Chapter 2

Figure 2.1 The model of blending inheritance predicts that progeny have phenotypes that are the intermediate of their parents. Here "pure" blue and white parents yield light blue progeny, but these intermediate progeny could never themselves be parents of progeny with pure blue or white phenotypes identical to those in the P1 generation. Crossing any shade of blue with a pure white or blue phenotype would always lead to some intermediate shade of blue. By convention, in pedigrees females are indicated by circles and males by squares, whereas P refers to parental and F to filial.

Figure 2.2 Mendel's crosses to examine the segregation ratio in the seed coat color of pea plants. The parental plants (P1 generation) were pure breeding, meaning that if self-fertilized all resulting progeny had a phenotype identical to the parent. Some individuals are represented by diamonds since pea plants are hermaphrodites and can act as a mother, a father, or can self-fertilize.

Figure 2.3 Mendel self-pollinated (indicated by curved arrows) the F2 progeny produced by the cross shown in Figure 2.2. Of the F2 progeny that had a yellow phenotype (three-quarters of the total), one-third produced all progeny with a yellow phenotype and two-thirds produced progeny with a 3 : 1 ratio of yellow and green progeny (or three-quarters yellow progeny). Individuals are represented by diamonds since pea plants are hermaphrodites.

Figure 2.4 Mendel's crosses to examine the segregation ratios of two phenotypes, seed coat color (yellow or green) and seed coat surface (smooth or wrinkled), in pea plants. The hatched pattern indicates wrinkled seeds while white indicates smooth seeds. The F2 individuals exhibited a phenotypic ratio of 9 round/yellow:

3 round/green :

3 wrinkled/yellow :

1 wrinkled/green.

Figure 2.5 Hardy–Weinberg expected genotype frequencies for AA, Aa, and aa genotypes (y axis) for any given value of the allele frequency (x axis). Note that the value of the allele frequency not graphed can be determined by $q = 1 - p$.

Figure 2.6 A De Finetti diagram for one locus with two alleles. The triangular coordinate system results from the requirement that the frequencies of all three genotypes must sum to one. Any point inside or on the edge of the triangle represents all three genotype frequencies of a population. The parabola describes Hardy–Weinberg expected genotype frequencies. The dashed lines represent the frequencies of each of the three genotypes between zero and one. Genotype frequencies at any point can be determined by the length of lines that are perpendicular to each of the sides of the triangle. A practical way to estimate genotype frequencies on the diagram is to hold a ruler parallel to one of the sides of the triangle and mark off the distance on one of the frequency axes. The point on the parabola is a population in Hardy–Weinberg equilibrium where the frequency of AA is 0.36, the frequency of aa is 0.16, and the frequency of Aa is 0.48. The perpendicular line to the base of the triangle also divides the bases into regions corresponding in length to the allele frequencies. Any population with genotype frequencies not on the parabola has an excess (above the parabola) or deficit (below the parabola) of heterozygotes compared to Hardy–Weinberg expected genotype frequencies.

Figure 2.7 A schematic representation of random mating as a cloud of gas where the frequency of A alleles is 14/24 and the frequency of a alleles is 10/24. Any given A has a frequency of 14/24 and will encounter another A with probability of 14/24 or an a with the probability of 10/24. This makes the frequency of an A–A collision $(14/24)^2$ and an A–a collision $(14/24)(10/24)$, just as the probability of two independent events is the product of their individual probabilities. The population of A and a alleles is assumed to be large enough so that taking one out of the cloud will make almost no change in the overall frequency of its type.

Figure 2.8 The original data for the DNA profile given in Table 2.2 and Problem box 2.1 obtained by capillary electrophoresis. The PCR oligonucleotide primers used to amplify each locus are labeled with a molecule that emits blue, green, or yellow light when exposed to laser light. Thus, the DNA fragments for each locus are identified by their label color as well as their size range in base pairs. (a) A simulation of the DNA profile as it would appear on an electrophoretic gel (+ indicates the anode side). Blue, green, and yellow label the 10 DNA-profiling loci, shown here in grayscale. Other DNA fragments are size standards (originally in red) with a known molecular weight used to estimate the size in base pairs of the other DNA fragments in the profile. (b) The DNA profile for all loci and the size-standard DNA fragments as a graph of color signal intensity by size of DNA fragment in base pairs. (c) A simpler view of trace data for each label color independently with the individual loci labeled above the trace peaks. A few shorter peaks are visible in the yellow, green, and blue traces of (c) that are not labeled as loci. These artifacts, called pull-up peaks, are caused by intense signal from a locus labeled with another color (e.g. the yellow and blue peaks in the location of the green-labeled amelogenin locus). A full-color version of this figure is available on the textbook website.

Figure 2.9 A χ^2 distribution with one degree of freedom. The χ^2 value for the Hardy–Weinberg test with MN blood group genotypes as well as the critical value to reject the null hypothesis are shown (see text for details). The area under the curve to the right of the arrow indicates the probability of observing that much or more difference between the observed and expected outcomes.

Figure 2.10 Corn cobs demonstrating yellow and purple seeds that are either wrinkled or smooth. For a color version of this image see Plate 2.10.

Figure 2.11 An allozyme gel stained to show alleles at the phosphoglucomutase or PGM locus in striped bass and white bass. The right-most three individuals are homozygous for the faster-migrating allele (FF genotype) while the left-most four individuals are homozygous for the slower-migrating allele (SS genotype). No double-banded heterozygotes (FS genotype) are visible on this gel. The + and − indicate the cathode and anode ends of the gel, respectively. Wells where the individuals samples were loaded into the gel can be seen at the bottom of the picture. Gel picture kindly provided by J. Epifanio.

Figure 2.12   The impact of complete positive genotypic assortative mating (like genotypes mate) or self-fertilization on genotype frequencies. The initial genotype frequencies are represented by D, H, and R. When either of the homozygotes mates with an individual with the same genotype, all progeny bear their parent's homozygous genotype. When two heterozygote individuals mate, the expected genotype frequencies among the progeny are one half heterozygous genotypes and one quarter of each homozygous genotype. Every generation the frequency of the heterozygotes declines by one-half while one-quarter of the heterozygote frequency is added to the frequencies of each homozygote (diagonal arrows). Eventually, the population will lose all heterozygosity although allele frequencies will remain constant. Therefore, assortative mating or self-fertilization changes the packing of alleles in genotypes but not the allele frequencies themselves.

Figure 2.13   The impact of various systems of consanguineous mating or inbreeding on heterozygosity, the fixation index (F), and the inbreeding coefficient (f) over time. Initially, the population has allele frequencies of $p = q = 0.5$ and all individuals are assumed randomly mated. Since inbreeding does not change allele frequencies, expected heterozygosity (H_e) remains 0.5 for all 20 generations. As inbreeding progresses, observed heterozygosity declines and the fixation index and inbreeding coefficient increase. Selfing is 100% self-fertilization whereas mixed mating is 50% of the population selfing and 50% randomly mating. Full sib is brother–sister or parent–offspring mating. Backcross is one individual mated to its progeny, then to its grand progeny, then to its great-grand progeny and so on, a mating scheme that is difficult to carry on for many generations. Change in the coefficient of inbreeding over time is based on the following recursion equations: selfing $f_{t+1} = {}^1\!/_2(1 - F_t)$; mixed $f_{t+1} = {}^1\!/_2(1 - F_t)(s)$ where s is the selfing rate; full sib $f_{t+2} = {}^1\!/_4(1 + 2f_{t+1} + f_t)$; backcross $f_{t+1} = {}^1\!/_4(1 + 2f_t)$ (see Falconer & MacKay 1996 for detailed derivations).

Figure 2.14   Average relatedness and autozygosity as the probability that two alleles at one locus are identical by descent. (a) A pedigree where individual A has progeny that are half-siblings (B and C). B and C then produce progeny D and E, which in turn produce offspring G. (b) Only the paths of relatedness where alleles could be inherited from A, with curved arrows to indicate the probability that gametes carry alleles identical by descent. Upper-case letters for individuals represent diploid genotypes and lower-case letters indicate allele copies within the gametes produced by the genotypes. The probability that A transmits a copy of the same allele to B and C depends on the degree of inbreeding for individual A, or F_A.

Figure 2.15   The possible patterns of transmission from one parent to two progeny for a locus with two alleles. Half of the outcomes result in the two progeny inheriting an allele that is identical by descent. The a and a′ refer to paths of inheritance in the pedigree in Fig. 2.14b.

Figure 2.16   A graphical depiction of the predictions of the dominance and overdominance hypotheses for the genetic basis of inbreeding depression. The line for dominance shows purging and recovery of the population mean under continued consanguineous mating expected if deleterious recessive alleles cause inbreeding depression. However, the line for overdominance as the basis of inbreeding depression shows no purging effect since heterozygotes continue to decrease in frequency. The results of an inbreeding depression experiment with mice show that litter size recovers under continued brother–sister mating as expected under the dominance hypothesis (Lynch 1977). Only two of the original 14 pairs of wild-caught mice were left at the sixth generation. Not all of the mouse phenotypes showed patterns consistent with the dominance hypothesis.

Figure 2.17   Maps for human chromosomes 18 (left) and 19 (right) showing chromosome regions, the physical locations of identified genes and open reading frames (labeled orf) along the chromosomes, and the names and locations of a subset of genes. Chromosome 18 is about 85 million bp and chromosome 19 is about 67 million bp. Maps from NCBI Map Viewer based on data as of January 2008.

Figure 2.18   A schematic diagram of the process of recombination between two loci, A and B. Two double-stranded chromosomes (drawn in different colors) exchange strands and form a Holliday structure. The crossover event can resolve into either of two recombinant chromosomes that generate new combinations of alleles at the two loci. The chance of a crossover event occurring generally increases as the distance between loci increases. Two loci are independent when the probability of recombination and non-recombination are both equal to ${}^1\!/_2$. Gene conversion, a double crossover event without exchange of flanking strands, is not shown.

Figure 2.19   The decay of gametic disequilibrium (D) over time for four recombination rates. Initially, there are only coupling ($P_{11} = P_{22} = {}^1\!/_2$) and no repulsion gametes ($P_{12} = P_{21} = 0$). Gametic disequilibrium decays as a function of time and the recombination rate ($D_{tn} = D_{t0}(1 - r)^n$) assuming a single large population, random mating, and no counteracting genetic processes. If all gametes were initially repulsion, gametic disequilibrium would initially equal −0.25 and decay to 0 in an identical fashion.

Figure 2.20   The decay of gametic disequilibrium (D) over time when both strong natural selection and recombination are acting. Initially, there are only coupling ($P_{11} = P_{22} = {}^1\!/_2$) and no repulsion gametes ($P_{12} = P_{21} = 0$). The relative fitness values of the AAbb and aaBB genotypes are 1 while all other genotypes have a fitness of ${}^1\!/_2$. Unlike in Fig. 2.19, gametic disequilibrium does not decay to zero over time due to the action of natural selection.

Figure 2.21   The decay of gametic disequilibrium (D) over time with random mating ($s = 0$) and three levels of self-fertilization. Initially, there are only coupling ($P_{11} = P_{22} = {}^1\!/_2$) and no repulsion gametes ($P_{12} = P_{21} = 0$). Self-fertilization slows the decay of linkage disequilibrium appreciably even when there is free recombination ($r = 0.5$).

Figure 2.22   Expected levels of gametic disequilibrium (ρ^2) due to the combination of finite effective population size (N_e) and the recombination rate (r). Chance sampling will maintain gametic disequilibrium if very few recombinant gametes are produced (small r), the population is small so that gamete frequencies fluctuate by chance (small N_e), or if both factors in combination cause chance fluctuations in gamete frequency (small $N_e r$).

Chapter 3

Figure 3.1 Beakers filled with micro-centrifuge tubes can be used to simulate the process of sampling and genetic drift. For a color version of this image see Plate 3.1.

Figure 3.2 The Wright–Fisher model of genetic drift uses a simplified view of biological reproduction where all sampling occurs at one point: sampling $2N$ gametes from an infinite gamete pool. In this case N diploid individuals ($N/2$ of each sex) generate an infinite pool of gametes where allele frequencies are perfectly represented, and a finite sample of $2N$ alleles is drawn from this gamete pool to form N new diploid individuals in the next generation. Genetic drift takes place only in the random sample of $2N$ gametes to form the next generation. Major assumptions include non-overlapping generations, equal fitness of all individuals, and constant population size through time. The model can easily be adjusted for haploid individuals or loci by assuming $2N$ individuals or sampling N gametes to form the next generation.

Figure 3.3 The results of genetic drift continued every generation in populations of $N = 4$ and $N = 20$. In the top panels, the six lines represent independent replicates or independent populations experiencing genetic drift starting at the same initial allele frequency ($p = 0.5$). The random nature of genetic drift can be seen by the zig-zag changes in allele frequency that have no apparent direction. Allele frequencies that reach the upper or lower axes represent cases of fixation or loss. In the bottom panels, the genotype frequencies are shown for the allele frequencies represented by the black and blue lines under the assumption of random mating within each generation. The changes in genotype frequencies are a consequence of changes in allele frequencies due to genetic drift.

