoplasy in which two alleles that are identical in state have different origins (e.g., alleles *11C* and *11D* in Figure 12.3). Therefore, the actual expected heterozygosity is less than predicted by Equation 12.1. Allozymes also tend to follow a stepwise model of mutation (Ohta & Kimura 1973), but homoplasy will be less common because of the smaller number of alleles present in a population due to the lower mutation rate.

The type of point mutation that occurs when a single nucleotide is replaced by another depends on whether a purine (A or G) is replaced by a

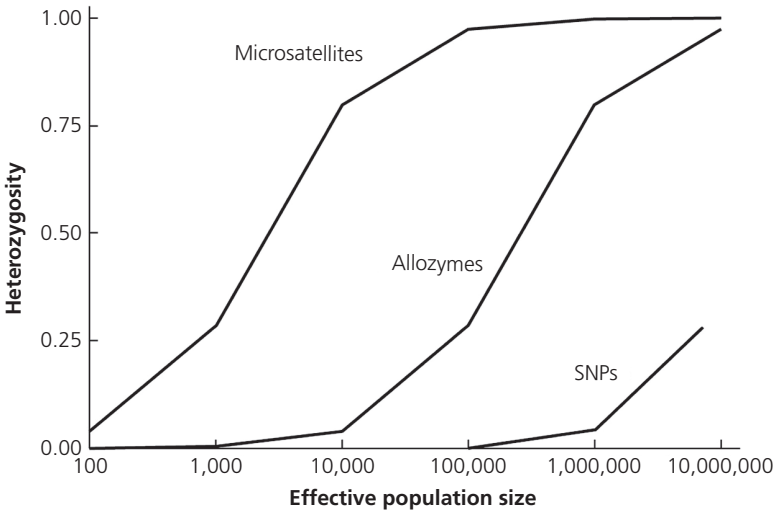


Figure 12.4 Expected heterozygosity in populations of different size using Equation 12.1 for microsatellites ($\mu = 10^{-4}$), allozymes ($\mu = 10^{-6}$), and SNPS ($\mu = 10^{-8}$).

