æ

3

THE GENETIC EFFECTIVE SIZE OF A POPULATION

19 May 2014

Throughout the preceding chapter, we assumed a population with an idealized set of Wright-Fisher features, including random mating within a homogeneous group of monoecious, self-compatible individuals with equal expected family sizes, discrete generations, and an absence of density fluctuations. Because almost all populations deviate from this ideal structure in one or more ways, and often substantially so, the relevance of the resultant theory might then seem in doubt. In fact, much of the theory of inbreeding and random genetic drift can be generalized to other types of population structures to a good approximation in a relatively simple manner. To accomplish this task, instead of relying on the total number of adult individuals (N) as a measure of population size, we construct a surrogate index that takes into account the deviations from the ideal model from a genetic perspective. Following the influential work of Wright (1931, 1938, 1939), such an index has become widely known as N_e, the **effective population size**. With such a reparameterization, essentially all of the results in the preceding chapter hold when N_e is substituted for N. Buri's experiment is a case in point – a set of populations with actual size $N = 16$ exhibited allele-frequency dynamics closely approximated by the expectations for an idealized Wright-Fisher population with an effective size of just ten individuals.

Because of its central role in defining levels of variation within populations, the rate of divergence among populations, and the efficiency of natural selection, N_e is one of the most important parameters in population genetics. Although N_e is not as easily measured as the total (census) population size, as will be seen in the following sections, it is at least in part defined by observable demographic and mating-system properties of populations (Latter 1959; Lande and Barrowclough 1987; Crow and Denniston 1988). A central goal of this chapter is to illustrate that nearly every violation of the assumptions underlying the Wright-Fisher model leads to a reduction in N_e relative to N, thereby implicating a stronger role for genetic drift in evolution than might be surmised from estimates of total population sizes. We will progressively consider aspects of the mating system (including the possibility of self-fertilization, variation in the sex ratio, and variance in family size), age structure, temporal variation, and spatial structure.

In addition, we will see that the effective size of a population is often strongly influenced by the structural aspects of genomes, independent of population demographic features. The physical linkage of genes on chromosomes ties their mutual fates together, creating stochastic fluctuations of allele frequencies out of chance associations with other loci under selection. Tight linkage to a deleterious mutation can result in loss of an otherwise neutral or beneficial allele from the population, whereas tight linkage to a beneficial mutation can result in hitchhiking to fixation. This influence of regional recombination rates and chromosomally restricted patches of selection has the counterintuitive effect of creating variation in N_e among different loci in the same population.

This chapter focuses entirely on the concept of N_e from a theoretical perspective, with a goal of providing the reader with a qualitative understanding of the mechanistic determinants of this key population-genetic parameter. Chapter 4 will provide an overview of methods for the estimation of N_e using molecular markers in natural populations, which require no knowledge of a population's demographic features. A more thorough review of a number of the topics we touch upon can be found in Caballero (1994), and although we focus on autosomal loci, the general principles are readily extended to sex-linked loci (Caballero 1995; Nagylaki 1995; Wang 1997; Charlesworth 2001).

GENERAL CONSIDERATIONS

An appreciation for the concept of effective population size can be gained by recalling that all members of an ideal monoecious population contribute equally to the total gamete pool, with each successful gamete uniting randomly with another gamete derived from the total population of N individuals. Under these conditions, the probability that two uniting gametes are derived from the same parent is simply $P = 1/N$. However, many factors, including self-incompatibility, limited dispersal, differential productivity of gametes, and selection, can cause P to deviate from $1/N$. To account for the joint influence of all of these factors and many others, we define P to be the reciprocal of the effective population size.

Using this definition of N_e, many of the results in the previous chapter can be generalized, at least to a first-order approximation. Consider, for example, the expected dynamics of the inbreeding coefficient for a diploid, monoecious population. The probability that two uniting gametes are derived from the same parent is now $1/N_e$, in which case there is a 50% chance that they each carry copies of the same gene (i.e., they are identical by descent, one generation removed) and a 50% chance that they carry copies of different genes. In the latter case, the uniting genes may still be identical by descent with probability f_{t-1} from previous inbreeding. Finally, there is a $1 - (1/N_e)$ probability that the uniting gametes are derived from different parents, in which case there is again a probability f_{t-1} that they are identical by descent from previous inbreeding. Summing up the three ways by which identity-

by-descent can arise between uniting gametes,

$$\begin{aligned} f_t &= \left(\frac{1}{N_e}\right)\left(\frac{1}{2}\right) + \left(\frac{1}{N_e}\right)\left(\frac{1}{2}\right) f_{t-1} + \left(1 - \frac{1}{N_e}\right) f_{t-1} \\ &= \frac{1}{2N_e} + \left(1 - \frac{1}{2N_e}\right) f_{t-1} \end{aligned} \tag{3.1}$$

This expression is identical in form to Equation 2.3 with N_e replacing N.

Under the above interpretation, N_e is the size of an ideal population that would exhibit the same amount of inbreeding as the population under consideration. Defined in this way, N_e is the **inbreeding effective size**. There are, however, numerous additional ways to define the effective size of a population. One can, for example, define the **variance effective size** as the N that when applied to Equation 2.14a yields the variance in allele-frequency change exhibited by a nonideal population. Crow (1954) has emphasized that the inbreeding effective size is most closely related to the number of parents (or the number of grandparents if selfing does not occur) because it is based upon the probability of uniting gametes coming from the same ancestor. In contrast, the variance effective size, which is associated with allele-frequency drift resulting from gamete sampling, is primarily a function of the number of offspring produced. Thus, in an expanding or declining population, the rates of inbreeding and allele-frequency drift can differ in any particular generation (Templeton 2006). However, for populations with stable size, inbreeding and variance effective sizes are generally equivalent, and in the long run this is even true for fluctuating populations because both measures depend on the same sequence of adult population sizes (Caballero 1994; Whitlock and Barton 1997). More discourse on these issues may be found in Kimura and Crow (1963b), Crow and Kimura (1970), Crow and Morton (1955), Crow and Denniston (1988), and Caballero (1994). Unless stated otherwise, the following discussion will assume populations of constant size.