Figure 3.4 The results of genetic drift with different initial allele frequencies. The two panels have identical population sizes ($N = 25$) but initial allele frequency is $p = 0.2$ on the right and $p = 0.8$ on the left. The chances of fixation are equal to the initial allele frequency, a generalization that can be seen by examining a large number of replicates of genetic drift. Consistent with this expectation, more replicates go to loss on the left and more reach fixation on the right. Even with the difference in initial allele frequencies, the random trajectory of allele frequencies is apparent.

Figure 3.5 Probability distributions for binomial random variables based on samples of $N = 4$ (left) and $N = 20$ (right) from populations where $p = q = 0.5$. These distributions describe the expected probability of each of the possible outcomes of the micro-centrifuge sampling experiment described in the text.

Figure 3.6 The standard error of the allele frequency $\left(\sigma = \sqrt{\dfrac{pq}{2N}} \right)$ for a binomial random variable for a sample size of $N = 10$ for a range of allele frequencies. The standard error of the allele frequency decreases as the allele frequency approaches fixation or loss. In the same way, genetic drift is less effective at spreading out the distribution of allele frequencies as alleles approach fixation or loss. The standard deviation is 0 when the allele frequencies are 0 or 1 since there is no genetic variation and any size sample will faithfully reproduce the allele frequencies in the source population.

Figure 3.7 Probability distributions for binomial random variables based on samples of $N = 20$ from populations where the allele frequency is 0.50, 0.75, or 0.95 (dark blue, white, and light blue bars, respectively). The range of probable outcomes with sampling depends on the allele frequency. As allele frequencies approach the boundaries of fixation or loss, there is a decreasing number of outcomes other than fixation or loss that are probable due to sampling error.

Figure 3.8 A schematic illustration of how the effects of genetic drift due to sampling error depend on allele frequencies in a population. The horizontal axis represents allele frequency and the width of the arrows represents the standard error of allele frequency $\left(\sigma = \sqrt{\dfrac{pq}{2N}} \right)$ at a given allele frequency. Larger standard errors for allele frequency are another way of saying that sampling error will cause a greater range of outcomes, equivalent to a larger effect of genetic drift.

Figure 3.9 The expected frequencies of populations with zero, one, or two A alleles over five generations genetic drift. Initially, all populations have one A and one a allele ($p = q = 0.5$). Each generation two gametes are sampled from each population under the Wright–Fisher model to found a new population. This distribution assumes a very large number of independent replicate populations.

Figure 3.10 Genetic drift modeled by a Markov chain. In this case, the sample size is two diploid genotypes ($2N = 2$) or four gametes per generation. Initial allele frequencies in all populations are $p = q = 0.5$. In one generation, sampling error shifts some proportion of the initial populations that contain two copies of each allele to states of zero, one, two, three, or four copies of one allele. Between generations one and two, sampling error again shifts some proportion of the initial populations to states of zero, one, two, three, or four copies of one allele. However, in generation one there are populations present with all allelic states. The arrows represent the possible allelic states produced by sampling error in the third generation for each of the states in the second generation. The bars in the histogram for the third generation are divided by horizontal lines to show the contributions of each second generation allelic state to the total frequency of populations with a given allelic state (some contributions are very small and are difficult to see). As the Markov process continues, the frequency distribution accumulates more and more of the populations at states of zero and four alleles, eventually reaching fixation or loss for all populations.

Figure 3.11 (a) Allelic states (or allele frequencies) for 107 *D. melanogaster* populations where 16 individuals (eight of each sex) were randomly chosen to start each new generation. Initially, all 107 populations had equal numbers of the wild-type and bw^{75} alleles (the latter causes homozygotes to have a red-orange and heterozygotes an orange body color so genotypes can be determined visually). The allelic states of the population rapidly spread out and many populations reached fixation or loss by the nineteenth generation. (b) The expected frequency of populations in each allelic state determined with a Markov chain model for a population size of 16 with 107 populations that initially have equal frequencies of two alleles. The experimental *D. melanogaster* populations show a higher rate of fixation and loss than the model populations, suggesting that the population size was actually less than 16 individuals each generation. The *D. melanogaster* data come from Table 13 in Buri (1956).

Figure 3.12   An imaginary Petri dish that confines ink particles so that they can move only to the left or to the right from their current position. The particles have a constant velocity, so each will move the distance δ in a fixed amount of time. If the direction of particle movement is random (equal probability of moving left or right at any moment in time), the mean position of particles does not change but the variance in particle position increases with time. The frequency of particles passing through an area, such as the plane at *x*, depends on the net balance of particles arriving minus those that are leaving, called the flux of particles. The flux is determined by both the rate of diffusion of particles and gradients in the concentration of particles (net movement of particles is from areas of higher concentration to areas of lower concentration). If the left and right boundaries capture particles, then the diffusion coefficient drops to 0 at those points and particles will accumulate. The process of diffusion for particles is analogous to the process of genetic drift for allele frequencies in an ensemble population where allele frequencies "diffuse" because of sampling error.

Figure 3.13   Probability densities of allele frequency for many replicate populations predicted using the diffusion equation. The initial allele frequency is 0.5 on the left and 0.1 on the right. Each curve represents the probability that a single population would have a given allele frequency after some interval of time has passed. The area under each curve is the proportion of alleles that are not fixed. Time is scaled in multiples of the effective population size, *N*. Both small and large populations have identically shaped distributions, although small populations reach fixation and loss in less time than large populations. The populations that have reached fixation or loss are not shown for each curve. For a color version of this image see Plate 3.13.

Figure 3.14   Average time that an allele segregates, takes to reach fixation, or takes to reach loss depending on its initial frequency when under the influence of genetic drift alone. Alleles remain segregating (persist) for an average of $2.8N$ generations when their initial frequency is $^1/_2$. Fixation or loss takes up to an average of $4N$ generations when alleles are initially very rare or nearly fixed, respectively. Since these are average times, alleles in individual populations experience longer and shorter fixation, loss, and segregation times. Time is scaled in multiples of the population size.

Figure 3.15   Sewall Wright (1889–1988) with a guinea pig in an undated photograph taken during his years as a professor at the University of Chicago. Starting in 1912 and throughout his career, Wright studied the genetic basis of coat colors and physiological traits in guinea pigs. Wright, along with J.B.S. Haldane and R.A. Fisher, established many of the early expectations of population genetics using mathematical analyses. Many of the conceptual frameworks in population genetics today were originated by Wright, especially those related to consanguineous mating, genetic drift, and structured populations. An often retold (although mythical) story was that Wright, who would sometimes carry a guinea pig with him, would on occasion absent-mindedly employ the animal to erase the chalk board while lecturing. Provine's (1986) biography details Wright's manifold contributions to population genetics and his interactions with other major figures such as Fisher. Photograph courtesy of Special Collections Research Center, University of Chicago Library. Archival Photofiles, Series 1, Wright, Sewall, Informal #2.

Figure 3.16   Schematic representation of a genetic bottleneck where census population size fluctuates across generations. The harmonic mean of census population size (*N*) is 25 and provides an estimate of the effects of genetic drift over three generations, or the effective population size (N_e). In other words, this population of fluctuating size would experience as much genetic drift as an ideal Wright–Fisher population with a constant population size of 25. The squares represent alleles present in each generation. In the first generation the alleles are equally frequent, but they end up at frequencies of 25 and 75%. Such an allelic state transition would be extremely unlikely if 200 gametes were sampled to found generations two and three. However, the observed allelic transition is expected 25% of the time in a sample of four gametes.

Figure 3.17   Distributions of family size. The variance equals the mean as expected for a Poisson distribution on the left. However, the center distribution has a few families that are very prolific while 75% of the families produce two or fewer progeny with most individuals failing to reproduce. The distribution on the right has less variance in family size than expected for a Poisson distribution with most families of size two. The Poisson distribution is taken as the standard with an effective size of 100. By comparison, the center distribution has a smaller effective population size and the distribution on the right a larger effective population size.

Figure 3.18   Autozygosity and allozygosity in a finite population where identity by descent is related to the size of the population. Finite populations accumulate genotypes containing alleles identical by descent through random sampling in a manner akin to mating among relatives. In this example, alleles in the ancestral gamete pool identical in state are not identical by descent. Sampling of alleles takes place to form the diploid genotypes of the next generation. By chance, the same allele can be sampled twice to form an autozygous genotype with probability $\frac{1}{2N_e}$. The chance of not sampling the same allele twice is the probability of all other outcomes or $1 - \frac{1}{2N_e}$. Autozygous genotypes must be homozygous but allozygous genotypes can be either homozygous or heterozygous.

Figure 3.19   The decline in heterozygosity as a consequence of genetic drift in finite populations. The solid lines show expected heterozygosity over time according to $H_t = \left(1 - \frac{1}{2N_e}\right)H_{t-1}$. The decrease in heterozygosity can also be thought of as an increase in autozygosity or the fixation index (*F*) through time under genetic drift. The dotted lines in each panel are levels of heterozygosity (2*pq*) in six replicate finite populations experiencing genetic drift. There is substantial random fluctuation around the expected value for any individual population.

Figure 3.20   Simulated allele frequencies in 10 replicate populations that experienced effective population sizes of 100, 10, 50, and 100 individuals across four generations. The variance in the change in allele frequency (Δ*p*) can be used to estimate the variance effective population size. The inbreeding effective population size can be estimated from the change in heterozygosity through time.

Figure 3.21   Isolation by distance is characterized by the declining probability of gamete dispersal with increasing geographic distance. The specific shape of the gamete dispersal by distance curve may vary (top), but it is often modeled using a normal distribution (bottom). In a normal distribution, about 95% of observations are expected to fall within two standard deviations from the mean. Empirical estimates of mating and progeny movement in a generation can be used to estimate the variance and thereby the standard deviation of overall dispersal in gametes in order to estimate the area of a genetic neighborhood.

Figure 3.22   An ideal two-dimensional normal distribution used to model the size of genetic neighborhoods and to estimate the breeding effective size (N_e^b) of demes within continuous populations. The radius of the distribution is twice the standard deviation in total gamete dispersal in a generation. The actual physical dimensions of the distribution could range from just a few meters to hundreds or thousands of kilometers.

Figure 3.23   Haploid (a) and diploid (b) reproduction in the context of coalescent events. In a haploid population, the probability of coalescence is $\frac{1}{2N}$ (dashed lines) whereas the probability that two lineages do not have a common ancestor in the previous generation is $1 - \frac{1}{2N}$ (solid lines). In a diploid population, the two gene or allele copies in one individual in the present time have one ancestor in the female population (N_f) and one ancestor in the male population (N_m). Coalescent events in the diploid population arise when the gene copies in males and females are identical by descent. The haploid model with $2N$ lineages is routinely used to approximate the diploid model with $N = N_f + N_m$ diploid individuals.

Figure 3.24   The distribution of the probabilities of coalescence over time predicted by $\left(1 - \frac{1}{2N}\right)^{t-1} \frac{1}{2N}$ (top) for a population of $2N = 4$. The bottom panel shows the approximations of these coalescence probabilities based on geometric and exponential distributions with a probability of "success" of $^1/_4$.

Figure 3.25   Six independent realizations of the coalescent tree for six lineages. All six trees are drawn to the same scale. Each genealogy exhibits coalescent events between random pairs of lineages. The differences in the height of the trees is due to random variation in waiting times. Because of this random variation, average times to coalescence for a given number of lineages are only approached in a large sample of independent trees. Because genealogies are drawn sideways here, tree "heights" are actually width, from left to right.

Figure 3.26   A schematic coalescent tree the shows average coalescence times and the relationship between time scales in continuous units and units of $2N$ generations. With a sample of k lineages, the expected or average time to coalescence is $\frac{2N}{\frac{k(k-1)}{2}}$ since each of the independent $\frac{k(k-1)}{2}$ pairs of lineages can coalesce. Dividing continuous time in generations (j) by $2N$ yields a continuous time scale ($t = j/(2N)$) that can be used to describe coalescent trees. Multiplying the continuous time scale by $2N$ gives time to coalescence in units of population size. E refers to expected and T refers to time to coalescence so that $E(T_n)$ is the expected or average time to coalescence for n lineages. The basic patterns seen in all coalescent trees apply to populations of all sizes, although the absolute time for coalescent events does depend on N. Values on the left are one realization of coalescent waiting times. Genealogy is not drawn to scale.

Figure 3.27   The distribution of times to a MRCA (or genealogy heights) for 1000 replicate genealogies starting with six lineages ($k = 6$). The distribution of total coalescence times has a large variance because the range of times is large and also asymmetric with a long tail of a few genealogies that take a very long time to reach the MRCA. The genealogies shown above the distribution are those for the tenth, fiftieth, and ninetieth percentile times to MRCA. In this example $N_e = 1000$.

Figure 3.28   The effects of a population bottleneck on gene genealogies. During the bottleneck the chance that two randomly sampled gene copies are derived from one copy in the previous generation $\left(\frac{1}{2N_e}\right)$ increases. This can also be thought of as a reduction in the overall height of a genealogical tree caused by the bottleneck since lineages that find their ancestors during the bottleneck lead to short branches. The overall effect of a bottleneck on coalescence among gene copies sampled in the present depends on the reduction in the effective population size and the duration. The arrows indicate the point in time when gene copies were sampled from the population.