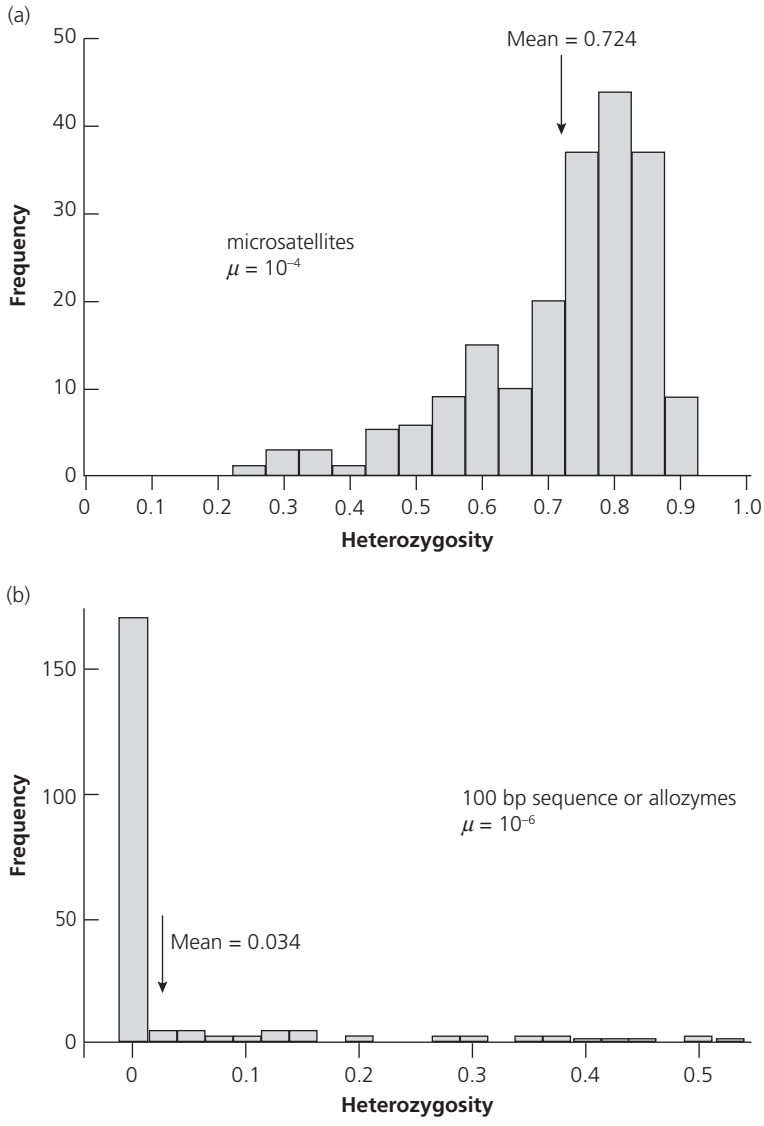


Figure 12.5 Simulated heterozygosities at 200 loci in a population with N_e = 10,000 and the IAM of mutation produced with the program *EASYPOP* (Balloux 2001). (a) Microsatellite loci with $\mu = 10^{-4}$; (b) Allozyme loci with $\mu = 10^{-6}$. A 100-bp DNA sequence would be expected to have a similar distribution to allozymes. The expected heterozygosities are (a) 0.800 and (b) 0.038 (Equation 12.1).

purine or by a pyrimidine (G or C), and vice versa. Mutations replacing a purine with a purine or a pyrimidine with a pyrimidine are called transition mutations. Mutations replacing a purine with a pyrimidine, or vice versa, are called transversion mutations. Although there are twice as many possible transversions as transitions, transitions are more common because of the similarity of the chemical structures substituted.

It is also important to remember that the mutation rates used here (μ) are the neutral mutation rates. Mutations in DNA sequence within some regions of the genome are likely not to be selectively neutral. Therefore, different regions of the genome will have different effective neutral mutation rates, even though the actual rate of molecular mutations may be the same. For example, mutations in protein coding regions may affect the amino

acid sequence of an essential protein and thereby reduce fitness. Transversion mutations within coding regions are more likely to result in amino acid substitutions than transitions. Such mutations will not be neutral and therefore will not contribute to the amount of variation maintained by our model of drift–mutation equilibrium considered here. In these regions, so-called **purging selection** stops these mutations reaching high frequencies in a population. In contrast, mutations in DNA sequence in regions of the genome that are not functional are much more likely to be neutral. This expectation is supported by empirical results; exons, which are the coding regions of protein loci, are much less variable than the introns, which do not encode amino acids (Graur & Li 2000).

12.2.2 Population subdivision

The process of mutation may also contribute to allele frequency divergence among populations (Ryman & Leimar 2008). The relative importance of mutation on divergence (e.g., F_{ST}) depends primarily upon the relative magnitude of the rates of migration and of mutation. Under the IAM of mutation with the island model of migration (Crow & Aoki 1984), the expected value of F_{ST} at equilibrium is approximately:

$$F_{ST} = \frac{1}{(1 + 4Nm + 4N\mu)} \qquad (12.2)$$

However, greater mutation rates will increase F_{ST} when new mutations are not dispersed at sufficient rates to attain equilibrium between genetic drift and gene flow. Under these conditions, new mutations may drift to substantial frequencies in the population in which they occur before they are distributed among other populations via gene flow (Neigel & Avise 1993). Divergence for markers like SNPs and allozymes with relatively low mutation rates is unlikely to be affected by differences in mutation rates unless the subpopulations are completely isolated.

In general, mutations will have an important effect on population divergence only when the migration rates are very low (say 10^{-3} or less) and the mutation rates are unusually high, as for microsatellite markers (10^{-3} or greater; Nichols & Freeman 2004; Epperson 2005; Wang 2015). However, as we saw in Section 9.8, F_{ST} will underestimate genetic divergence at loci with very high within-deme heterozygosities (H_S; Hedrick 1999). Large differences in H_S caused by differences in mutation rates among loci (e.g., Steinberg et al. 2002) can result in discordant estimates of F_{ST} and G_{ST} among microsatellite loci. This may result in an underestimation of both the degree of genetic divergence among populations if all loci are pooled for analysis and the estimation of F_{ST} (Olsen et al. 2004b). See Section 9.8.2 and Wang (2015) for a solution to this problem.