Random genetic drift is a fundamental evolutionary property of all populations resulting from two stochastic processes. First, in sexual species, segregation during meiosis leads to the random transmission of alleles from heterozygous parents, as there is a 50% chance for each alternative allele to be inherited by any given offspring. Second, variation in family size encourages lineages of alleles that happen to be contained within large families to expand at the expense of others. As we will see in the following sections, such variation arises by numerous mechanisms, including simple sampling of gametes, spatial population structure, variation among the sexes, and trapping in various genetic backgrounds by linkage.

MONOECY

To illustrate the mathematical approach to deriving expressions for N_e, we first generalize the concept of a monoecious, self-compatible population to allow for arbitrary gamete production by different individuals. Let k_i be the number of gametes that the ith parent contributes to offspring that survive to maturity, μ_k and σ_k^2 be the mean and variance of successful gamete production per individual, and N_{t-1} be the number of reproducing parents. Assuming that mating is random and **isogamous** (so that there is no distinction between male and female gametes), there

are $k_i(k_i-1)$ ways in which the gametes of parent i can unite with each other, and summing over all parents, $\sum_{i=1}^{N_{t-1}} k_i(k_i-1)$ total ways by which gametes can unite by self-fertilization. Because a total of $N_{t-1}\mu_k$ successful gametes are produced, the expected fraction of zygotes derived from the same parent is

$$P_t = \frac{1}{N_e} = \frac{\sum_{i=1}^{N_{t-1}} k_i(k_i-1)}{N_{t-1}\mu_k(N_{t-1}\mu_k-1)} \tag{3.2}$$

This expression can be simplified greatly by noting that $\sum_{i=1}^{N_{t-1}} k_i(k_i-1)/N_{t-1} = E(k^2) - \mu_k = \sigma_k^2 + \mu_k(\mu_k-1)$, and that because all zygotes are derived from two gametes, $N_{t-1}\mu_k = 2N_t$. Substituting into Equation 3.2 and inverting,

$$N_e = \frac{2N_t - 1}{(\sigma_k^2/\mu_k) + \mu_k - 1} \tag{3.3}$$

This shows that for a randomly mating monoecious population with discrete generations, the effective population size is a function of three measurable quantities: the actual population size (N_t), and the mean (μ_k) and variance (σ_k^2) of successful gamete production. All other things being equal, variance in gamete production causes a reduction in N_e, as this inflates the representation of a fraction of the population in the descendant gene pool. Because such variation is expected to be the rule in natural populations occupying environments that are heterogeneous with respect to resource availability, on this basis alone, we can generally expect N_e to be less than the number of reproductive adults.

The above expression simplifies greatly under a number of conditions. For example, populations that are stable in size have on average two successful gametes per parent ($\mu_k = 2$), and

$$N_e = \frac{4N-2}{\sigma_k^2 + 2} \tag{3.4}$$

If we further assume that each parent produces the same number of *potential* gametes (returning us to the ideal random-mating population), an explicit statement can also be made about σ_k^2. In this case, for any particular draw from the gamete pool, the variance in the number of gametes derived from a particular parent (0 or 1) is $(1/N)[1-(1/N)]$ (from the properties of a binomial distribution), and because a total of $2N$ gametes are drawn, $\sigma_k^2 = 2[1-(1/N)]$. Note that this result is very close to the Poisson expectation of $\sigma_k^2 = 2$ (the variance of a Poisson equals the mean), the slight deviation resulting because we assume a fixed population size. Substitution into Equation 3.4 yields $N_e = N$, showing that the inbreeding effective size of an ideal random-mating population is indeed equal to the number of reproductive adults in the previous generation.

In contrast, in the opposite and extreme situation in which all parents produce *exactly* two progeny (such that $\sigma_k^2 = 0$ and $\mu_k = 2$), $N_e = 2N - 1 \simeq 2N$. This shows that the elimination of variance in family size results in an effective population size that is twice the actual number of breeding adults, a feature that is often exploited in breeding schemes to minimize the amount of inbreeding.

Example 3.1. Hedgecock (1994) suggested that many marine organisms with high fecundities and broadcast spawning may have effective population sizes that are orders of magnitude smaller than the absolute number of potential breeders. This situation can arise if vagaries in oceanographic conditions are such that only a small fraction of adults produce gametes at points in time and space that allow recruitment to the next generation. Suppose the total adult population size is N, whereas only N_b individuals contribute equally to the breeding pool. Such a situation is sustainable if reproductive adults can individually produce an average of $2N/N_b$ gametes (many marine species are capable of producing tens of thousands of gametes). Given such a situation, N_b individuals have expected family sizes of $2N/N_b$, whereas $(N - N_b)$ have zero expected reproductive success, which implies a variance in expected family size among all N individuals of $4[(N/N_b) - 1]$ (Hedrick 2005). Using the logic outlined in the text, the additional variance in reproductive success among spawning individuals resulting from random gamete sampling is equal to

$$\left\{\left(\frac{N_b}{N}\right)\cdot 2N\cdot\left(\frac{1}{N_b}\right)\left[1-\left(\frac{1}{N_b}\right)\right]\right\}+\left\{\left[1-\left(\frac{N_b}{N}\right)\right]\cdot 2N\cdot 0\right\}=2\left[1-\left(\frac{1}{N_b}\right)\right]$$

Summing these two sources of family-size variance gives $\sigma_k^2 = [(4N - 2)/N_b] - 2$, and substituting into Equation 3.4, we obtain $N_e = N_b$. Thus, provided a species has a high enough gamete production to generate N surviving progeny from a small number of adults, the effective population size can be only a tiny fraction of N.

There are a number of ways in which the breeding systems of a monoecious species can deviate from the assumptions made in the preceding derivations. Consider, for example, hermaphroditic species with self-incompatibility, in which case identity-by-descent for pairs of uniting gametes comes through grandparents rather than parents. If we now let k_i be the number of successful gametes for individual i in generation $t-2$, there are $2k_i(k_i-1)$ ways in which pairs of genes from i can unite through matings in the parental generation $t-1$ (the 2 arising because we assume that each individual can serve as a mother or father). Because there are $N_{t-2}\mu_k/2$ parents in generation $t-1$, there are $2(N_{t-2}\mu_k/2)[(N_{t-2}\mu_k/2)-1]$ ways of drawing different parents, and $4\cdot 2(N_{t-2}\mu_k/2)[(N_{t-2}\mu_k/2)-1]$ ways of drawing gene pairs (the 4 because each parent carries two genes). Therefore, the probability that a pair of genes in a generation-t individual are derived from the same grandparent is

$$P_t = \frac{1}{N_e} = \frac{\sum_{i=1}^{N_{t-2}} k_i(k_i-1)}{N_{t-2}\mu_k(N_{t-2}\mu_k-2)} \tag{3.5}$$

Employing the same substitutions used for Equation 3.3,

$$N_e = \frac{2(N_{t-1}-1)}{(\sigma_k^2/\mu_k)+\mu_k-1} \tag{3.6}$$

and for constant N (which implies $\mu_k = 2$),

$$N_e = \frac{4(N-1)}{\sigma_k^2+2} \tag{3.7}$$

For populations that are moderately large and stable in size, Equations 3.4 and 3.7 give essentially the same answer, $N_e \simeq 4N/(\sigma_k^2 + 2)$, demonstrating that the prohibition of selfing has a negligible influence on N_e unless the total population size is tiny. The reason for this is that under random mating the increment in inbreeding resulting from self-fertilization is a transient event that can be completely undone in the following generation.