Figure 3.29   The effects of exponential population growth or shrinkage on coalescent genealogies. The upper panels show change in population size over time with exponential growth according to $N(t) = N_0 e^{-rt}$ with $r = \pm 0.1$, yielding relatively slow exponential population growth. The two genealogies illustrate examples of waiting times that might be seen under strong exponential population growth (left) and shrinkage (right). With strong exponential population growth coalescent times are longest in the present when the population is the largest, leading to genealogies characterized by long branches near the present and very short branches in the past around the time of the MCRA. With exponential population shrinkage, coalescence times are greatest in the past near the MRCA when the population was larger and shortest near the present when the population is at its smallest size. The genealogy on the left was obtained using equation 3.91 with $r = 100$.

Chapter 4

Figure 4.1   An example of population structure and allele-frequency divergence produced by limited gene flow. The total population (large ovals) is initially in panmixia and has Hardy–Weinberg expected genotype frequencies. Then the stream that runs through the population grows into a large river, restricting gene flow between the two sides of the total population. Over time allele frequencies diverge in the two subpopulations through genetic drift. In this example, you can imagine that the two subpopulations drift toward fixation for different alleles but neither reaches fixation due to an occasional individual that is able to cross the river and mate. Note that there is random mating (panmixia) within each subpopulation so that Hardy–Weinberg expected genotype frequencies are maintained within subpopulations. However, after the initial time period genotype frequencies in the total population do not meet Hardy–Weinberg expectations.

Figure 4.2   The plant *Linanthus parryae*, or desert snow, is found in the Mojave Desert region of California. (a) *L. parryae* can literally cover thousands of hectares of desert during years with rainfall sufficient to allow widespread germination of dormant seeds present in the soil. This tiny plant has either blue or white flowers. (b) In some locations most plants have blue or most have white flowers whereas in other locations more equal frequencies of the two flower colors are found. Reproduced with permission of Barbara J. Collins. For a color version of this image see Plate 4.2.

Figure 4.3   Isolation by distance causes spatial structuring of allele and genotype frequencies. In these pictures, a population is represented in two dimensions with each point on a grid representing one diploid individual. The colors represent an individual with a heterozygous (blue) or homozygous (black and white) genotype at each point. In (a) there is random mating over the entire population (the mating neighborhood is 99×99 squares) whereas in (b) there is strong isolation by distance (the mating neighborhood is 3×3 squares). The population with isolation by distance (b) develops and maintains spatial clumping of genotypes and therefore spatial clumping of allele frequencies. There is no such spatial structure in the population with random mating (a). In the simulation that produced these pictures, the grid is initially populated at random with genotypes at Hardy–Weinberg expected frequencies and $p = q = {}^1/_2$. Each generation every individual chooses a mate at random within its mating neighborhood and replaces itself with one offspring. The offspring genotypes are determined by Hardy–Weinberg probabilities for each combination of parental genotypes.

Figure 4.4   (*opposite*) Moran's I for simulated populations like those in Fig. 4.3. To estimate Moran's I, the 100×100 grid was simulated for 200 generations and was then divided into square subpopulations of 10×10 individuals. The frequency of the A allele within each subpopulation is y_i and the mean allele frequency over all subpopulations is $\bar{y}$ in equation 4.1. The distance classes are the number of subpopulations that separate pairs of subpopulations. As expected, the simulations with strong isolation by distance (3×3 mating neighborhood) show correlated allele frequencies in subpopulations that are close together. However, the simulations with panmixia (99×99 mating neighborhood) show no such spatial correlation of allele frequency. The fluctuation of I at the largest distances classes in both figures is random variation due to very small numbers of individuals compared. Each line is based on an independent simulation of the 100×100 population.

Figure 4.5   An individual *Corythophora alta* tree found at the Biological Dynamics of Forest Fragments Project field sites north of Manaus, Brazil. The plot shows the relative locations of the individual trees that make up the candidate parent population within a 9 ha forest plot at Cabo Frio. All trees can be both maternal parents and candidate paternal parents since the trees are hermaphrodites capable of self-fertilization.

Figure 4.6   Illustration of the hierarchical nature of heterozygosity in a subdivided population. $H_1 = \dfrac{3}{10}$ and $H_2 = \dfrac{3}{10}$ to give an average observed heterozygosity of $H_1 = \dfrac{1}{2}\left[\dfrac{3}{10} + \dfrac{3}{10}\right] = 0.30$. If p is the frequency of one allele (open circles) and q the frequency of the alternate allele (filled circles), then $p_1 = 13/20 = 0.65$ and $q_1 = 1 - p_1 = 0.35$ while $p_2 = 7/20 = 0.35$ and $q_2 = 1 - p_2 = 0.65$. The average expected heterozygosity in the two subpopulations is then $H_S = {}^1/_2[2(0.65)(0.35) + (2(0.35)(0.65)] = 0.455$. In the total population the average allele frequencies are $\bar{p} = {}^1/_2(0.65 + 0.35) = 0.50$ and $\bar{q} = {}^1/_2(0.35 + 0.65) = 0.50$ giving an expected heterozygosity in the total population of $H_T = 2\bar{p}\bar{q} = 2(0.5)(0.5) = 0.5$.

Figure 4.7   Allele frequencies at a diallelic locus for populations that consist of six subpopulations. Allele frequencies within subpopulations are indicated by shading. On the left, individual subpopulations are either fixed or lost for one allele. On the right, all subpopulations have identical allele frequencies of $p = q = 0.5$. In both cases, the total population has an average allele frequency of $\bar{p} = 0.5$ and an expected heterozygosity of $H_T = 2\bar{p}\bar{q} = 0.5$. In contrast, the average expected heterozygosity for subpopulations is $H_S = \overline{2pq} = 0.5$ on the right and $H_S = \overline{2pq} = 0.0$ on the left. $F_{ST} = 1.0$ on the left since the subpopulations have maximally diverged allele frequencies. $F_{ST} = 0.0$ on the right since the subpopulations all have identical allele frequencies. Divergence of allele frequencies among subpopulations produces a deficit of heterozygosity relative to the Hardy–Weinberg expectation based on average allele frequencies for the total population.

Figure 4.8   Allele frequencies, hierarchal heterozygosities, and fixation indices from a simulation of a subdivided population. Each subpopulation contained 10 individuals. The rate of gene flow was $m = 0.2$ in (a) and $m = 0.01$ in (b). The allele frequencies are shown for six randomly chosen subpopulations out of the 200 subpopulations in the simulation. The heterozygosities and fixation indices were calculated from all 200 subpopulations.

Figure 4.9   The distribution of F_{ST} values for 1000 replicate neutral loci in a finite island model of 200 subpopulations where each subpopulation contains 10 individuals and the rate of gene flow is 10% of each subpopulation ($m = 0.10$). In the distribution, 95% of the replicate loci show F_{ST} values between 0.1459 and 0.2002 whereas the average of all 1000 replicate loci is 0.1586 (based on the average of H_T and H_S then used to calculate F_{ST}). Replicate loci exhibit a range of F_{ST} values since allele frequencies among subpopulations are partly a product of the stochastic process of genetic drift. In an infinite island model with $N_e m = 1.0$ the expected value of F_{ST} is 0.2.

Figure 4.10   A graphical demonstration of the Wahlund effect for a diallelic locus in two demes. If there is random mating within subpopulations (H_1 and H_2) and in the total population (H_T), the heterozygosity of each falls on the parabola of Hardy–Weinberg expected frequency. The average heterozygosity of subpopulations (H_S) is at the mid-point between the deme heterozygosities. Therefore, H_S can never be greater than H_T based on the average allele frequency (the mid-point between the deme allele frequencies p_1 and p_2). Greater variance in allele frequencies of the demes is the same as a wider spread of deme allele frequencies in the two-deme case.

Figure 4.11   A hypothetical example of how the Wahlund effect relates variation in allele frequency between subdivided populations and genotype frequencies in a single panmictic population. Initially, the two subpopulations have different allele frequencies and therefore different frequencies of homozygous recessive albino phenotypes. The average frequency of the albino phenotype is 8% in the subpopulations. When the populations fuse, the allele frequencies become the average of the two subpopulations. However, the genotype frequencies are not the average of the two subpopulations. Rather, homozygotes become less frequent and heterozygotes more frequent than their respective subpopulation averages. In the fused population, the degree to which the frequencies of both homozygotes combined and the heterozygotes differ from their subpopulation averages is the same as the variance in allele frequency between the two subpopulations.

Figure 4.12   Classic models of population structure make different assumptions about the paths and rates of gene flow among subpopulations. (a) In the continent-island model, gene flow is essentially unidirectional from a very large population to a smaller population. The continent population is so large that allele frequencies are not impacted by emigration or drift whereas allele frequencies in the small population(s) are strongly influenced by immigration. (b) The island model has equal rates of gene flow exchanged by all populations regardless of the number of populations or their physical locations. (c, d) Stepping-stone models restrict gene flow to populations that are either adjacent or nearby in one (c) or two (d) dimensions and thereby incorporate isolation by distance. Gene-flow models can also incorporate the extinction and re-colonization of subpopulations, a feature commonly added to stepping-stone model populations. Each panel shows the rate of gene flow indicated by the arrows if m percent of each population is composed of migrants and $1 - m$ is composed of non-migrating individuals each generation.

Figure 4.13   Allele frequency in the island population for a diallelic locus under the continent-island model of gene flow. The island population allele frequencies (p^{island}) over time are shown for six different initial values (solid lines). The continent population has an allele frequency of $p^{continent} = 0.5$ shown by the dashed line. In the left-hand panel $m = 0.1$ and in the right-hand panel $m = 0.05$. Equilibrium is reached more slowly when the rate of gene flow is lower. In contrast, the difference in allele frequencies between the island and continent does not affect time to equilibrium for a given rate of gene flow. Note that the time scales in the two graphs differ.

Figure 4.14   Allele frequency in the two-island model of gene flow for a diallelic locus. Dashed lines in each panel highlight gene-flow-weighted average or equilibrium allele frequencies. Starting from allele frequencies of 0.9 and 0.2 and with equal rates of gene flow ($m = 0.1$), the subpopulations approach an equilibrium allele frequency of $\bar{p} = (0.9 + 0.2)/2 = 0.55$ (left panel). With initial allele frequencies of 0.9 and 0.2 but asymmetric rates of gene flow ($m_1 = 0.1$ and $m_2 = 0.05$), the subpopulations approach an equilibrium allele frequency of $\bar{p} = (0.9 \times 0.05 + 0.2 \times 0.1)/0.15 = 0.433$ (right panel). Equilibrium is reached more slowly in the case of asymmetric rates of gene flow on the right because the average rate of gene flow is lower. Note that the time scales in the two graphs differ.

Figure 4.15   Expected levels of fixation among subpopulations depend on the product of the effective population size (N_e) and the amount of gene flow (m) in the infinite island model of population structure. Each line represents expected F_{ST} for loci with different probabilities of autozogosity (from bottom to top $\frac{1}{2N_e}$, $\frac{1}{N_e}$, and $\frac{2}{N_e}$). Marked divergence of allele frequencies among subpopulations ($F_{ST} \geq 0.2$) are expected when $N_e m$ is below 1 for biparentally inherited nuclear loci with an autozygosity of $\frac{1}{2N_e}$. Y-chromosome or mitochondrial loci (autozygosity $= \frac{2}{N_e}$) are examples where marked divergence among populations is expected at higher levels of $N_e m$.

Figure 4.16   A hypothetical genealogy for two demes. Initially there are three lineages in each deme. The very first event going back in time is the migration of a lineage from deme one into deme two. Immediately after this migration occurs, the chance of coalescence in deme two increases since there are more lineages and the chance of coalescence in deme one decreases since there are fewer lineages. Continuing back in time, a coalescence event occurs in deme one and then a coalescence event occurs in deme two. The lineage that migrated out of deme one migrates back into deme one by chance. Coalescence to the single most recent common ancestor (MRCA) of all lineages cannot occur until the final two lineages are brought together in a single deme by migration.

Figure 4.17   Sample configurations for two lineages and two demes (a) and three lineages and three demes (b). Lineages are represented by the circles and the separation between demes is represented by a dotted vertical line. Only one possibility is given for each sample configuration even though some configurations can occur in multiple ways. For example, (0,1) can occur when both lineages are in the left-hand deme or when both lineages are in the right-hand deme.

Figure 4.18   The possible events that can occur when two lineages are in the same deme (0,1) or when two lineages are in two different demes (2,0) along with their probabilities of occurring. The separation between demes is represented by a dotted vertical line. Two lineages can coalesce only when they are in the same deme. The probability of coalescence (a), migration of one lineage such that the two lineages are in different demes (b), and migration that places both lineages in the same deme (c) determine the overall chances that two lineages coalesce. The chance that both lineages migrate (with probability m^2) is not shown in (b) and applies when there are three or more demes.