We saw in the previous section that long-term N_e can be estimated using the amount of heterozygosity in a population if we know the mutation rate. However, we also know from Chapter 9 that the amount of gene flow affects the amount of genetic variation in a population. Therefore, estimates of N_e using Equation 12.1 may be overestimates because they reflect the total N_e of a series of populations connected by gene flow rather than the N_e of the local population (Guest Box 7). Consider two extremes. In the first, a population on an island is completely isolated from the rest of the members of its species ($mN = 0$). In this case, estimates of N_e using Equation 12.1 will reflect the local N_e. In the other extreme, a species consists of a number of local populations that are connected by substantial gene flow (say $mN = 100$); in this case the estimates of N_e using Equation 12.1 will reflect the combined N_e of all populations (Table 12.2).

12.3 Harmful mutations

Most mutations that affect fitness have a detrimental effect (Figure 12.2). Natural selection acts to keep these mutations from increasing in frequency. Consider the joint effects of mutation and selection at a single locus with a normal allele (A_1) and a mutant allele (A_2) that reduces fitness as shown below:

A_1A_1	A_1A_2	A_2A_2
1	$1 - hs$	$1 - s$

Table 12.2 Estimates of effective population size (N_e) with computer simulations using Equation 12.1 in a series of 20 subpopulations (local N_e = 200) that are connected by different amounts of gene flow with an island model of migration (*EASYPOP*, Balloux 2001). A mutation rate of 10^{-4} was used to simulate the expected heterozygosities at 100 microsatellite loci. The simulations began with no genetic variation in the first generation and ran for 10,000 generations. F_{ST}* is the expected F_{ST} with this amount of gene flow corrected for a finite number of populations (Mills and Allendorf 1996). $\widehat{N}_e$ is the estimated effective population size based upon the mean expected local heterozygosity (H_S) using Equation 12.1.

mN	H_T	H_S	F_{ST}	F_{ST}*	$\widehat{N}_e$
0	0.814	0.076	0.907	1.000	205
0.5	0.665	0.477	0.283	0.311	2,274
1.0	0.635	0.516	0.187	0.184	2,667
2.0	0.621	0.558	0.100	0.101	3,156
5.0	0.618	0.592	0.041	0.043	3,630
10.0	0.606	0.594	0.020	0.022	3,665

where s is the reduction in fitness of the homozygous mutant genotype and h is the degree of dominance of the A_2 allele. A_2 is recessive when $h = 0$, dominant when $h = 1$, and partially dominant when h is between 0 and 1.

If the mutation is recessive ($h = 0$), then at equilibrium:

$$q^* = \sqrt{\frac{\mu}{s}} \tag{12.3}$$

When A_2 is partially dominant, q will generally be very small and the following approximation holds:

$$q^* \approx \frac{\mu}{hs} \tag{12.4}$$

See Lynch et al. (1999) for a consideration of the importance of mildly deleterious mutations in evolution and conservation.

12.4 Advantageous mutations

Genetic drift plays a major role in the survival of advantageous mutations even in extremely large populations. That is, most advantageous mutations will be lost during the first few generations because new mutations will always be rare. The initial frequency of a mutation will be one over the total number of gene copies at a locus (i.e., $q = 1/2N$). Even a greatly advantageous allele that is recessive will have the same probability of initial persistence in a population because the advantageous homozygotes will not occur until the allele happens to drift to a relatively high frequency. For example, a new mutation will have to drift to a frequency over 0.20 before even 5% of the population will be homozygotes with the selective advantage. Therefore, the great majority of advantageous mutations that are recessive will be lost.

Dominant advantageous mutations have a much greater chance of surviving the initial period because their fitness advantage will immediately be effective in heterozygotes that carry the new mutation. However, even most dominant advantageous mutations will be lost within the first few generations because of genetic drift. For example, over 80% of dominant advantageous mutations with a selective advantage of 10% will be lost within the first 20 generations (Crow & Kimura 1970, p. 423). This effect can be seen in a simple example. Consider a new mutation that arises that increases the fitness of the individual that carries it by 50%. Even if the individual that carries this mutation contributes three progeny to the new generation, there is a 0.125 probability that none of the progeny will carry the mutation because of the vagaries of Mendelian segregation ($0.5 \times 0.5 \times 0.5 = 0.125$).