A second potential complication is that in most hermaphroditic species, there is a distinction between male and female gametes (**anisogamy**), so that even with selfing only a fraction of potential gamete pairs are capable of spawning a successful zygote. When mating is random but selfing is prohibited, the effective population size is the same under isogamy and anisogamy, and Equation 3.6 still applies (Crow and Denniston 1988). However, with selfing permitted,

$$N_e = \frac{N_{t-1}}{(4\sigma_{o,p}/\mu_k^2) + 1} \tag{3.8}$$

where $\sigma_{o,p}$ is the covariance of the numbers of successful male (p, pollen) and female (o, ovule) gametes per parent (Crow and Denniston 1988). If $\sigma_{o,p}$ is positive, as might be expected in a spatially heterogeneous environment where some individuals acquire more resources than others, the effective population size will be less than the observed size. However, if $\sigma_{o,p}$ is negative, as might be expected when there is a tradeoff between male and female function, N_e can exceed N_{t-1}. This results because a negative covariance in male and female gamete production reduces the variance in family size.

Example 3.2. Heywood (1986) estimated that σ_k^2/μ_k^2 for *seed* production is on the order of 1 to 4 in a number of annual plants (including self-compatible species). Unfortunately, the value of σ_k^2 for *total* gamete production requires additional information on successful pollen production, which is extremely difficult to acquire due to problems in ascertaining paternity. For heuristic purposes, however, let us assume a stable monoecious population. This necessarily implies mean seed and pollen production are both equal to one, and $\mu_k = \mu_o + \mu_p = 2$, as each parent must produce two successful gametes (on average, one male and one female). We will also assume a three-fold higher standard deviation for successful pollen relative to seed production, so that $\sigma_p = 3\sigma_o$, and a perfect correlation between ovule and pollen production. Because the correlation between the number of female and male gametes produced per individual is defined to be $\sigma_{o,p}/(\sigma_o\sigma_p)$, the latter assumption implies $1 = \sigma_{o,p}/[\sigma_o \cdot 3\sigma_o]$. Assuming random mating, what is N_e? Substituting $\sigma_{o,p} = 3\sigma_o^2$ into Equation 3.8, we obtain $N_e = N/[(12\sigma_o^2/\mu_k^2) + 1]$. Thus, for σ_o^2/μ_k^2 in the range of 1 to 4, N_e is between 2% and 8% of the census number (N).

DIOECY

As in the case of monoecy with self-incompatibility, when the sexes are separate, inbreeding always needs to be defined with reference to the grandparental generation,

which is the earliest generation back to which the two genes of an individual can coalesce. Separate sexes also introduce the possibility of different levels of inbreeding through males and females, as might be expected, for example, in polygynous species in which most females mate with a relatively small segment of the male population.

If O is the individual of interest, with M and F being its mother and father, there are two ways by which O may derive two genes from the same grandparent: 1) M and F may share the same mother (with probability $1/N_{ef}$, where $1/N_{ef}$ is the effective number of females); or 2) M and F may share the same father (with probability $1/N_{em}$, where $1/N_{em}$ is the effective number of males). In either case, because both parents transmit to O a gene from the shared ancestor with probability 0.5, the probability that O inherits both genes from the shared grandparent is 1/4. Thus, the total probability that O inherits two genes from the same grandparent is

$$P = \frac{1}{N_e} = \frac{1}{4N_{em}} + \frac{1}{4N_{ef}} \tag{3.9}$$

What do we mean by the effective numbers of males and females? Assuming random mating (including no prohibition of mating between sibs), the effective number of each sex can be derived by the same method used to obtain Equation 3.3. Skipping the intermediate steps, we simply note that

$$N_{es} = \frac{\mu_{sk} N_{s,t-2} - 1}{(\sigma^2_{sk}/\mu_{sk}) + \mu_{sk} - 1} \tag{3.10}$$

where s denotes the sex (m or f), and μ_{sk} and σ^2_{sk} are the mean and variance of gamete production by sex s (Crow and Denniston 1988). Latter (1959) provides a more elaborate expression for N_{es} that explicitly accounts for the variance and covariance of male and female progeny production. Letting ϕ be the sex ratio (proportion of females),

$$N_{em} = \frac{4N_{m,t-2}}{2 + \sigma^2_{mm} + \frac{2(1-\phi)}{\phi}\sigma_{mm,mf} + \left(\frac{1-\phi}{\phi}\right)^2 \sigma^2_{mf}} \tag{3.11a}$$

$$N_{ef} = \frac{4N_{f,t-2}}{2 + \sigma^2_{ff} + \frac{2\phi}{1-\phi}\sigma_{fm,ff} + \left(\frac{\phi}{1-\phi}\right)^2 \sigma^2_{fm}} \tag{3.11b}$$

where for male parents, σ^2_{mm} is the variance of male progeny number, σ^2_{mf} is the variance of female progeny number, and $\sigma_{mm,mf}$ is the covariance of male and female progeny number, with similar definitions for female parents. There are a variety of situations in which these types of specifications may be useful. For example, if parents produce a fixed number of offspring, the covariances $\sigma_{mm,mf}$ and $\sigma_{fm,ff}$ between numbers of sons and daughters must be negative, whereas these terms can be positive if parents differ in the resources available for overall progeny production.