Figure 4.19   Genealogies for six lineages initially divided evenly between two demes when the migration rate is low (a) and when the migration rate is high (b). When migration is unlikely, coalescent events within demes tend to result in a single lineage within all demes before any migration events take place. There is then a long wait until a migration event places both demes in one deme where they can coalesce. When migration is likely, lineages regularly move between the demes, and lineages originally in the same deme are as likely to coalesce as lineages initially in different demes. These two genealogies are examples and substantial variation in coalescence times is expected. In (a) $M = 4N_e m = 0.2$ and in (b) $M = 4N_e m = 2.0$. The two genealogies are not drawn to the same scale. MRCA, most recent common ancestor.

Chapter 5

Figure 5.1   A hypothetical distribution of the effects of mutations on phenotypes that ultimately impact the Darwinian fitness of genotypes. Mutations that have a mean fitness less than the mean fitness of the population ($\bar{w}$) are decreased in frequency by natural selection. The shaded area around $\bar{w}$ indicates the zone where mutations have small effects on fitness relative to the effects of genetic drift (the width of the neutral zone depends on the effective population size). The shaded area near zero mean fitness indicates mutations that cause failure to reproduce or are lethal. Lethals are more common since it is a category that includes many degrees of severity resulting from diverse causes. The fitness effects of mutations are inherently difficult to measure because of the rarity of mutation events, the small effect of many mutations, and the dependence of fitness on environmental context.

Figure 5.2   The results of the classic *Drosophila melanogaster* mutation accumulation experiment carried out by Mukai et al. (1972). The experiment maintained three distinct sets of mutation-accumulation populations with 25 lines each. The left-hand panel shows the change in mean viability over time and the right-hand panel shows the change in the variance among replicate independent lines. Each point is the value obtained from one set of mutation-accumulation populations. Mutation of any type makes the lines diverge genetically and increases the variance. Mean viability declines over time as deleterious mutations are more common than advantageous mutations. Redrawn from Figure 2 in Mukai et al. (1972).

Figure 5.3   The probability that a novel mutation is lost from a population due to Mendelian segregation. A neutral allele is eventually lost from the population while a beneficial mutation has a probability of about twice its selective advantage of fixation. The cumulative probability over time is described by $e^{c(x-1)}$ where x is the probability of loss in the generation before and c is the degree of selective advantage, if any. This expected probability assumes an infinitely large population that has a Poisson-distributed variance in family size.

Figure 5.4   The frequencies over time of new mutations that each have an initial frequency of $\frac{1}{2N}$. In this example, one new mutation is introduced into the population every 30 generations and $N_e = 10$. All of the mutations except one (solid line) go to loss within a few generations. The one allele that does go to fixation takes a relatively long time to do so compared with the time to loss. At the start of the simulation the ancestral allele has a frequency of 1 (not shown). When a new mutation reaches fixation, the original ancestral allele is lost and the new mutation becomes the ancestral allele.

Figure 5.5   R.A. Fisher's geometric model of mutations fixed by natural selection. (a) Axes for two hypothetical phenotypes that determine fitness with maximum fitness when both phenotypes have the values at the center point marked with the black dot. An individual (or the mean phenotype of a population) with a phenotypic value is some distance from the maximum fitness. The dashed circle shows a perimeter of equal fitness around the point of maximum fitness. Although only two phenotypes define fitness in this example, the dashed circle of equal fitness would be a sphere with three phenotypes and an *n*-dimensional hyperspace if *n* phenotypes contribute to fitness. (b) Two mutations with smaller or larger phenotypic effects. The phenotypic effect of the mutations could be in any direction around the current phenotype (circles with radius *m*). Mutations with smaller effects have a better chance of moving the phenotype toward the maximum fitness (more of the area of the mutation-effect circle is to the left of the dashed line of equal fitness).

Figure 5.6   The probability that a mutation is fixed by natural selection depends on the magnitude of its effect on fitness. Using the geometric model of mutation and assuming that fitness is determined by many phenotypes, Fisher showed the probability that a mutation improves fitness approaches $^1/_2$ as the effect of a mutation approaches 0 (top panel). This result comes about because smaller mutations have a better chance of moving the phenotype toward the optimum than do larger mutations (see Fig. 5.5). Kimura pointed out that mutations with small effects on fitness are also the most likely to be fixed or lost due to genetic drift rather than by natural selection. Combining the chance that a mutation moves the phenotype toward higher fitness *and* the chance that a mutation has a large-enough fitness difference to escape genetic drift suggests that mutations with intermediate effects are most likely to be fixed by natural selection (bottom panel). Both models assume that mutations of any effect on fitness are equally likely to occur.

Figure 5.7   Simulation results show the action of Muller's Ratchet in increasing the number of deleterious mutations in the absence of recombination. Initially, all haploid individuals in the population have zero mutations. Mutations occur randomly over time and continually reduce the frequency of individuals with fewer mutations. Genetic drift causes sampling error and the stochastic loss of mutation classes with few individuals. Individuals with more mutations are less likely to reproduce, due to natural selection against deleterious alleles. Once the category with fewest mutations (e.g. the zero mutations class) is lost due to genetic drift and mutation, there is no process that can repopulate it. Therefore, the distribution of the number of mutations continually moves to the right but can never move back to the left. The simulation parameters were $N_e = 200$ and $\mu = 0.06$, each mutation reduced the chance of reproduction by 1%, and each individual had 100 loci.

Figure 5.8   Patterns of mutational change in DNA sequences under the infinite sites (a) and finite sites (b) models. Base-pair states created by a mutation are in blue lower-case letters. In the infinite sites model sequences that are identical in state at the same site are identical by descent because mutations only occur once at each site. In contrast, the finite sites model shows how multiple mutations at the same site act to obscure the history of identity based on comparisons of site differences among DNA sequences. The ellipses (. . .) that surround the sequences in (a) indicate that each sequence has infinitely many sites of which only 10 are displayed.

Figure 5.9   Expected change in allele frequency due to irreversible or one-way mutation for a diallelic locus for five initial allele frequencies. Here the chance that an A allele mutates into an a allele (or the per locus rate of mutation) is 0.00001. This rate of mutation is high compared with estimates of the per-locus mutation rate (see Table 5.1). The expected equilibrium allele frequency is $p = 0$ since there is no process acting to replace A alleles in the population. The population has not reached equilibrium even after 100,000 generations have elapsed. Changes in allele frequency due to mutation alone occur over very long time scales.

Figure 5.10   Expected change in allele frequency due to reversible or two-way mutation for a diallelic locus for five initial allele frequencies. Here the chance that an A allele mutates to an a allele (A $\rightarrow$ a) is 0.0001 and the chance that an a allele mutates to an A allele (a $\rightarrow$ A) is 0.00005. These mutation rates are toward the high end of the range of estimated mutation rate values (see Table 5.1). The expected equilibrium value is $p = 0.333$ according to equation 5.22, an allele frequency that is reached only after tens of thousands of generations. The time to equilibrium is proportional to the absolute magnitudes of the mutation rates while the equilibrium value depends only on a function of the ratio of the mutation rates.

Figure 5.11   Expected homozygosity (F or autozygosity, solid line) and heterozygosity (H or allozygosity, dashed line) at equilibrium in a population where the processes of both genetic drift and mutation are operating. The chance that two alleles sampled randomly from the population are identical in state depends on the net balance of genetic drift working toward fixation of a single allele in the population and mutation changing existing alleles in the population to new states. A critical assumption is the infinite alleles model, which guarantees that each mutation results in a unique allele and thereby maximizes the allozygosity due to mutations.

Figure 5.12   Haploid reproduction in the context of coalescent and mutation events. In a haploid population the probability of coalescence is $\dfrac{1}{2N}$ (solid lines) while the probability that the two lineages do not have a common ancestor in the previous generation is $1 - \dfrac{1}{2N}$ (dashed lines). The process of mutation can also occur simultaneously (stars), changing the state of alleles (filled circles to unfilled circles). Compare with Figure 3.23.

Figure 5.13   A genealogy constructed under the simultaneous processes of coalescence in a single finite population and mutation. Working backward in time from the present, both mutation and coalescence events can occur. The blue dots represent mutation events, each assigned at random to a lineage present when the event occurred. Mutation events alter the state of a lineage, causing divergence from the ancestral state of the most recent common ancestor (MRCA) of all the lineages in the present.

Figure 5.14   A genealogy constructed under the simultaneous processes of coalescence in a single finite population and mutation. Here the infinite alleles model of mutation is assumed to determine the allelic state of each lineage in the genealogy. Arbitrarily assigning allelic state A to the most recent common ancestor (MRCA), each mutational event then alters the state of the lineage experiencing the mutation. Each mutation changes the allelic state of the lineage to a new allele not present in the population, giving rise to a variety of allelic states among the lineages in the present.

Figure 5.15   A genealogy constructed under the simultaneous processes of coalescence in a single finite population and mutation. Here the infinite sites model of mutation for DNA sequences is assumed to determine the allelic state of each lineage in the genealogy. Arbitrarily assigning the DNA sequence ACTGCTAGCA to the most recent common ancestor (MRCA), each mutational event then alters the DNA sequence of the lineage experiencing the mutation. Each mutation occurs at a random site in the DNA sequence that has not previously experienced a mutation (bases in blue lower-case letters), giving rise to differences in the DNA sequences among the lineages in the present. Here each base is equally likely to be produced by a mutation, although there are numerous models to specify the pattern of nucleotide changes expected by mutation. Under the finite sites model of mutation, each site in the DNA sequence could experience mutation repeatedly.

Chapter 6

Figure 6.1   Population growth in two genotypes with clonal reproduction, starting out with equal numbers of individuals and therefore equal proportions in the total population. Genotype A grows 3% per generation ($\lambda = 1.03$) and genotype B grows 1% per generation ($\lambda = 1.01$). (a) Individuals of both genotypes increase in number over time. (b) Because the genotypes grow at different rates, their relative proportions in the total population change over time. The solid line shows the initial equal proportions. Eventually, genotype A will approach 100% and genotype B 0% of the total population. Values are plotted for every third generation.

Figure 6.2   Allele frequencies at the protease locus over time in the HIV populations in two patients undergoing protease inhibitor (ritonavir) treatment (Doukhan & Delwart 2001). Alleles found at very low frequencies before drug treatment come to predominate in the HIV population after drug treatment, due to natural selection among HIV genotypes for drug resistance. Alleles are bands observed in denaturing-gradient gel electrophoresis (DGGE), a technique that is capable of discriminating single-base-pair differences among different DNA fragments. DGGE was used to identify the number of different protease locus DNA sequences present in a sample of HIV particles. Protease inhibitor treatment began on day 0. Dr. E. Delwart kindly provided the original data used to draw this figure.

Figure 6.3   A diagram of the life cycle of organisms showing some points where differential survival and reproduction among genotypes can result in natural selection. Viability is the probability of survival from zygote to adult, mating success encompasses those traits influencing the chances of mating and the number of mates, and fecundity is the number of gametes and progeny zygotes produced by each mating pair. Gametic compatibility is the probability that gametes can successfully fuse to form a zygote whereas meiotic drive is any mechanism that causes bias in the frequency of alleles found in gametes. Most models of natural selection assume a single fitness component such as viability. In reality, all of these components of fitness can influence genotype frequencies simultaneously.

Figure 6.4   The change in genotype and allele frequencies caused by viability selection against the aa genotype exhibiting the recessive phenotype. The top panel shows the change in genotype frequencies over time and the bottom panel shows the frequency of the dominant allele (A) over time. The colored, dashed line in the bottom panel corresponds to the allele frequencies in the top panel. Because of changes in genotype frequency caused by natural selection, the frequency of the dominant allele rapidly approaches fixation from all five initial allele frequencies. In this illustration $w_{AA} = w_{Aa} = 1.0$ while $w_{aa} = 0.8$, meaning that eight individuals with the aa genotype are expected to survive to reproduce for every 10 individuals with the AA or Aa genotype that survive to reproduce each generation. Genotype frequencies assume random mating.

Figure 6.5   The change in the genotype and allele frequency of a completely dominant allele (A) when natural selection acts against the AA and Aa genotypes exhibiting the dominant phenotype. Notice that the frequency of the A allele decreases slowly at first when the A allele is common in the population since the aa genotype is infrequent. The colored, dashed line in the bottom panel corresponds to the allele frequencies in the top panel. In this illustration $w_{AA} = w_{Aa} = 0.8$ while $w_{aa} = 1.0$. Genotype frequencies assume random mating.

Figure 6.6   Allele frequencies over time for three types of gene action with a low initial allele frequency. In all three cases the equilibrium allele frequency is fixation or near fixation for the A allele. With complete dominance, natural selection initially increases the allele frequency very rapidly. The approach to fixation for the A allele slows as aa homozygotes become rare since heterozygotes harbor a alleles that are concealed from natural selection by dominance. Natural selection initially changes the frequency of a recessive allele very slowly since homozygote recessive genotypes are very rare. As the recessive homozygotes become more common, allele frequency increases more rapidly. With additive gene action the phenotype of the heterozygote is intermediate between the two homozygotes so all genotypes differ in their viability. Additive gene action has the most rapid overall approach to equilibrium allele frequency. The degree of dominance is represented by the dominance coefficient, h. In this illustration the selection coefficient is $s = 0.1$.