Gene flow and spread of globally advantageous mutations may be an important cohesive force in evolution (Rieseberg & Burke 2001). Ehrlich & Raven (1969) argued in a classic paper that the amounts of gene flow in many species are too low to prevent substantial differentiation among subpopulations by genetic drift or local adaptation so that local populations are essentially independently evolving units in many species. We saw in Chapter 9 that even one migrant per generation among subpopulations can cause all alleles to be present in all subpopulations. However, even much lower amounts of gene flow can be sufficient to cause the spread of an advantageous allele (say $s >$ 0.05) throughout the range of a species (Rieseberg & Burke 2001). The rapid spread of such advantageous alleles may play an important role in maintaining

Box 12.1 Detection of bottlenecks with the heterozygosity excess test

There are a variety of tests available to detect past population bottlenecks (see Peery et al. 2012). For example, we saw in Example 6.2 with a graphical qualitative approach that the absence of rare alleles in a population sample indicates the effects of a recent bottleneck. The heterozygosity excess test is a quantitative test of this same effect based upon Equation 12.1. As we have seen, heterozygosity at selectively neutral loci results from an equilibrium between mutation and genetic drift (Equation 12.1). The expected heterozygosity also can be calculated from the observed number of alleles and the sample size of individuals, assuming neutrality and mutation–drift equilibrium (Cornuet & Luikart 1996). In nonbottlenecked populations that are near mutation–drift equilibrium, the expected heterozygosity equals the measured Hardy–Weinberg (HW) heterozygosity. But if a population has experienced a recent bottleneck, the mutation–drift equilibrium is transiently disrupted and the heterozygosity measured at a locus will exceed the heterozygosity computed from the number of alleles sampled (Maruyama & Fuerst 1985). That is, bottlenecks generate a "heterozygosity excess" because alleles are generally lost faster than heterozygosity during a bottleneck (Section 6.4).

For example, Ramstad et al. (2013) tested for bottlenecks using the heterozygosity excess test with 15 microsatellite loci in four human-founded island populations of little spotted kiwi, including the Long Island population in Example 6.4. Little spotted kiwi were once found throughout the North and South Islands of New Zealand, but they declined rapidly in the 1800s, and were extinct on the North Island by 1900 owing to introduced predators and an enormous trade in their skins for export to Europe. They remained common on the South Island until the early 1900s, but they then declined rapidly owing to predation from stoats, cats, and dogs, and were virtually extinct by the 1980s. They were saved from extinction by translocation of five birds from the South Island of New Zealand to Kapiti Island in 1912. The Kapiti Island population now numbers over 1,000 birds and has provided founders for several new populations. All four of these populations displayed significant heterozygosity excess ($P < 0.025$; Figure 12.6).

The difference between an excess of heterozygotes and heterozygosity excess has sometimes been confused in the literature. The former compares the number of observed heterozygotes with that expected under HW proportions given the allele frequencies of the population. The latter compares mean expected heterozygosity observed in a population relative to that expected of a nonbottlenecked population with the same number of alleles under mutation–drift equilibrium.

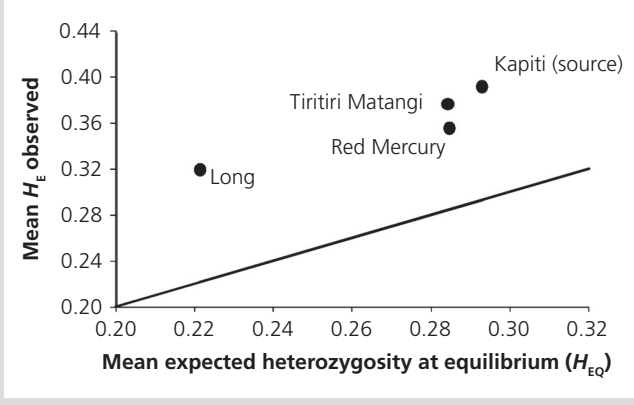


Figure 12.6 Relationship between mean expected heterozygosity observed (H_E) and expected at mutation–drift equilibrium (H_{EQ}) at 15 microsatellite loci for four island populations of little spotted kiwi. The line represents equality between H_E and H_{EQ}. All four populations displayed significant heterozygosity excess ($P < 0.025$), with the most extreme signal in the sample from Long Island. From Ramstad et al. (2013).