Further simplification of Equation 3.10 is possible when certain assumptions are met. Consider, for example, the case in which members of the same sex produce equal numbers of gametes so that the variation in family size is a simple consequence of the random union of gametes. It then follows from the development of

the monoecy model that $N_{em} = N_{m,t-1}$ and $N_{ef} = N_{f,t-1}$. Rearranging Equation 3.9,

$$N_e = \frac{4N_{m,t-1}N_{f,t-1}}{N_{m,t-1} + N_{f,t-1}} = 4\phi(1-\phi)N_{t-1} \tag{3.12}$$

In this case, N_e attains a maximum of N_{t-1} when the sex ratio is balanced ($\phi = 0.5$), but with skewed sex ratios, N_e is influenced much more strongly by the density of the rarer sex. For example, in a highly polygynous species, as $\phi \to 1$, $N_e \to 4(1-\phi)N_{t-1} \simeq 4N_{m,t-1}$, namely four times the number of males.

Finally, the results for dioecy can also be linked to those for the monoecy model in the following informative way. The mean gamete production for the whole population is $\mu_k = (1-\phi)\mu_{mk} + \phi\mu_{fk}$, or equivalently because all individuals have a father and a mother, $\mu_k = 2(1-\phi)\mu_{mk} = 2\phi\mu_{fk}$. The variance of gamete production across the entire population is $\sigma_k^2 = (1-\phi)\sigma_{mk}^2 + \phi\sigma_{fk}^2 + \phi(1-\phi)(\mu_{mk} - \mu_{fk})^2$. Using these expressions, Equation 3.9 is essentially equivalent to Equation 3.7 (Kimura and Crow 1963b), showing that the effective size of an ideal population with separate sexes is the same as that for a monoecious, self-incompatible population with the same population properties μ_k and σ_k^2.

In natural populations, where individuals inhabit different environments that influence the availability of resources and mates, it is likely that the variance in progeny production will exceed the mean (i.e., $\sigma_k^2 > \mu_k$), in which case N_e will be less than that predicted by Equation 3.12. For example, in a summary of data on lifetime reproductive success in female birds, Grant (1990) found that σ_{fk}^2/μ_{fk} ranged from 1.2 to 4.2. Assuming a stable population size ($\mu_{fk} = 2$) and substituting into Equation 3.10, the female effective population size for these species is found to be 40 to 90% of the actual number of females. Nonrandom variation in family sizes appears to be the rule even in laboratory populations. For example, caged populations of *Drosophila* typically exhibit effective sizes on the order of 10% of the census size of the adult population (Briscoe et al. 1992). Observations from natural populations of other animals suggest an average N_e/N ratio for single generations on the order of 0.7 (Crow and Morton 1955; Nunney and Elam 1994).

AGE STRUCTURE

Because the previous formulae were obtained under the assumption of discrete generations, they provide estimates of N_e for explicit generational intervals. Such expressions are reasonable for organisms such as annual plants (ignoring the problem of seed banks; Nunney 2002) or univoltine insects, but for species that reproduce at different ages, as is the case for most vertebrates and perennial plants, overlapping generations raise some additional complications. Nevertheless, as first pointed out by Hill (1972, 1979), there is a simple correspondence between the effective sizes of populations with and without age-structure.

In the previous formulations, N was the number of potential reproductive individuals entering the population in each generation. For age-structured populations, we must consider instead N_b, the total number of newborns entering the population during each unit of time, as well as the number of time units per generation. The

latter quantity, known as the **generation time** (T), is the average age of parents giving birth, which in turn is a function of the age-specific schedules of survival and reproduction. For an ideal monoecious population,

$$T = \frac{\sum_{i=1}^{\tau} i l_i b_i}{\sum_{i=1}^{\tau} l_i b_i} \tag{3.13}$$

where l_i is the probability of surviving to age i, b_i is the expected number of offspring produced by parents of age i, and τ is the maximum reproductive age. The quantity $l_i b_i$ denotes the expected number of births by an individual of age i, discounting for prior mortality. For a dioecious population, T is further complicated by the need to average over males (m) and females (f),

$$T = \frac{T_{mm} + T_{mf} + T_{fm} + T_{ff}}{4} \tag{3.14}$$

where T_{mf}, for example, is the average age of male parents of daughters. The average generation length is the natural time scale for the evolutionary analysis of age-structured populations. Letting $N = N_{eb}T$, with $N_{eb} = 4\phi_b(1-\phi_b)N_b$ being the effective size of the newborn age class and ϕ_b being the sex ratio of newborns, all of the preceding formulae for discrete-generations apply provided the structure and size of the population are stable. However, we are still left with the rather substantial problem of estimating σ_k^2, which now depends on variation in longevity as well as variation in fertility.

Felsenstein (1971), Johnson (1977), and Emigh and Pollak (1979) have shown how the variance in offspring production can be expressed in terms of the age-specific parameters l_i and b_i. Again making the assumption that the population is stable in size, sex ratio, and age composition, the effective size of an age-structured population with separate sexes is

$$N_e = \frac{N_{eb}T}{1 + (1-\phi_b)\sum_{i=1}^{\tau_f}\left(\frac{1}{l_{i+1}^f} - \frac{1}{l_i^f}\right)\left(\sum_f\right)^2 + \phi_b \sum_{i=1}^{\tau_m}\left(\frac{1}{l_{i+1}^m} - \frac{1}{l_i^m}\right)(\sum_m)^2} \tag{3.15}$$

where $(\sum_s)^2 = \left(\sum_{j\geq i+1}^{\tau_s} l_j^s b_j^s\right)^2$, τ_s is the maximum age of reproduction for sex s, and $s = m$ or f (Emigh and Pollak 1979). An analogous expression is available for monoecious populations (Felsenstein 1971). While the derivations underlying these expressions rely on the assumption that gametes are drawn randomly from the members within age classes, no assumptions are made with regard to the preference of matings between age classes.

Despite their complicated structure, demographic formulae such as Equation 3.15 are useful for analyzing the sensitivity of a population's effective size to modifications in the life-history schedule. Nevertheless, the Emigh-Pollak equation has some practical difficulties. First, it rests on the assumption of a stable population structure. Such situations are rare in nature because of temporal changes

in the environment. Johnson (1977) and Choy and Weir (1978) have derived dynamical equations to resolve these difficulties, and the entire subject is reviewed by Charlesworth (1994). Second, Equation 3.15 has been derived under the assumption that the age-specific mortality and birth rates of individuals are uncorrelated, i.e., that individuals with an elevated likelihood of survivorship do not have elevated or reduced birth rates. This will not be true for populations in which energetic tradeoffs exist between different life-history characters. The problem needs further investigation.