Figure 6.7   The change in the genotype and allele frequency when there is underdominance for fitness and natural selection acts against individuals with Aa genotypes. The equilibrium allele frequency depends on the initial allele frequency. Starting below 0.5 populations head toward loss while starting above 0.5 populations go to fixation. There is an unstable equilibrium at an initial allele frequency of exactly 0.5. From any initial allele frequency the population converges on a minimum frequency of heterozygotes. The colored, dashed line in the bottom panel corresponds to the allele frequencies in the top panel. In this illustration $w_{AA} = w_{aa} = 1.0$ and $w_{Aa} = 0.9$. Genotype frequencies assume random mating.

Figure 6.8   The change in the genotype and allele frequencies when there is overdominance for fitness and natural selection favors individuals with Aa genotypes. From any initial allele frequency the population converges on a maximum frequency of heterozygotes. This corresponds to equal allele frequencies with random mating. The colored, dashed line in the bottom panel corresponds to the allele frequencies in the top panel. In this illustration $w_{AA} = w_{aa} = 0.9$ and $w_{Aa} = 1.0$. Genotype frequencies assume random mating.

Figure 6.9   The strength of natural selection influences the rate of change in genotype and allele frequencies. In this illustration, selection acts against the recessive homozygote (aa). The top panel shows strong natural selection where viability of the aa genotype is 10–50% less than that of the other genotypes. The bottom panel shows weak natural selection where viability of the aa genotype is 1–0.1% less than that of the other genotypes. Note the vastly different time scales in the two plots.

Figure 6.10   Mean fitness in a population ($\bar{w}$) and change in allele frequency over a single generation (Δp) as a function of allele frequency for directional selection. Directional selection reaches allele frequency equilibrium at either fixation or loss, the point of highest mean fitness. Positive values of Δp (above the dashed line) indicate that allele frequency is increasing under selection while negative values of Δp (below the dashed line) indicate that allele frequency is decreasing under selection. The change in allele frequencies is faster when average fitness changes more rapidly (the slope of $\bar{w}$ is steeper). Here $w_{AA} = w_{Aa} = 1.0$ and $w_{aa} = 0.8$ for selection against a recessive phenotype and $w_{AA} = w_{Aa} = 0.8$ and $w_{aa} = 1.0$ for selection against a dominant phenotype.

Figure 6.11   Mean fitness in a population ($\bar{w}$) and change in allele frequency over a single generation (Δp) as a function of allele frequency for balancing and disruptive selection. Natural selection changes allele frequencies to increase the average fitness in each generation, eventually reaching an equilibrium when the mean fitness is highest. The change in allele frequencies is faster when average fitness changes more rapidly (the slope of $\bar{w}$ is steeper). The dashed line in the plots of Δp by p shows where allele frequencies stop changing ($\Delta p = 0$) and thus are allele frequency equilibrium points. With underdominance for fitness, Δp is zero when $p = 0.5$ and so defines an equilibrium point marked by the circle. However, this equilibrium point is unstable since Δp on either side of $p = 0.5$ changes allele frequencies *away* from the equilibrium point (below $p = 0.5$ Δp is negative leading toward loss and above $p = 0.5$ Δp is positive leading toward fixation). In contrast, with overdominance Δp on either side of $p = 0.5$ changes allele frequencies *toward* the equilibrium point (below $p = 0.5$ Δp is positive and above $p = 0.5$ Δp is negative) and thus $p = 0.5$ is a stable equilibrium point. Here $w_{AA} = w_{aa} = 1.0$ and $w_{Aa} = 0.7$ for underdominance and $w_{AA} = w_{aa} = 0.7$ and $w_{Aa} = 1.0$ for overdominance.

Figure 6.12   Sir Ronald A. Fisher (1890–1963) photographed in 1943, was a pioneer in the theory and practice of statistics. He invented the techniques of analysis of variance and maximum likelihood as well as numerous other statistical tests and methods of experimental design. Fisher's 1930 book *The Genetical Theory of Natural Selection* established a rigorous mathematical framework that coupled Mendelian inheritance and Darwin's qualitative model of natural selection and is one of the foundation works of modern population genetics. Much of Fisher's work stressed the effectiveness of natural selection in changing gene frequencies in infinite, panmictic populations. Photograph courtesy of the Master and Fellows of Gonville and Caius College, Cambridge.

Figure 6.13   A graphical illustration of R.A. Fisher's fundamental theorem of natural selection. The curved lines represent the product of the homozygote frequencies ($P = p^2$ and $R = q^2$) as a constant proportion of the square of the product of the allele frequencies ($Q = pq$) or $\lambda = Q^2/PR$. Hardy–Weinberg genotype frequencies produced by random mating represent the special case of $\lambda = 1$ (solid colored line). Mean fitness is represented by the grayscale gradient with darker tones representing higher mean fitness. In this illustration, genotype frequencies start out at z_1. Suppose that natural selection over one generation changes genotype frequencies to point z_3 (under the conditions that genotype AA has the highest fitness and additive gene action, for example). This change in genotype frequencies can be decomposed into two distinct parts. One part is the change from z_1 to z_2 moving along the curve where λ is constant but allele frequencies change from p to p'. The other part is the change in the genotype frequencies (changing the value of λ) that occurs by moving vertically on the De Finetti diagram from z_2 to z_3 but keeping allele frequencies constant. The fundamental theorem says that the change in the mean fitness by natural selection is proportional to the change in allele frequency alone. Processes other than natural selection, such as mating system, dictate the change in genotype frequencies. When natural selection moves the genotype frequencies along a curve of constant λ, then the total change in mean fitness is completely due to changes in allele frequency and genetic variation in fitness is completely additive. Modified from Edwards (2002).

Chapter 7

Figure 7.1   A fitness surface made by including mean fitness on a De Finetti plot of the three genotype frequencies for a diallelic locus. The colored lines indicate the possible trajectories of genotype frequencies as natural selection increases the mean fitness of the population. The fitness values are $w_{AA} = 1.0$, $w_{Aa} = 0.6$, and $w_{aa} = 0.2$ so the highest mean fitness is found in the lower left apex when the population is fixed for the AA genotype. This highest fitness point can be reached by continually increasing mean fitness from any initial point on the surface. Gene action is additive because alleles have a constant impact fitness regardless of the allele they are paired with in a genotype. An A allele always contributes 0.5 and an a allele 0.1 toward the fitness of a genotype.

Figure 7.2   Fitness surfaces for the A, S, and C alleles at the human hemoglobin β gene when malaria is common. The surface in (a) corresponds to the top set of fitness values in Table 7.1 and (b) shows the surface for the bottom set of values. The tracks of circles represent generation-by-generation allele frequency trajectories due to natural selection over 50 generations calculated with equation 7.3. In (a), when the initial frequency of the C allele is relatively high, the equilibrium of natural selection is the fixation of the CC genotype. In contrast, when the C allele is initially rare (a frequency of less than about 7%) selection reaches an equilibrium with only the A and S alleles segregating and the C allele going to loss. In (b), selection will eventually fix the CC genotype from any initial frequency of the C allele. However, when the C allele is at low frequencies, the increase in the C allele each generation is extremely small so that selection would take hundreds of generations to fix the CC genotype. The six initial allele frequency points, shown as open circles, are identical for the two surfaces.

Figure 7.3   A fitness surface for two loci that each have two alleles where gene action is additive. The blue dots show generation-by-generation allele frequencies based on equations 7.11–7.14 for seven different initial sets of four gamete frequencies. When recombination is a weak force ($r = 0.05$), equilibrium allele frequencies are dictated by natural selection and all initial gamete frequencies eventually reach the highest mean fitness point (a). In contrast, when recombination is a strong force ($r = 0.5$) then equilibrium allele frequencies depend on initial gamete frequencies (b). When recombination is strong, equilibrium allele frequencies may not correspond to the highest mean fitness. Relative fitness values are $w_{AABB} = 0.9$, $w_{AABb} = 0.8$, $w_{AAbb} = 0.7$, $w_{AaBB} = 0.7$, $w_{AaBb} = 0.6$, $w_{Aabb} = 0.5$, $w_{aaBB} = 0.5$, $w_{aaBb} = 0.4$, and $w_{aabb} = 0.3$. The seven initial allele frequency points, shown as open circles, are identical for the two surfaces.

Figure 7.4   A fitness surface for two loci that each have two alleles where gene action exhibits epistasis. When recombination is a weak force ($r = 0.05$), equilibrium allele frequencies are dictated by natural selection. Equilibrium allele frequencies depend on initial gamete frequencies since the two highest mean fitness points are separated by a fitness valley (a). When recombination is strong ($r = 0.5$), allele frequencies change such that mean fitness actually decreases for a time before increasing again to eventually reach the lower of the two mean fitness peaks (b). The two initial gamete frequencies in the upper right of the surface reach an equilibrium point where fitness is not maximized and there is gametic disequilibrium ($D = 0.041$). Relative fitness values are $w_{AABB} = 0.61$, $w_{AABb} = 0.58$, $w_{AAbb} = 0.50$, $w_{AaBB} = 1.0$, $w_{AaBb} = 0.77$, $w_{Aabb} = 0.50$, $w_{aaBB} = 0.64$, $w_{aaBb} = 0.62$, and $w_{aabb} = 0.92$. The seven initial allele frequency points, shown as open circles, are identical for the two surfaces.

Figure 7.5   The relative fitness of each genotype (w_{xx}) and the change in allele frequency (Δp) across all frequencies of the A allele under frequency-dependent natural selection. There is a stable equilibrium point at $p = 0.5$ in this particular case, even though the heterozygote has the lowest fitness. Two unstable equilibria at fixation and loss are marked with open circles. Here the relative fitness values are $w_{AA} = 1 - s_{AA}p^2$, $w_{Aa} = 1 - s_{Aa}2pq$, and $w_{aa} = 1 - s_{aa}q^2$ with $s_{AA} = s_{Aa} = s_{aa}$.

Figure 7.6   The results of density-dependent natural selection on the numbers of individuals of different genotypes (N_{AA}, N_{Aa}, and N_{aa}) and allele frequencies in a population of total size N. At the upper limit of N, the equilibrium allele and genotype frequencies are determined by the genotype with the highest carrying capacity (K). In contrast, the genotype with the highest growth rate (r) has the greatest impact on allele frequency when the population is small. In this example $K_{AA} = 10,000$, $K_{Aa} = 9000$, and $K_{AA} = 8000$ with $r_{AA} = 0.2$, $r_{Aa} = 0.25$, and $r_{aa} = 0.3$.

Figure 7.7   The expected distribution of allele frequencies for a very large number of replicate finite populations under natural selection where there is overdominance for fitness ($w_{AA} = w_{aa} = 1 - s$ and $w_{Aa} = 1$). In an infinite population the expected allele frequency at equilibrium is 0.5. However, in finite populations the equilibrium allele frequency will depend on the balance of natural selection and genetic drift. This balance is determined by the product of the effective population size and the selection coefficient ($N_e s$). Low values of $N_e s$ mean that selection is very weak compared to drift and each population reaches fixation or loss. High values of $N_e s$ mean that selection is strong compared to drift and most populations reach an equilibrium allele frequency near 0.5. Here forward and backward mutation rates are equal ($\mu = \nu = 0.00001$).

Figure 7.8   Frequency of the *Bar* allele in 108 replicate *D. melanogaster* populations over 10 generations (Wright & Kerr 1954). Each population was founded from four males and four females. The *Bar* locus is found on the X chromosome and so is hemizygous in males, making the effective population size equivalent to about six individuals. The eyes of *D. melanogaster* individuals homozygous for the wild-type allele are oval, but heterozygotes and homozygotes for the partially dominant *Bar* allele have bar-shaped eyes with a reduced number of facets. Females homozygous for the Bar allele produced 37% of the progeny compared to females homozygous or heterozygous for the wild-type allele. Despite this strong natural selection against *Bar*, three populations fixed for *Bar* by the end of the experiment. Compare with the similar example in Figure 3.11 where the locus is selectively neutral.

Figure 7.9   The absolute value of the change in allele frequency due to mutation ($\Delta q_{mutation}$) and due to natural selection ($\Delta q_{selection}$) when there is selection against a recessive homozygote. Mutation continually makes new copies of the recessive allele while selection continually works toward loss of the recessive allele. The equilibrium allele frequency occurs when the processes of mutation and selection exactly counteract each other. Here $s = 0.1$ and $\mu = 1 \times 10^{-6}$ so the expected equilibrium is $q_{equilibrium} = 0.0032$ as shown by the vertical arrow.