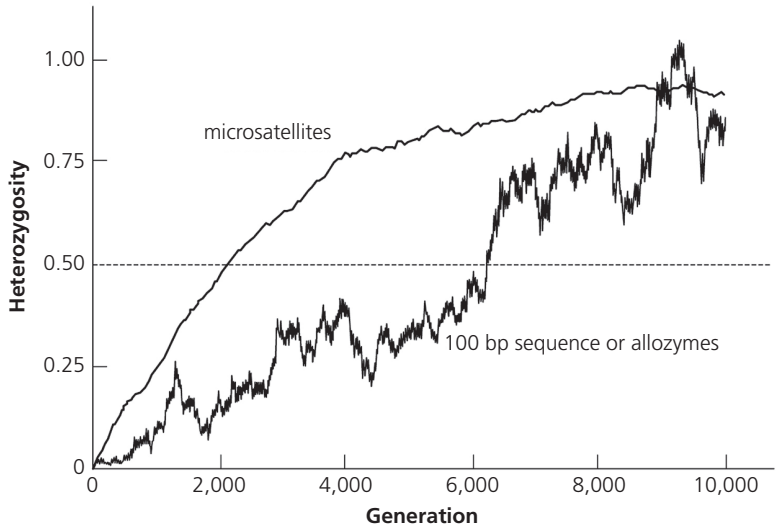


Figure 12.7 Simulated recovery of heterozygosity at 100 loci in a population of 5,000 individuals following an extreme bottleneck using *EASYPOP* (Balloux 2001). The initial heterozygosity was zero. The mutation rates are 10^{-4} for microsatellites and 10^{-6} for a 100-bp DNA sequence or for allozymes. Heterozygosity is standardized as the mean heterozygosity over all 100 loci divided by the expected equilibrium heterozygosity using Equation 12.1 (0.670 and 0.020, respectively).

the genetic integration of subpopulations connected by small amounts of genetic exchange.

12.5 Recovery from a bottleneck

We have seen that population bottlenecks will have a greater effect on allelic diversity than on heterozygosity. This effect can be used to detect the presence of previous bottlenecks (Box 12.1). The rate of recovery of genetic variation from the effects of a bottleneck will depend primarily on the mutation rate (Lynch 1996). The equilibrium amount of neutral heterozygosity in natural populations (Equation 12.1) will be approached at a time scale equal to the shorter of $2N_e$ or $1/(2\mu)$ generations (Kimura & Crow 1964).

Mutation rates affect how quickly genetic diversity will accumulate following a strong bottleneck. Microsatellites have a mutation rate of ~10^{-3} per generation, while a 100-bp DNA sequence or an allozyme will have mutation rates of ~10^{-6} per generation. Simulations of 100 typical loci show the expected heterozygosity at microsatellite loci returned to 50% of that expected at equilibrium after 2,000 generations in populations of 5,000 individuals (Figure 12.7). It took approximately three times as long at the loci with mutation rates of 10^{-6} per generation. In this case, $1/(2\mu)$ is 5,000 generations for microsatellites and 100 times that for the other markers. However, $2N_e$ is 10,000 for all types of markers.

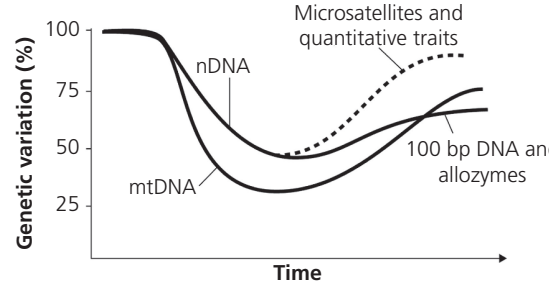


Figure 12.8 Diagram of relative expected effects of a severe population bottleneck on different types of genetic variation. The smaller N_e for mtDNA causes more genetic variation to be lost during a bottleneck. The rate of recovery following a bottleneck is largely determined by the mutation rate (Section 12.5).