Substantial simplification of Equation 3.15 can be achieved under some conditions. For example, if year-to-year survival is age-independent, and if the mating system can be described in simple terms, N_e can be defined as a function of a small number of parameters, eliminating the need for refined age-specific schedules of survivorship and fecundity. Using this approach, Nunney (1993) concluded that N_e in animals with overlapping generations is typically on the order of $N/2$ to N, although his analysis ignores the important influence of variation in N across generations (as will be seen in a following section).

Example 3.3. While complete age-specific survivorship and reproductive schedules are available for the females of many natural populations, male promiscuity often imposes enormous practical difficulties for paternity ascertainment. Thus, the variance in male reproductive success is generally unknown. However, a long-term study on the behavior and demography of the red deer (*Cervus elaphus*) by Clutton-Brock et al. (1982) allows at least a crude estimate of N_e by use of the Emigh-Pollak equation, as shown in the table below. The study population was roughly constant in density for two decades, and observations on known individuals provide information on the age-specific rates of mortality and reproduction for both sexes. The sex ratio at birth (ϕ_b) averaged 0.43 over several years, so $N_{eb} = 0.98N_b$.

The age-specific survival rates, l_i, in the following table are extracted directly from Clutton-Brock et al. (1982), while b_i^f and b_i^m are estimated from behavioral and demographic observations of the authors and are adjusted downward to maintain a stable population size. Columns marked (1) and (2) are $[(1/l_{i+1}^s)-(1/l_i^s)]$ and $(\sum_{j>i+1}^{\tau_s} l_j^s b_j^s)^2$, and column (3) is the product of (1) and (2).

Age	Females					Males				
i	l_i^f	b_i^f	(1)	(2)	(3)	l_i^m	b_i^m	(1)	(2)	(3)
1	1.00	0.00	0.33	0.25	0.08	1.00	0.00	0.45	5.97	2.69
2	0.75	0.00	0.12	0.25	0.03	0.69	0.00	0.22	5.97	1.31
3	0.69	0.00	0.02	0.25	0.01	0.60	0.00	0.03	5.97	0.18
4	0.68	0.18	0.04	0.24	0.01	0.59	0.00	0.03	5.97	0.18
5	0.66	0.26	0.07	0.22	0.02	0.58	0.00	0.03	5.97	0.18
6	0.63	0.33	0.05	0.21	0.01	0.57	0.34	0.03	5.92	0.18
7	0.61	0.34	0.06	0.19	0.01	0.56	0.26	0.03	5.88	0.18
8	0.59	0.40	0.06	0.17	0.01	0.55	0.60	0.10	5.59	0.56
9	0.57	0.42	0.03	0.16	0.01	0.52	0.53	0.12	5.30	0.64
10	0.56	0.34	0.07	0.14	0.01	0.49	0.79	0.40	3.94	1.58
11	0.54	0.46	0.03	0.13	0.00	0.41	0.53	0.42	3.11	1.31

12	0.53	0.42	0.20	0.08	0.02	0.35	0.45	1.69	1.00	1.69
13	0.48	0.45	0.19	0.04	0.01	0.22	0.08	2.60	0.63	1.64
14	0.44	0.40	0.23	0.01	0.00	0.14	0.20	3.97	—	—
15	0.40	0.25	0.20	—	—	0.09	—	—	—	—
16	0.37	0.00	—	—	—	0.05	—	—	—	—
					0.23					12.32

The summations in the denominator of Equation 3.15 reflect the variation in lifetime reproductive success of females and males. As outlined in the table, these terms are equal to 0.23 and 12.32, respectively, indicating a great inequity between the reproductive properties of the sexes. This results because male red deer appropriate harems, and older males are much more successful at it than young ones. The few males that live to an old age may father up to two dozen offspring in their lifetimes, whereas males that die before the age of five ($\sim$ 40% of newborn males) have no reproductive success at all. On the other hand, almost all females reproduce to some degree once they have attained reproductive maturity.

Substituting the sums from the table into Equation 3.15, the effective population size is found to be $0.98N_bT/[1 + (1 - 0.43)(0.23) + 0.43(12.32)] = 0.15N_bT$. Thus, the effective size of this population is $\sim$ 15% of the number of offspring produced by the population/generation. The mean generation time through females and males is 9.47 and 9.18 years, so $T \simeq 9.32$, and the annual number of offspring produced by the population is $N_b \simeq 270$. Thus, $N_e \simeq 0.15 \times 270 \times 9.32 = 378$.

VARIABLE POPULATION SIZE

Most populations vary in density from generation to generation, often dramatically so, and this raises practical problems in the implementation of the previous theory. As noted in Equation 2.7, with variable population size, the expected loss of heterozygosity over t generations is no longer $[1 - (1/2N_e)]^t$ but a product of t terms, each incorporating the effective population size of a particular generation such that

$$H_t = H_0 \prod_{i=0}^{t-1} \left(1 - \frac{1}{2N_{e,i}}\right) \tag{3.16}$$

Thus, it is informative to evaluate the size of an ideal population of constant size with the same expected heterozygosity after t generations as a population with variable size over the same period. An approximate answer can be obtained by noting that with moderately large $N_{e,i}$, Equation 3.16 simplifies to

$$H_t \simeq H_0 \exp\left(-\sum_{i=0}^{t-1} \frac{1}{2N_{e,i}}\right) \tag{3.17}$$

which may be compared to

$$H_t \simeq H_0 e^{-t/2N_e}$$

for the ideal case of constant effective size. Equating the exponents of these two expressions,

$$N_e^* \simeq \frac{t}{1/N_{e,0} + 1/N_{e,1} + \cdots + 1/N_{e,t-1}} \tag{3.18}$$

Thus, the long-term effective size N_e^* is approximately equal to the harmonic mean of the generation-specific effective sizes. An asterisk is placed on N_e to remind the reader that the inbreeding projected by N_e^* strictly pertains to generation t. Other generations may exhibit more or less loss of variation than anticipated depending upon the actual pattern of temporal changes in $N_{e,i}$.