Figure 7.10   Haploid reproduction with the possibility of coalescence and natural selection events. Each haploid lineage replicates itself and is included in the next generation if there are no coalescence events (solid lines). A lineage makes an extra copy of itself (dashed line) and has the potential to displace one copy of another lineage. If the lineage making the extra copy of itself (open circle) has a higher-fitness haplotype than a randomly chosen lineage (blue circle), then it will displace the lineage of the lower fitness haplotype. Therefore, the outcome of a lineage-duplication event that may result in natural selection depends on the haplotype states of the specific lineages involved. The solid lines are continuing branches and the dashed line is an incoming branch. Compare with Figures 3.23 and 5.12.

Figure 7.11   The ancestral selection graph used to include natural selection in the genealogical branching model. In (a), the waiting times between events and the types of events are determined until the MRCA is reached by working backward in time from six lineages in the present. Branching and coalescence events due to natural selection (dashed lines) and mutation events (blue circles) are identified. Natural selection causes the addition of one "incoming" branch to the number of lineages that can coalesce and this incoming branch can then coalesce with any lineage. In (b), a haplotype state is assigned to the ultimate ancestor and allelic states are traced forward in time to determine the outcome of mutation and natural selection events. At each of the selection events the state of the continuing branch and the incoming branch are compared. In this example A is the fitter haplotype and it displaces the a haplotype when continuing and incoming branches coalesce. When the haplotypes of continuing and incoming branches are identical there is no change in haplotype state. In (c), the haplotype states of the lineages in the present are assigned once all of the selection events have been resolved. In this example selection causes a slight increase in the total branch length because the selection event at #3 displaces a shorter branch.

Figure 7.12   A genealogy where balancing natural selection is modeled by type switching. Every generation, lineages of one type (here A and B) may switch to the other type with rate μ. Twice the expected number of the $2N$ total lineages in the population that switch types each generation is then $\nu = 4N\mu$. Since lineages can only coalesce when they are of the same type, type switching increases the average time to coalescence. This is analogous to natural selection favoring heterozygotes because overdominance also extends the segregation times of alleles. Genealogical trees that result from balancing selection modeled as type switching tend to have longer branches compared to genealogies that result from genetic drift or directional natural selection.

Figure 7.13   An example ancestral selection graph showing how natural selection operates in the genealogical branching model. The genealogy on the left shows incoming and continuing branches and their haplotype states. The genealogy on the right shows the final pattern of branch states once all natural selection events have been resolved and states reassigned as necessary. In this illustration, the A haplotype is at a selective advantage over the a haplotype. Haplotype A increases in frequency when the continuing and incoming branches are resolved. See Problem box 7.2.

Chapter 8

Figure 8.1   Motoo Kimura (on left) and James Crow in 1986 on the occasion of Kimura being awarded an honorary doctoral degree at the University of Wisconsin, Madison, WI, USA. Kimura pioneered the use of diffusion equations to determine quantities such as the average time until fixation or until loss for neutral mutations. Based on these foundations, he proposed the neutral theory of molecular evolution in 1968. Kimura and Crow collaborated to develop some of the basic expectations for neutral genetic variation. Crow mentored and encouraged many people who became influential contributors to population genetics, including Kimura. Photograph kindly provided by J. Crow.

Figure 8.2   The fate of selectively neutral mutations in a population. New mutations enter the population at rate µ and an initial frequency of $\frac{1}{2N}$. Allele frequency is a random walk determined by genetic drift. The time that a new mutation segregates in the population, or the dwell time of a mutation, depends on the effective population size. However, the chance that a new mutation goes to fixation (equal to its initial frequency) is also directly related to the effective population size. These two effects of the effective population size cancel each other out for neutral alleles. The neutral theory then predicts that the rate of fixation is µ and therefore the expected time between fixations is $1/\mu$ generations. For that subset of mutations that eventually fix, the expected time from introduction to fixation is $4N_e$ generations. After Figure 3.1 in Kimura (1983a).

Figure 8.3   The dwell time for new mutations is different if fixation and loss is due to genetic drift or natural selection. With neutral mutations (b), most mutations go to loss fairly rapidly and a few mutations eventually go to fixation. For eventual fixation or loss of neutral mutations the path to that outcome is a random walk, implying that the time to fixation or loss has a high variance. For mutations that fix because they are advantageous (a), directional selection fixes them rapidly in the population. Therefore under directional selection alleles segregate for a shorter time and there is less polymorphism than with neutrality. For mutations that show overdominance for fitness, natural selection favoring heterozygote genotypes maintains several alleles in the population indefinitely. Therefore balancing selection greatly increases the segregation time of alleles and increases polymorphism compared to neutrality. Both cases of natural selection (a and c) are drawn to show negative selection acting against most new mutations. If new mutations are deleterious then the time to loss is very short and there is very little random walk in allele frequency since selection is nearly deterministic.

Figure 8.4   The process of divergence for two DNA sequences that descended as identical copies of an ancestral sequence. Each sequence experiences neutral mutations, some of which are eventually fixed by genetic drift. These fixed mutations replace all other alleles and are therefore substitutions (indicated by blue lower-case letters). As substitutions accumulate, the two sequences diverge from the ancestral sequence as well as from each other. In this example, the two sequences are eventually divergent at five of 12 nucleotide sites due to substitutions. The dashed line indicates complete isolation of the two populations containing the derived sequences.

Figure 8.5   The probability of eventual fixation for a new mutation under the neutral and nearly neutral theories. Under the nearly neutral theory the probability of fixation depends on the balance between natural selection and genetic drift, expressed in the product of the effective population size and the selection coefficient ($N_e s$). When negative selection operates against a deleterious allele, the selection coefficient and $N_e s$ are negative. Values of $N_e s$ near 0 yield a fixation probability close to that predicted by neutral theory. Only when the absolute value of $N_e s$ is large does natural selection exclusively determine the probability of fixation. Neutral theory assumes that neutral mutations are not influenced by selection and have a constant probability of fixation dictated by the effective population size. In this example the initial allele frequency is 0.001, or the frequency of a new mutation at a diploid locus in a population of 500. After Ohta (1992).

Figure 8.6   A DNA sequence electropherogram. See the text web page for a full-color version of this figure.

Figure 8.7   Substitutions that occur repeatedly at the same nucleotide site lead to saturation of nucleotide changes as time since divergence from a common ancestor increases. The rate of substitutions does not change and the total number of substitutions continues to increase over time, as shown by the dashed line in the top panel representing the true number of substitutions. In contrast, multiple substitutions at the same sites leads to a slowing and leveling off in the estimate of divergence (solid line, top panel). Therefore, the amount of divergence leads to the perception that the rate of divergence decreases over time. The bottom panel shows divergence and saturation at the mitochondrial cytochrome *c* oxidase subunit II gene among bovine species (ungulates including domestic cattle, bison, water buffalo, and yak) that diverged between 2 and 20 million years ago. In the top panel $\alpha = 1 \times 10^{-6}$ (α is explained overleaf). The bottom panel data are from Janecek et al. (1996) and the line is a quadratic regression fit.

Figure 8.8   The three types of event that a single nucleotide site may experience over two generations. A nucleotide site may initially have a G, for example. In case 1 a single substitution event in generation one changes the G to an A, C, or T nucleotide, the nucleotide also present at generation two. A *p* distance measure of divergence will accurately count the number of substitutions in case 1. In cases 2 and 3 the nucleotide site still retains the same nucleotide it had initially, giving the impression that there have been no substitutions. In case 2 this impression is accurate. However, in case 3 there have been two substitution events that are not accounted for in a simple *p* distance measure of divergence.

Figure 8.9   The probability that a nucleotide site retains its original base pair under the Jukes–Cantor model of nucleotide substitution. If a nucleotide site originally has a G base, for example, the probability of the same base being present declines steadily over time. If a nucleotide site was initially not a G (it was an A, C, or T), the probability that a G is present at the site increases over time. The probability that a given base is present always converges to 25% because that is the probability of sampling a given base at random if the probability of substitution to each nucleotide is equal. In the top panel $\alpha = 1 \times 10^{-6}$ whereas in the bottom panel $\alpha = 1 \times 10^{-5}$.

Figure 8.10   The hierarchy of nucleotide-substitution models that can be used to correct apparent divergence between DNA sequences to better estimate the actual number of substitutions that have occurred. The Jukes–Cantor (JC) model is the simplest and assumes that there is just one rate of substitution that applies to all nucleotide changes and is constant among nucleotide sites. Other nucleotide substitution models include an increasing number of parameters to represent more features of DNA sequence evolution, in particular variable rates of substitution among various categories of nucleotides. If nucleotide-substitution rates are variable among different sites, this variation can be modeled by a gamma distribution indicated by the Greek letter Γ. Nucleotide-substitution models: JC, Jukes–Cantor (Jukes & Cantor 1969); F81, Felsenstein 81 (Felsenstein 1981); K80, Kimura 80 (Kimura 1980); HKY, Hasegawa–Kishino–Yano (Hasegawa et al. 1985); SYM, symmetrical model (Zharkikh 1994); GTR, general time reversible (Rodriguez et al. 1990). Figure after Posada and Crandall (1998).

Figure 8.11   A hypothetical sample of four DNA sequences that are each 10 nucleotides long. There a total of three segregating sites ($S = 3$) or three-tenths of a segregating site per nucleotide ($p_S = 0.3$). The nucleotide diversity is calculated by summing the nucleotide sites that differ between each unique pair of DNA sequences. In this example there are 1.67 average pairwise nucleotide differences or 0.167 average pairwise nucleotide differences per nucleotide site.

Figure 8.12   Rates of nucleotide change in the NS gene that codes for "nonstructural" proteins based on 11 human influenza A virus samples isolated between 1933 and 1985. The number of years since isolation and DNA sequence divergence from an inferred common ancestor are positively correlated. The pattern of increasing substitutions as time since divergence increases is expected under the molecular clock hypothesis. The observed rate of substitution was approximately 1.9×10^{-3} substitutions per nucleotide site per year, a very high rate compared to most genes in eukaryotes. The line is a least-squares fit. Data from Buonagurio et al. (1986).

Figure 8.13   (*left*) Rates of protein evolution as amino acid changes per 100 residues in fibrinopeptides, hemoglobin, and cytochrome *c* over very long periods of time. Rates of divergence are linear over time for each protein, as expected for a molecular clock. Different proteins have different clock rates due to different mutation rates and degrees of functional constraint imposed by natural selection. Amino acid changes between pairs of taxa with unknown divergence times are plotted on dashed lines with the same slope as lines through points for taxa with estimated divergence times. The six points with unknown divergence times for hemoglobin represent divergence of ancestral globins into hemoglobins and myoglobins in the earliest animals, events that the molecular clock estimates to have happened between 1.1 billion and 800 million years ago. Data from Dickerson (1971).

Figure 8.14   A schematic phylogenetic tree that can be used to date divergence events under the assumption of a constant rate of divergence over time or a molecular clock. T_1 is the time in the past when species C and the ancestor of species A and B diverged. T_2 is the time in the past when species A and B diverged. If either T_1 or T_2 are known, the rate of molecular evolution per unit of time can be estimated from observed sequence divergences. This rate of divergence can then be used to estimate the unknown amount of time that elapsed during other divergences.

Figure 8.15   Substitution patterns under a Poisson process. The top panel shows the probability distribution for the number of substitutions that might occur during one time interval. $N(t)$ between 0 and 9 all have probabilities of greater than 0.01. The bottom panel shows the cumulative number of substitutions under a Poisson process for five independent trails. Each trail is akin to an independent lineage experiencing substitutions. The average number of substitutions is approximately 40 (four multiplied by the number of time intervals) but there is variation among the lineages. In both panels the rate of substitution is the same at $\lambda = 4$.

Figure 8.16   Two representations of rate at which substitution events (circles) occur over time. Mutations might occur with metronome-like regularity, showing little variation in the time that elapses between each mutation event. If substitution is a stochastic process, an alternative view is that the time that elapses between substitutions is a random variable. The Poisson distribution is a commonly used distribution to model the number of events that occur in a given time interval, so the bottom view is often called the Poisson molecular clock. Note that in both cases the number of substitutions and time elapsed is the same so that the average substitution rate is identical.

Figure 8.17   An illustration of the history of two DNA sequences that might be sampled from two species in the present time to estimate the rate of substitutions. The history is like a water pipe in an upside-down Y shape. The tube at the top contains the total population of lineages in the ancestral species, eventually splitting into populations of lineages that compose two species. The time when two lineages diverged from a common ancestor is not necessarily the same as the time of speciation. Therefore, a population process governing polymorphism operates for T generations in the ancestral species while a divergence process operates for t generations in the diverged species. The polymorphism process initially dictates the number of nucleotide changes between two sequences. Later, the divergence process dictates the number of nucleotide changes between two sequences. In two DNA sequences sampled in the present it is impossible to distinguish which process has caused the nucleotide changes observed.

Figure 8.18   Patterns of nucleotide changes that are possible when comparing DNA (or amino acid) sequences from two lineages and an outgroup. The letters i, j, and k are used to represent the identity of the nucleotide found at the same nucleotide site in each of the three sequences. For example, iij indicates that the first two lineages have an identical base pair and the third lineage has a different base pair. Tajima's 1*D* relative rate test utilizes substitutions that can be unambiguously assigned to one lineage (iji and ijj). If rates of substitution are identical for lineage 1 and 2, $E(n_{iji}) = E(n_{ijj})$. The lineages where the substitution took place are ambiguous for the patterns jji and ijk. The pattern iii indicates identical nucleotides in all three sequences and therefore no substitution events.