As we saw in Section 12.1.3, the estimated mutation rate (V_m) for phenotypic characters affected by many loci (quantitative characters) is much higher than expected for a single gene. Therefore, we would expect quantitative genetic variance for phenotypic characters to be restored at rates comparable with those of microsatellites (Lande 1996).

Thus, recovery of microsatellite variation following a severe bottleneck may be a good measure of the recovery of polygenic variation for fitness traits.

Figure 12.8 provides a simplistic representation of the effects of a severe population bottleneck on different sources of genetic variation. Microsatellites, allozymes, and quantitative traits are all expected to lose genetic variation at approximately the same rates. However, mtDNA will lose genetic variation more rapidly because of its smaller N_e. The rates of recovery of variation will depend upon the mutation rates for these different sources of genetic variation.

Guest Box 12 Mutation, inbreeding depression, and adaptation

Philip W. Hedrick

Mutation is the original source of genetic variation, whether it be advantageous variants, neutral variants, or detrimental variants. Most mutations have virtually no effect on fitness or a slightly detrimental one and, as a result, the majority of mutations are lost due to the large chance effects of genetic drift for alleles at low frequency. However, there is an important group of mutations that have a higher deleterious impact on fitness and another group which have an advantageous impact on fitness. Both these latter categories of mutations are particularly important in conservation because deleterious variants are the cause of inbreeding depression (Hedrick & Garcia-Dorado 2016) and advantageous variants are the basis of future adaptation.

Let us first examine an example of deleterious variation and the genetic basis of inbreeding depression from a genomic survey of 28 offspring produced by selfing from a single eucalyptus tree in a species (*Eucalyptus grandis*) that ordinarily does not self-fertilize (Hedrick et al. 2016). In this parental tree, 9,560 SNPs were heterozygous and were examined in the progeny. If there were no selection, 50% would be expected to be homozygous for one of the two parental alleles, or identical by descent, at each of these loci, and 50% of the progeny would be expected to be heterozygous. However, in these 28 progeny, only 34% of these markers were homozygous and 66% were heterozygous, a difference from expectations that was present on all 11 chromosomes (Myburg et al. 2014).

As an example of these results, Figure 12.9 gives the observed proportion of the three genotypes for each of the 1,019 SNPs along chromosome 1. Except for a short region on the far right end of this chromosome, the proportion of the two homozygotes combined is much less than 0.5 and averages around 30%, while the proportion of heterozygotes is much greater than 0.5 and averages around 70%. These effects appear to be the result of strong selection at many genes along the chromosome that cause high mortality when made homozygous by one generation of self-fertilization. The greatest reduction in homozygosity is expected for markers that are closely linked to genetic mutants with high detrimental effects. It is likely that mutations at more than 100 genes over the genome, many with a substantial effect on viability, are contributing to inbreeding depression in this example.

Mutation importantly is also the source of new genetic variation for future adaptation. An example is in the flightless Galápagos cormorant, only first discovered in 1898, and

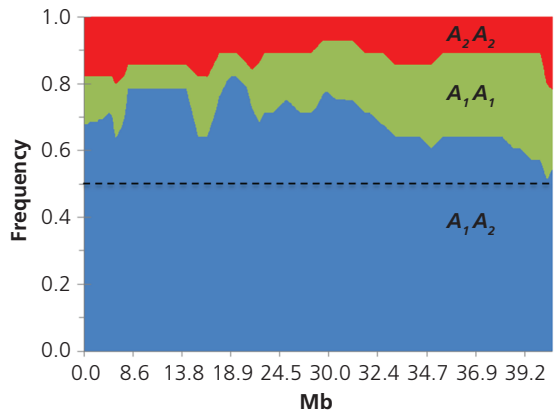


Figure 12.9 The observed proportions of heterozygotes (blue) and the two homozygotes (red and green) for 1,019 SNPs along chromosome 1 for 28 progeny produced by self-fertilization from a single parent in *Eucalyptus grandis*. The broken line gives the expected proportion of heterozygotes in the absence of selection. From Hedrick et al. (2016).