Example 3.4. To see how population bottlenecks have especially pronounced effects on N_e^*, consider a population whose effective size regularly fluctuates between 10 and 100. From Equation 3.18, $N_e^* = 2/(0.1 + 0.01) = 18.2$. Thus, the total loss of heterozygosity from this population every two generations is equivalent to that expected for an ideal random mating population with a constant effective size of 18, which is much closer to the expectation for a constant population size of 10 than 100. Frankham (1995) and Vucetich et al. (1997) have shown that N_e^* is frequently in the range 10 to 20% of N_e for a diversity of natural populations of animals. In the extreme case where the effective population size is effectively infinite for $t-1$ generations and N_e for one generation, $N_e^* = tN_e$.

PARTIAL INBREEDING

Although most of the previous formulations assumed a random union of gametes, the frequency of mating between relatives often exceeds that expected under random mating. Many plants, for example, produce a significant proportion of offspring by self-fertilization. If the total population size were infinite, a fixed proportion of matings between relatives would simply lead to an equilibrium condition wherein the production of new inbreeding each generation is balanced by the breakdown of old inbreeding through outcrossing (Wright 1951, 1969; Hedrick 1986; Hedrick and Cockerham 1986). However, such an equilibrium does not exist for finite populations, where allele frequencies are subject to random genetic drift. Here we consider the consequences of partial selfing in monoecious populations and of partial full-sib mating in populations with separate sexes, in both cases assuming an otherwise randomly mating population. Both subjects are evaluated in considerable detail by Caballero and Hill (1992) and Wang (1996).

Assuming a constant number of adults, for a population in which a random proportion β of progeny from each family is a product of self-fertilization,

$$N_e = \frac{2(2-\beta)N}{\sigma_k^2 + 2(1-\beta)} \tag{3.19}$$

(Crow and Denniston 1988; Caballero and Hill 1992; Wang 1996). Further assuming that the numbers of selfed and outcrossed progeny per parent are independent

Poisson variables, $\sigma_k^2 = 2(1+\beta)$, and Equation 3.19 reduces to

$$N_e = \frac{(2-\beta)N}{2} \tag{3.20}$$

If selfing is random, so that $\beta = 1/N$, then $N_e = N - 0.5$, a result that can also be obtained directly from Equation 3.3 by letting $\sigma_k^2 \simeq \mu_k \simeq 2$. Under obligate self-fertilization, a mode of reproduction in some plants and hermaphroditic animals, $\beta = 1$ and $N_e = N/2$. Self-fertilization results in a reduction in the effective population size because the nonindependence imposed by inbreeding reduces the effective number alleles per locus within individuals.

For the case of species with separate sexes, with β being the fraction of offspring derived by full-sib mating, and assuming equal numbers of males and females and Poisson-distributed family sizes,

$$N_e \simeq \frac{(4-3\beta)N}{4-2\beta} + 1 \tag{3.21}$$

(Wang 1995). Three special cases are of interest here. For a population derived entirely by full-sib mating, $\beta = 1$ and $N_e = N/2$, as in the case of complete selfing. With complete avoidance of sib-mating, $\beta = 0$ and $N_e = N+1$. Finally, with the type of population structure assumed, there are $N/2$ families, so under random mating, the probability of full-sib mating is simply $\beta = 2/N$, which implies $N_e \simeq N + 0.5$.

SPECIAL SYSTEMS OF MATING

It is important to bear in mind that, when applied to matters of genetic variation, the equations for N_e given above are appropriate for predicting the *expected* loss of heterozygosity resulting from inbreeding at the population level. When variation in pedigree structure exists among individuals, as will almost always be the case in nature, the actual degree of inbreeding will generally *vary* among loci within individuals as well as among individuals within the population. Given a cumulative level of inbreeding (loss of heterozygosity) at a locus equal to f, identity by descent will be binomially distributed among individuals with mean f and variance $\sigma_f^2 = f(1-f)$. With completely linked loci, this is also the total variance in f, as there will be no variation in f among loci. However, for unlinked loci, the realized inbreeding at each locus need not be the same. Weir et al. (1980) found that the coefficient of variation of $(1-f)$ among individuals is approximately $(3N_e)^{-1/2}$ for randomly mating monoecious populations, $(6N_e)^{-1/2}$ for randomly mating but monogamous, dioecious populations, and $(12N_e)^{-1/2}$ for monoecy with selfing excluded and for dioecy with random mating. These asymptotic values are reached in only a few generations. Thus, provided the population size and number of constituent loci are moderately large, the variation in inbreeding is negligible for most practical purposes (see also Franklin 1977; Cockerham and Weir 1983).

In contrast to the usual situation, as first pointed out by Wright (1921), one can envision (and implement) systems of mating involving fixed relationships such that all members of the population have exactly the same average inbreeding coefficient over all loci. Consider first the most extreme form of inbreeding – obligate self-fertilization, a mode of reproduction in some plants and hermaphroditic animals.

Because all individuals are reproductively isolated under this scheme of mating, a collection of such lines is equivalent to a series of populations, each consisting of a single individual, and after t generations of selfing, the expected fraction of heterozygotes at any locus is reduced to

$$H_t = H_0(1/2)^t \tag{3.22}$$

where H_0 denotes the initial level of variation. After $t = 10$ generations, only 0.1% of the initial heterozygosity remains (Figure 3.1).

–Insert Figure 3.1 Here–

The next most intense system of inbreeding involves continuous brother-sister mating. Starting with unrelated parents, it takes a generation of full-sib mating before alleles identical by descent can appear in the same individual. In one of the first applications of matrices in population genetics, Haldane (1937) showed that thereafter

$$H_t \simeq H_0(0.81)^t \tag{3.23a}$$

Thus, starting from a non-inbred base population, twelve generations of full-sib mating will result in a loss of 90% of the initial heterozygosity. Written in terms of the inbreeding coefficient, such that $H_t = (1 - f_t)H_0$, the exact recursion equation under full-sib mating is

$$f_t = \frac{1}{4}(1 + 2f_{t-1} + f_{t-2}) \tag{3.23b}$$

Moving on, with a constant population size of four breeding adults, the minimum relationship between individuals is that of double-first cousins (Figure 3.2, left). Starting with four unrelated individuals, it then takes three generations for alleles identical by descent to appear in the same individual, and thereafter

$$H_t \simeq H_0(0.92)^t \tag{3.24}$$

(Wright 1921). The number of generations required for the loss of 90% heterozygosity is now 30 (Figure 3.1).