Figure 8.19   Estimates of the scaled mutation rate θ are estimated differently using nucleotide diversity ($\hat{\theta}_\pi$) and the number of segregating sites ($\hat{\theta}_S$) depending on the location of mutations in a genealogy. Each mutation makes a single segregating site under the assumptions of the infinite alleles model no matter where it occurs. However, mutations on internal branches will appear in multiple pairwise comparisons and cause π to be larger (a). In contrast, mutations that occur on external branches (b) that cause a nucleotide change in only a single lineage contribute less to π. Each mutation is counted four times (d_{13}, d_{23}, d_{14}, and d_{24}) in (a) but three times (d_{12}, d_{23}, and d_{24}) in (b) when computing π.

Figure 8.20   Differences in the shape of genealogies are the basis of Tajima's *D* test. In the standard coalescent model of genealogical branching the probability of coalescence is constant per lineage over time. The standard coalescent therefore gives expected branch lengths when all alleles are selectively neutral and the effective population size is constant (center). Changes in the effective population size over time (population growth, population bottlenecks) change the probability of coalescence over time as well. Natural selection also alters the probability of coalescence based on the fitness of alleles each lineage bears. Changes in the effective population size and natural selection alter the expected time to coalescence and therefore the expected branch lengths in a genealogical tree. If the chance of coalescence is greater in the present than in the past (right), most coalescent events occur near the present and internal branches are long in comparison with external branches. If the chance of coalescence is smaller in the present than in the past (left), most coalescent events occurred in the past and external branches are long in comparison with internal branches. Since the chance of a mutation is constant over time, lineages with longer branches are expected to experience more mutations.

Figure 8.21   The basis of the mismatch distribution. (a) A neutral genealogy that bears multiple mutation events. Each mutation event is represented by a circle and the number of the random nucleotide site that mutated assuming the infinite sites mutation model. The six lineages in the present can be separated into two groups (called A and B) based on their ancestral lineage when there were only two lineages in the population. (b) The DNA sequences for each lineage are shown based on the 30 base-pair sequence assigned to the most recent common ancestor (MRCA) with mutations shown in lower-case letters. (c) The number of nucleotide sites that are different or mismatched between pairs of DNA sequences. (d) The mismatch distribution shown is a histogram of the mismatches for the 15 pairs of DNA sequences compared. Neutral genealogies from populations with constant N_e through time tend to show bimodal mismatch distributions. The cluster of observations with few mismatches results from sequence comparisons between recently related lineages (comparisons within group A or group B). In contrast, sequences from distantly related lineages that do not share the same ancestor when $k = 2$ (comparisons between groups A and B) tend to have more mismatches.

Figure 8.22   The impact of natural selection on an advantageous mutation as well as on associated nucleotide sites, often called a selective sweep. Imagine a single population that contains five distinct DNA sequences without recombination because reproduction is clonal. Each DNA sequence is distinguished by a number of neutral mutations (blue circles) and has a frequency given by the histogram on the left. Initially, the population has polymorphism since the population is composed of intermediate frequencies of each DNA sequence. At time 0, the third DNA sequence experiences a mutation that is strongly advantageous, indicated by the star. Natural selection acts to increase the frequency of the advantageous mutation over time, until the population approaches fixation for the third DNA sequence. Once selection has swept the advantageous mutation to near fixation the population has very little polymorphism. This is because only those original neutral mutations that were linked to the advantageous allele on the same DNA sequence remain in the population. Thus, positive selection on one site also sweeps away polymorphism at linked nucleotide sites if gametic disequilibrium is maintained. The figure assumes that positive natural selection is strong and increases the frequency of the third DNA sequence rapidly such that no new mutations appear in the population.

Figure 8.23   Plots of nucleotide diversity within *D. melanogaster* populations (a) and divergence between *D. melanogaster* and *D. simulans* (b) by the coefficient of exchange, a measure of the recombination rate, for numerous loci. The nucleotide diversity at a locus decreases along with the recombination rate (a). The correlation of polymorphism and the recombination rate at a locus could be explained by neutral theory if loci with lower recombination rates also happen to have lower mutation rates. Under this neutral hypothesis, divergence rates would also be correlated with recombination rates since both polymorphism and divergence increase as the mutation rate increases. An alternative explanation for the correlation of recombination and polymorphism is the action of natural selection on beneficial mutations that has caused hitch-hiking and a reduction of polymorphism due to selective sweeps. Divergence rates at a subset of the loci (b) suggest that divergence and recombination rates are independent, rejecting the neutral hypothesis for these data. Data from Begun and Aquadro (1992).

Chapter 9

Figure 9.1   Examples of continuous quantitative trait distributions. The top panels show the distributions of body length and weight in a sample of 3-year-old striped bass (*Morone saxatilis*). The bottom panels are pupal mass distributions for mosquitoes raised in laboratory or field conditions. Each panel gives the sample size (n), mean, variance, and coefficient of variation (CV) that quantify the phenotypic distribution. Striped bass data are from L. Pieper (unpublished results). Mosquito pupal mass data from Armbruster and Conn (2006) and P. Armbruster (unpublished results).

Figure 9.2   The phenotypic distribution for a trait determined by two Mendelian loci. This hypothetical phenotype might be something like flower color, ranging between nearly white and deep blue. The genotypes are those expected in a large number of progeny from a cross between two doubly heterozygous (AaBb) parents. Alternatively, the genotype frequencies are those expected in a large number of progeny from a parental population with Hardy–Weinberg genotype frequencies where mating is random and all allele frequencies are $1/2$. Each a or b allele in a genotype causes $1/4$ unit of pigment in the phenotype whereas each A or B allele in a genotype causes $2^1/4$ units of pigment in the phenotype. As the number of loci contributing to the trait increases, the expected frequency of any individual genotype decreases and the phenotypic distribution will become smoother. For this phenotypic distribution the mean = 5, the variance = 4, and the coefficient of variation (CV) = 40.

Figure 9.3   The effect of environmental variation on phenotypic variation. The phenotypic distribution on the left is produced by two Mendelian loci with all allele frequencies equal to $1/2$ as in Figure 9.2. If the environment causes some variation in the phenotype expressed by each genotype, then the distribution of phenotypes produced by polygenic variation becomes both smoother and wider. In this illustration, environmental variation causes 50% of the individuals of each genotype to randomly increase or decrease one unit in phenotypic value. Although the average effect of environmental variation here is a zero change in phenotypic value, the phenotypic variance increases.

Figure 9.4   Phenotypic distributions for a trait determined by two loci in populations with low additive genetic variation (V_A). Genetic variation in phenotypes is additive since each a or b allele in a genotype contributes $1/4$ unit of pigment while each A or B allele in a genotype contributes $2^1/4$ units of pigment irrespective of genotype. However, total genetic variation is relatively low since allele frequencies are near fixation and loss (the frequency of a and b alleles is 0.1 while the frequency of A and B alleles 0.9). Compare with Fig. 9.2 where allele frequencies are all equal to $1/2$ and allelic effects are additive.

Figure 9.5   Phenotypic distributions for a trait determined by two loci with either complete dominance or epistasis. (a) Due to dominance, the majority of phenotypes are 9 units of pigment and no individuals display phenotypes with 3 or 7 units of pigment. In this example of two diallelic loci, the frequency of heterozygotes is at a maximum because all allele frequencies are $1/2$. In general, total phenotypic variation in the population can be increased or decreased by dominance. (b) The bottom phenotypic distribution shows an example of epistasis identical to the two-locus system that determines coat color in Labrador retriever dogs. One dominant locus controls pigment color (BB and Bb genotypes have black coats and bb genotypes have brown coats) while a second completely dominant locus controls the presence or absence of pigment in hairs (AA and Aa genotypes have pigmented hair and aa genotypes have unpigmented hair). In this example of dominance by dominance epistasis, phenotypic expression of the genotype at the coat color locus depends on the genotype at the pigmentation locus. In both graphs, the genotype frequencies are those expected under Hardy–Weinberg assumptions where all allele frequencies are $1/2$. The mean and variance are based on a population of 1000 individuals.

Figure 9.6   Examples of phenotypic variation due to genetic (V_G), environmental (V_E), and genotype-by-environment ($V_{G \times E}$) causes shown in norm-of-reaction plots. In all graphs, the phenotypic values of four genotypes within each of two environments (here called A and B) are plotted. Lines connect the phenotypic values of one genotype measured in the two environments. (a) Genotypic variation, where the four genotypes have different phenotypic values but the phenotypic value of each genotype does not change between environments. (b) Environmental variation: all genotypes have identical phenotypes (lines are staggered so that each can be seen) but the phenotype changes between environments. (c) Both genotypic and environmental variation: genotypes differ in phenotype and genotypes have different phenotypes in the different environments. (d) and (e) Genotype-by-environment interaction, with genotypes differing in the phenotypic value expressed in two or more environments. One type of genotype-by-environment interaction is characterized by lines connecting the genotypes that are not parallel (d), leading to changes in the phenotypic variance. In a second type of genotype-by-environment interaction, the rank order of phenotypic values exhibited by genotypes changes across environments and leads to crossing lines in norm-of-reaction plots (e).

Figure 9.7   The longest stem and number of stems phenotypes for seven *Achillea* genotypes originally sampled at Aspen Valley, California, cloned from vegetative clippings, and then transplanted at three elevations. *Achillea* phenotypes show that genetic (V_G), environmental (V_E), and genotype-by-environment ($V_{G \times E}$) interaction contribute to total phenotypic variation (V_P). Data are from Table 11 and the photograph from Figure 17 of Clausen et al. (1948). Used with permission of Carnegie Institution of Washington.

Figure 9.8   Broad-sense and narrow-sense heritabilities for five blood-pressure-related quantitative traits in humans. Each pie chart divides the total phenotypic variation (V_P) into its causal components of dominance (V_D) and additive (V_A) genotypic variance, as well as environmental variance (V_E). These heritabilities were estimated in a small population of Hutterites, a self-reliant, communal group of Anabaptists that traces its origins to followers of Jakob Hutter who fled Austria in the sixteenth century to escape religious persecution. Today, Hutterites live in Canada and North America. The 806 individuals in this study are descendants of 64 ancestors so that many individuals have a non-zero probability of sharing a genotype that is identical by descent because their parents are distantly related. This improves the precision of estimates of dominance variance. Estimates from Abney et al. (2001).

Figure 9.9   General types of natural selection on quantitative traits. Directional selection occurs when phenotypic values at the upper or lower end of the distribution have the highest fitness. Stabilizing selection occurs when intermediate trait values have the highest fitness. Disruptive selection occurs when trait values at the edges of the phenotypic distribution have the highest fitness. Response to directional selection increases or decreases the population mean phenotype. Response to stabilizing or disruptive selection does not change the mean but decreases or increases the variance of the phenotypic distribution.

Figure 9.10   A hypothetical example of directional selection and response to selection. In the parental generation those individuals with a phenotypic value of 12.0 or greater are allowed to mate. The selection differential is $s = 12.5 - 10.0 = 2.5$. Random mating among this subset of the parental population produces a distribution of progeny phenotypic values with a mean value of 11.0. The response to selection is $R = 11.0 - 10.0 = 1.0$. The realized heritability is therefore $h^2 = 1.0/2.5 = 0.40$. Response to selection is proportionate to the degree to which parental phenotypic values are caused by the phenotypic effects of alleles. Parental phenotypic values being caused by genotypes or the environment do not lead to response to selection since these causes of phenotypic value are not inherited by offspring.

Figure 9.11   Parent–offspring regressions used to estimate heritability (h^2) under the assumption of strict additivity (a) or complete dominance (b). The resemblance or covariance between the mid-parent phenotypic value and the mean progeny phenotypic value is greater on average with additivity (slope = 1.0) than with dominance (slope = 0.667). The slope of the regression line is equal to the heritability since it measures resemblance of parents and offspring. The mid-parent phenotypic value is the average phenotypic value of two parents that mate and produce progeny. Phenotypic values are given in parentheses next to each genotype. Each dot represents the mean phenotypic value of many progeny that result from a given parental mating. Both panels assume Hardy–Weinberg expected genotype frequencies in the parental populations, allele frequencies of $p = q = 0.5$, and that $V_E = 0$.