Guest Box 12 Continued

present only on the two youngest and most western of the Galápagos Islands (Hedrick 2019). The Galápagos cormorant is the only flightless cormorant of the ~40 recognized species of cormorants. Figure 12.10 shows the body and wing structure of the flightless Galápagos cormorant on the right and the closely related flighted double-crested cormorant on the left. Note that as well as rudimentary wings, about one-third of the size expected for a cormorant of that size, the flightless cormorant has a much larger body than the double-crested cormorant.

The western shores of the two most western Galápagos Islands are quite nutrient rich and appear to be the only suitable Galápagos environments that provide abundant and reliable food for the Galápagos cormorant. It is assumed that flightless cormorants descended from a flighted cormorant, colonized this new niche, and subsequently adapted to it by losing the ability to fly and becoming highly adapted to foraging in this environment where there are no native mammalian predators.

In an effort to determine the genetic mutants resulting in flightlessness, Burga et al. (2017) identified function-altering variants at multiple genes, none of which were found in related flighted cormorants. These variants are at genes that influence limb development and bone growth, cause genetic deformities in humans, and are all probably new mutations. All of the variants they identified at 11 different genes were fixed for the potentially function-altering variants in the flightless cormorants. In other words, mutations at multiple genes in the ancestor of the flightless cormorant, which were probably detrimental for a cormorant with flight, became adaptive and were fixed in this new environment.

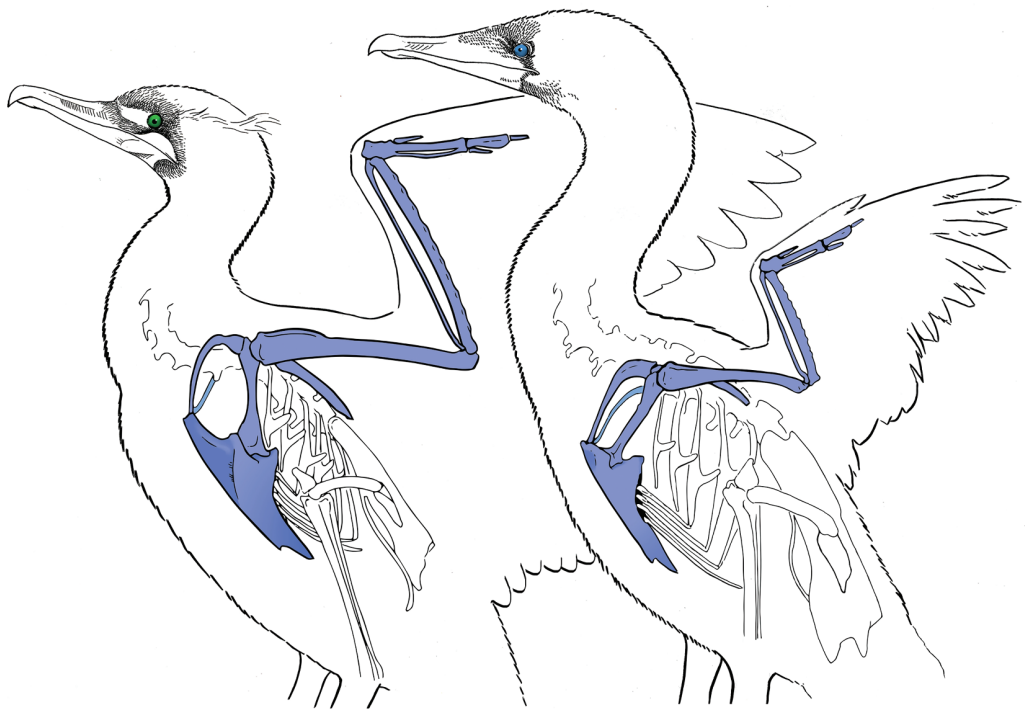


Figure 12.10 An illustration of a double-crested cormorant on the left and the flightless cormorant from the Galápagos Islands on the right showing the rudimentary wings and larger body size of the flightless cormorant. From Hedrick (2019); drawing by Katie Bertsche.