–Insert Figure 3.2 Here–

These types of results are of special interest to managers of small, captive populations of endangered species and/or breeding stocks viewed as genetic resources for the future. Here we consider just one of the many practical questions that arise in these areas. Given a limited number of founders and an upper ceiling on the number of individuals that can be maintained, what is the optimal breeding scheme for minimizing the erosion of genetic variation? Wright (1921) suggested that the best way to minimize the loss of heterozygosity from a small population would be to restrict matings to pairs of individuals with the least degree of relatedness. Such a breeding scheme, known as **maximum avoidance of inbreeding** or MAI, is

exemplified by all three of the special mating systems just noted – in each case, matings occur between the most distantly related individuals within each line. An added advantage of MAI is that for a population size of $N = 2^m$, m generations pass before any inbreeding occurs at all. For example, with $N = 64$, $m+1 = 7$ generations would pass before two copies of a founding gene could appear in the same individual under a maximum avoidance scheme. Once the inbreeding begins, the proportion of heterozygosity lost each generation is very nearly constant, approaching an asymptotic value of $1/(4N - m - 1)$ (Robertson 1964), which with $N = 4$ and $m = 2$ under double-first cousin mating, yields Equation 3.24.

Note that when N is large, $m \ll N$, and the asymptotic rate of loss of heterozygosity is $\simeq 1/(4N)$ per generation under MAI. Comparing this expression with Equation 2.4c, it can be seen that this mating scheme has the same effect as doubling the size of a random mating population. This result is not strictly a consequence of the avoidance of inbreeding, but again is the outcome of all families producing equal numbers of offspring. In fact, even under a random mating scheme, if family sizes are equilibrated, provided $N \geq 4$, the erosion of heterozygosity is

$$H_t \simeq H_0 \left(1 - \frac{1}{4N}\right)^t \tag{3.25}$$

where t is the number of generations after the onset of inbreeding (Wright 1951). This can be seen by returning to many of the formulae in the earlier sections of this chapter and setting $\sigma_k^2 = 0$. Under the idealized scheme of random mating discussed earlier, variances in allele frequencies arise from variance in the number of progeny left by each individual and from segregational variance resulting from the sampling of alleles within individuals. For randomly mating populations of even moderate size, about half of the total sampling variance of allele frequency arises from each source, so that equilibration of family size reduces the total sampling variance by 50%.

The central point of the preceding discourse is that, for a fixed census size, three factors can potentially be manipulated to reduce the rate of inbreeding: avoidance of matings among relatives; equilibration of family sizes; and a sex ratio as close to one as possible. In some domesticated animals, the latter is a problem as females have only one or two offspring per year. In such cases, Gowe et al. (1959) suggest that when the sex ratio of contributing parents is r females to each male, the loss of genetic variance is minimized if every male contributes exactly one son and r daughters, and every female leaves one daughter and also contributes a son with probability $1/r$. Wang (1997) improved on this scheme with the constraint that a female contributing a son does not contribute a daughter, with another female from the same male family instead contributing two daughters.

Kimura and Crow (1963a) noted that Wright's intuition that MAI minimizes the long-term loss of genetic variation is actually not quite correct, pointing out that a **circular mating** (CM) scheme (Figure 3.2) ultimately leads to a lower rate of loss of heterozygosity. Under this breeding design, females and males are arranged such that each of them is mated to two "neighbors," with the last individual in the linear array being mated with the first, thereby completing the circle. Nevertheless, although circular mating ultimately reduces the rate of loss of heterozygosity relative to MAI, it is inferior in the early generations of mating, and even with small N, it

may take > 100 generations before its superiority is realized. Thus, because most of the initial genetic variation in a population will generally have been lost by this time, the practical advantages of circular mating are actually quite negligible.

The major limitation of both the MAI and CM schemes is that they only impede the loss of genetic variation – ignoring new mutations, any randomly mating finite population will ultimately become homozygous at every locus. However, Robertson (1964) obtained the more general (and counterintuitive) result that the rate of loss of overall genetic variation from a population actually declines as the relatedness between mates increases. In the extreme, genetic diversity can be preserved indefinitely by subdividing a population into several isolated lines. Although the individual lines are all expected to become homozygous eventually, different lines will become fixed for different sets of genes, with the overall level of preservation of genetic diversity being defined by the number of inbred lines. In effect, complete inbreeding gives rise to a condition equivalent to each family having preserved the equivalent of one gamete from the base population. For example, for a locus with initial allele frequency p_0, assuming a large number of families, the allele frequency in the total collection of lines would remain close to p_0, so subsequent random mating of the lines would render the heterozygosity close to its original state, $2p_0(1-p_0)$. It must be emphasized, however, that these arguments assume that intense inbreeding in small lines has no consequences that might endanger line survival. In reality, very small lines are likely to die out occasionally just by accident, and extreme inbreeding also often has serious deleterious effects on fitness (LW Chapter 10). The gradual replacement of extinct lines by members of surviving lines will lead to further loss of variation.

POPULATION SUBDIVISION

Although few species exhibit internal isolation as extreme as that just noted, many (probably most) species occupy spatially structured environments. Such structure causes local inbreeding, as mates are more related than random pairs from the entire population. One of the simplest models of population structure is Wright's (1951) island model, introduced in Chapter 2 (Figure 2.8). Here, the metapopulation consists of d demes, each consisting of a fixed number N of randomly mating individuals with idealized Wright-Fisher properties, with each deme contributing an equal fraction m of its genes to the migrant pool, which is then equally distributed among the remaining demes. Under this model, an equilibrium level of population divergence is eventually reached (Equation 2.47), at which point the increase in divergence resulting from within-deme genetic drift is balanced by the exchange of alleles by migration.

Recalling from Chapter 2 that the mean coalescence time for an ideal, random mating population is $2N$ generations, Equation 2.40a (giving the mean coalescence time under this population structure) implies an effective size for the overall metapopulation of

$$N_e = Nd + \frac{(d-1)^2}{4dm} \tag{3.26}$$

In this simplest case, we see that the effective size of a metapopulation exceeds the

sum of the demic effective sizes (Nd) by an amount approaching $d/(4m)$ when the number of demes is large. With low migration rates ($m \ll 1$), this inflation can be substantial. For example, if $1/(4m) > N$, N_e exceeds twice the total number of breeding adults ($2Nd$). This confirms Robertson's (1964) argument that population subdivision can reduce the overall rate of loss of variation by drift as unique alleles are sequestered within individual demes. However, numerous authors have pointed out that this inflation of N_e in the ideal island model is a special consequence of the absence of variation in deme productivity (analogous to the consequences of constant family sizes within a single population, noted above).