Figure 9.12   Simulations of directional selection on a quantitative trait with strictly additive genetic variance under three scenarios (parts (a), (b), and (c)) for the genetic architecture of the trait. From top to bottom in each part, the graphs show the plus allele frequencies (in b and c 10 loci are shown of the total 100 loci that influence the trait), the genotypic and phenotypic mean values for the trait, genotypic and phenotypic trait variance (V_P and V_G), and the narrow-sense heritability. In (a), there are 10 loci each with equal and large effects on the trait (10% of V_G). In (b), 100 loci have equally small effects (each 1% of V_G) on the trait. In (c), there are two loci with large effects (20% of V_G each) and 98 loci with small effects, as well as recurrent mutation between plus and minus alleles ($\mu = 0.001$ mutations gamete^{-1} generation^{-1}). In (c), the initial allele frequencies for the two loci of large effect (dashed allele frequency lines) and eight loci of small effect are initially 0.1 while the remaining 90 loci are fixed for the minus allele. A selection plateau is reached in (a) when all of the loci causing genotypic variation in the trait reach fixation (V_G and h^2 drop to 0). Simulations in (b) and (c) never reach selection plateaus nor exhaust all trait variation because at least some loci that influence trait variation remain segregating. Genetic variation is maintained in (b) because selection in finite populations is not able to fix alleles at all loci with small effects. Even though the two loci with major effects fix rapidly in (c), recurrent mutation at the many loci with small effects maintains some additive genetic variation for continued response to selection. Selection was accomplished by forming the next generation with the 50 individuals with the largest phenotypic values. In all simulations the maximum genotypic value was 10.0, $V_E = 0.1$, and the truncation point for natural selection each generation was the 50th percentile of phenotypic value. QTL, quantitative trait loci.

Figure 9.13   Phenotypic means for oil and protein content for high and low selected lines of the Illinois Long-Term Selection experiment initiated in 1896. Response to selection has been nearly linear over time, consistent with additive genetic variation that is caused by a relatively large number of loci each with a small effect on the phenotype. The low-oil-selected line was discontinued in generation 89 because oil content was too low to be measured reliably and plants had poor viability. Data kindly provided by J.W. Dudley and S.P. Moose.

Figure 9.14   Long-term selection for muscle mass in mice (measured as protein content per individual) over 70 generations. (a) The phenotypic mean over time, with a pronounced asymptote, or selection plateau, that indicates a diminishing response to selection over time. The total phenotypic variance, the additive genetic variance, and the realized heritability are shown in (b). Even though the selection differential is constant over time, the heritability declines steadily, as expected for a quantitative trait where genetic variation is caused by relatively few loci. The dip in protein content during generations 18–20 was an artifact due to an environmental effect. While the phenotypic variance increases over the experiment (b), this is caused largely by the increase in the phenotypic mean (the coefficient of variation of V_P stays nearly constant). Redrawn from Bünger, L., Renne, U., Dietl, G., and Kuhla, S. (1998) Long-term selection for protein amount over 70 generations in mice. *Genetical Research* 72: 93–109. Reprinted with the permission of Cambridge University Press.

Figure 9.15   The F2 or recombinant inbred line design for QTL mapping assuming one QTL (Q) and one genetic marker locus (M). The top phenotypic distribution represents the variance in value in the population that is subject to QTL mapping. Individuals at the edges of this phenotypic distribution are then sampled to start lines to be inbred for five (self-fertilization) to 10 (sib mating) generations to achieve high homozygosity. Individuals are then sampled from inbred lines with diverged phenotypes to form the P1 generation. The progeny of the P1 individuals are all double heterozygotes and form the F1 generation. Gamete frequencies produced by F1 individuals depend on recombination rates between loci. Random mating among F1 individuals produces the F2 individuals which are genotyped for the marker loci and measured for their phenotypic values.

Figure 9.16   Interval mapping utilizes two marker loci (A and B) that sit on either side of a QTL. An individual with the genotype shown can produce two types of gametes if there is no recombination, four type of gametes if there is a single recombination event, and two types of gametes from a double recombination event. The expected gamete frequencies are a function of the recombination rates.

Figure 9.17   The difference in phenotypic mean values for each of 17 genetic marker loci versus the position of each marker locus on a chromosome. In this hypothetical example, there are two genetic markers that lie closest to QTLs indicated by two peak values for marker-class mean differences. The phenotypes of the homozygote classes for the marker loci at 33, 41, and 85 map units (or centimorgans, cM) differ by over one phenotypic standard deviation. The QTL near 33 cM increases the mean value while the QTL near 85 cM decreases the mean value. Marker-class mean value differences greater than ±1 phenotypic standard deviation (blue dashed lines) are considered statistically meaningful in this example. Marker-class mean differences smaller than one standard deviation could be different due to chance alone. The close proximity of these two QTLs with opposite effects on the trait would lead to reduced estimates of QTL effects at all marker loci. This occurs because each marker locus is partly linked to a QTL of both positive and negative effect, so the perceived marker class mean difference is the net effect of the two QTLs. Genetic

marker locus positions are established by observed recombination rates. One map unit or centimorgan distance along a chromosome is equal to a 1% recombination rate. LOD score, log of odds score.

Figure 9.18   Distributions of estimated QTL effect sizes from several animal species. The distributions show the difference in phenotypic values for alternative homozygous genotypes (or the equivalent of $2a$) at a QTL in units of phenotypic standard deviations. QTLs with larger effect sizes explain a large proportion of the additive variance in a quantitative trait. The distribution in (a) represents QTLs identified from multiple phenotypes in dairy cattle while the distribution in (b) represents QTLs identified in multiple phenotypes in pigs. The distribution in (c) shows QTLs located on the third chromosome that influence sternopleural bristle number in *D. melanogaster*. Data in (a) and (b) from Hayes and Goddard (2001); data in (c) from Shrimpton and Robertson (1988).

Figure 9.19   The genotypes and distribution of phenotypic values for a trait caused by three codominant diallelic loci in a randomly mating population where all alleles at each locus have a frequency of $1/2$.

Chapter 10

Figure 10.1   The genotypic scale of measurement for quantitative traits. The variables $+a$ and $-a$ define genotypic values of the A_1A_1 and A_2A_2 homozygotes, respectively. The genotypic value of the heterozygote is defined as d and is measured relative the midpoint ($d = A_1A_2$ genotypic value minus the midpoint). The degree of dominance is expressed as the ratio d/a. These genotypic measurements are illustrated for the *IGF*1 gene in dogs, which is a major gene contributing to body-size differences. Since the degree of dominance is unknown for *IGF*1 genotypes, the genotypic value of the heterozygote is hypothetical.

Figure 10.2   The derivation of the average change in value caused by replacing one A_2 allele with an A_1 allele in those genotypes sampled at random from the population that contain at least one A_2 allele. The mean change in value caused by an allelic replacement forms the basis of the average effect once it is multiplied by the frequency of the allele that is replaced.

Figure 10.3   Illustration of dominance deviation for the *IGF*1 gene in dogs. The dominance deviation is the difference between the genotypic value (measured relative to the population mean, M) and the breeding value. The dominance deviation is a consequence of the heterozygote genotypic value not falling at the midpoint. Panels (a) and (b) correspond to cases (c) and (d), respectively, in Tables 10.3, 10.5, and 10.7.

Figure 10.4   The relationship between average effect of an allele replacement, genotypic values, breeding values, and dominance deviations. The solid line represents the least-squares regression of genotypic value on number of A_1 alleles in a genotype in a population of 100 individuals with Hardy–Weinberg genotype frequencies. The slope of the line is equal to the effect of an allelic replacement or α. There is no dominance in (a) and partial dominance in (b). The breeding value is equal to the genotypic value adjusted by the effects of dominance on the progeny population mean as measured by the dominance deviation. All values are deviations from the population mean.

Figure 10.5   The additive (V_A) and dominance (V_D) components of the total genotypic variance (V_G). In (a) there is no dominance ($d = 0.0$). Panel (b) has partial dominance ($d = 5.25$) and panel (c) has complete dominance ($d = 10.5$). In (b) and (c) the dominance component of the total genotypic variance is greatest when allele frequencies are equal because the frequency of heterozygotes is greatest. In all cases, the value of $+a = 10.5$.

Figure 10.6   The expected covariance in genotypic values for relatives based on probabilities that individuals share alleles (r) and genotypes (u) that are identical by descent. (a) The pedigree for the general case. (b) Half siblings (half brothers and sisters) share parent B in common and are thus unilineal relatives. (c) Full siblings (full brothers and sisters) share both parents in common and are thus bilineal relatives. In general, bilineal relatives include dominance components in their expected covariances.

Chapter 11

Figure 11.1   The caricatures of the mutation fitness spectrum drawn to illustrate the extreme views of pan-selectionism and pan-neutralism. (a) Under a pan-neutralist view almost all mutations have little or no impact on fitness and so are selectively neutral. (b) Under a pan-selectionist view there would be almost no mutations that are selectively neutral, many mutations that are selected against, and a relatively high frequency of beneficial mutations. Neither of these extremes is supported by most observations, except perhaps in isolated cases. This illustration is inspired by figures in Turner (1992) and Crow (1972).

Figure 11.2   Sewall Wright's original adaptive landscape diagram. The high-fitness points on the surface are indicated by + while the low fitness points are indicated by −. Original caption: "Diagrammatic representation of the field of gene combinations in two dimensions instead of many thousands. Dotted lines represent contours with respect to adaptiveness." From Wright (1932); reproduced with permission.

Figure 11.3   Wright's schematic representation of the simultaneous action of multiple population genetic processes leading to equilibrium distributions of allele frequencies. Each distribution represents the allele frequencies of many replicate populations or an ensemble distribution (in (A) numerous populations that make up a species are independent while in (B) subpopulations are interdependent in an island model). The magnitude and direction of the effects of a process on the distribution of allele frequencies is indicated by arrows bearing letters. Solid arrows indicate stronger processes and dashed arrows indicate weaker processes. For example, in the left-most distribution of (B) strong migration relative to drift maintains subpopulation allele frequencies with little divergence while weak genetic leads to a modest spread of subpopulation allele frequencies around the average allele frequency of the total population. In all panels, the frequency of the wild-type allele (x) is given on the x axis and the frequency of populations with a given allele frequency is given on the y axis. The probability on the y axis is given by the equation at the top of (A) and (B) for the allele frequency on the x axis given values for the population parameters. Original caption: "Random variability of a gene frequency under various specified conditions." From Wright (1932); reproduced with permission.

Figure 11.4   Wright's representation of the action of drift/mutation balance (dictated by the magnitude of 4NU), drift/selection balance (dictated by the magnitude of 4NS), and drift/migration balance in the island model (dictated by the magnitude of 4nm). Wright's parameters are N for effective population size, U for mutation rate, S for the selection coefficient in directional selection, and 4nm for the effective migration rate in the infinite island model. The word "inbreeding" is used in the population sense where finite population size leads to genetic drift. Original caption: "Field of gene combinations occupied by a population within the general field of possible combinations. Type of history under specified conditions indicated by relation to initial field (heavy broken contour) and arrow." From Wright (1932); reproduced with permission.

Appendix

Figure A.1   The frequency distribution of hypothetical weights for 200 jelly beans. The mean is 5.06 and the variance is 1.06.
Figure A.2   An abstract representation of 16 mouse populations. The total population (the entire rectangle) is composed of a series of smaller, discrete populations (the individual ellipses), sometimes called subpopulations. Each subpopulation has its own frequency of the A allele, indicated by $\hat{p}_i$. The entire population also has an exact value for the frequency of the A allele, which can be estimated by sampling some subpopulations and then taking the average and variance of the $\hat{p}_i$ values. Here eight subpopulations are sampled to estimate the allele frequency in the total population as $\bar{p} = 0.4976$, the standard deviation (SD) = 0.0272, and the standard error (SE) of the mean = 0.0096.
Figure A.3   Two frequency distributions of 100 data points each with nearly identical means (0.498 on the left and 0.506 on the right) but different degrees of variance among the observations (the variance is 0.0293 on the left and 0.0025 on the right). The lines at the top indicate the position of the mean (vertical line) and two standard deviations on either side of the mean (horizontal dashed line) and two standard errors on either side of the mean (horizontal solid line). Like allele frequencies, each distribution is on the interval of 0 to 1.
Figure A.4   Examples of joint distributions between two variables, x and y. The covariance between x and y is zero in (a), positive in (b), and negative in (c). In all three panels the variance of x remains constant at 0.0912. Trends between the variables can be expressed as correlation coefficients (ρ) or as the slopes of the least-squares regression of y on x (b).

Plate

Plate 2.10   Corn cobs demonstrating yellow and purple seeds that are either wrinkled or smooth.
Plate 3.1   Beakers filled with micro-centrifuge tubes can be used to simulate the process of sampling and genetic drift.
Plate 3.13   Probability densities of allele frequency for many replicate populations predicted using the diffusion equation. The initial allele frequency is 0.5 on the left and 0.1 on the right. Each curve represents the probability that a single population would have a given allele frequency after some interval of time has passed. The area under each curve is the proportion of alleles that are not fixed. Time is scaled in multiples of the effective population size, N. Both small and large populations have identically shaped distributions, although small populations reach fixation and loss in less time than large populations. The populations that have reached fixation or loss are not shown for each curve.
Plate 4.2   The plant *Linanthus parryae*, or desert snow, is found in the Mojave Desert regions of California. *L. parryae* can literally cover thousands of hectares of desert during years with rainfall sufficient to allow widespread germination of dormant seeds present in the soil. This tiny plant has either blue or white flowers. In some locations most plants have blue or most have white flowers whereas in other locations more equal frequencies of the two flower colors are found. Reproduced with permission of Barbara J. Collins.