The next simplest type of island model allows for extinction and recolonization of individual demes. Each generation, a fraction e of the demes goes extinct, but immediate recolonization ensures the maintenance of a fixed number (d) of demes. Tracing back to Slatkin (1977) and Maruyama and Kimura (1980), most attempts to model this process have assumed that the newly colonized deme is immediately restored to size N in a single generation, with recolonization involving k immigrants either derived from a single random deme (the **propagule-pool model**) or from a random pool of migrants from the entire metapopulation (the **migrant-pool model**). This simple modification results in a reduction in N_e for the metapopulation by inducing variation in productivity among the demes surviving in each generation – demes that contribute to a colonization event experience a burst of productivity relative to demes that do not.

Using the logic noted above, approximate expressions for mean coalescence times for metapopulations experiencing extinction and recolonization derived by Pannell and Charlesworth (1999), yield formulae for the metapopulation N_e. Such expressions are functions of the relative rates of extinction and migration, and the size and type of colonizing pool, and we only give two examples for the propagule-pool model. If the extinction rate is smaller than the migration rate ($e \leq m$) and much smaller than the relative size of the colonizing pool ($e \ll k/N$),

$$N_e \simeq Nd\frac{[4m + (1/N)]}{4(e + m)} \tag{3.27a}$$

a result also obtained by Maruyama and Kimura (1980) and Whitlock and Barton (1997). Under these conditions, local extinctions are sufficiently rare that within-deme variation is able to recover substantially by migration between bottleneck events, and although $N_e < Nd$, it approaches the latter value as $e \to 0$ unless N is tiny. On the other hand, if the extinction rate is relatively high, such that $e \gg m$ and $e \gg k/N$,

$$N_e \simeq \frac{d}{4e} \tag{3.27b}$$

In this case, extinctions are so frequent that local demes (reestablished from a small number of colonists and experiencing little immigration) are almost completely inbred, and the total effective size is independent of the number of individuals per deme and simply defined by deme number and average deme longevity (e^{-1} generations).

A more general expression for the effective size of a metapopulation under the island model was derived by Whitlock and Barton (1997). Assuming large d, so that

$d \simeq d - 1$ as a first-order approximation,

$$N_e \simeq \frac{d(1 + 4Nm)}{4m + 2\sigma_K^2(1 + 2m)} \tag{3.28}$$

where σ_K^2 denotes the among-deme variance in the number of gametes contributing to the next generation. Comparing this expression to Equation 3.26, it can be seen that $\sigma_K^2 > 0$ will always cause a reduction in N_e. In addition, for N_e to be less than the total metapopulation size $2Nd$, σ_K^2 need only be larger than $1/(2N)$. This amount of among-deme variance in gamete production is trivial, as even under ideal conditions in which individual family sizes are Poisson distributed, the variance in total deme productivity will be on the order of $2/N$. This follows from the fact, noted above, that within an ideal random-mating deme, $\sigma_k^2 = 2$, so that at the deme level $\sigma_K^2 = 2/N$. Thus, it appears that population subdivision will almost always result in a reduction in N_e, provided the individual demes are not completely isolated, a point first made by Wright (1940; see also Nunney 1999).

SELECTION, RECOMBINATION, AND HITCH-HIKING EFFECTS

Up to now, we have assumed that alleles at the locus under consideration are immune to selective processes, which is of course unrealistic in many cases. In fact, selection generally causes a still further reduction in N_e by inflating the among-family variance in offspring production. Evaluating the magnitude of such effects is complicated by the fact that unlike family-size variation induced by environmental heterogeneity, which can be erased in a single generation, genetic variation in fitness is sustained across generations. Such heritable transmission will elevate the genetic representation of some individuals in future generations beyond the expectations with drift alone. This phenomenon was initially mentioned by Morley (1954), who noted in sheep flocks exposed to selection that "the genetically superior individuals will tend to be most inbred". The processes that we will examine are analogous to those that occur in spatially structured populations with random extinction, except that now specific alleles can become trapped in genetic backgrounds that are destined for elimination or fixation.

Before proceeding, it must be emphasized that because N_e is defined in the context of hypothetically neutral loci that serve as benchmarks for the pure drift process, our concern here is with not so much with the specific loci under selection, but with the effects of such selection on the dynamics of neutral-allele frequencies elsewhere in the genome. Because the long-term effects of selection depend on the frequency of recombination between selected loci and their associated neutral markers as well as on the mode of selection, the issues are quite technical. Our goal is to simply provide a heuristic overview of why the effects of selection almost always lead to a substantial reduction in N_e. Chapter 8 examines many of these issues in greater detail.

Effects From Selection at Unlinked Loci

Robertson (1961) first considered the influence of a constant selection regime on the long-term dynamics of a neutral locus assumed to be entirely unlinked to any selected loci. In addition to any baseline variance in gamete production among individuals that might exist for environmental reasons (our previous σ_k^2), in the first generation of selection there will also be an among-family genetic variance in relative fitness, σ_w^2, associated with the differential contributions of individual families. Here the relative fitness (w_i) of the ith family is simply the expected contribution to the next generation relative to the average in the population, such that the mean relative fitness $\sum_{i=1}^{N/2} w_i/(N/2) = 1$, where we assume a balanced sex ratio and $N/2$ families. For populations with features in accordance with the standard Wright-Fisher model, a single generation of selection will then reduce the effective population size to

$$N_e \simeq \frac{4N}{\sigma_k^2 + 2 + 4\sigma_w^2} \tag{3.29a}$$

which is identical in form to Equation 3.4, except for the additional variance associated with selection in the denominator. Note that because σ_w^2 is the variance in relative fitness among families, and the average family size is two, $4\sigma_w^2$ is the genetic variance in actual family size. This expression demonstrates that the random association of neutral alleles with families with different genetic endowments has the same qualitative effect as environmental differences in family sizes (σ_k^2).

Of course, it would be a rare situation in which selection just operated for a single generation. However, Robertson (1961) had the additional insight that with subsequent generations of selection, new stochastic associations between selected and unselected loci will arise each generation, while old associations are lost at rate 0.5 with free recombination. This yields a long-term cumulative contribution to the among-family genetic variance in relative fitness proportional to $[1 + (1/2) + (1/4) + (1/8) + \cdots]^2\sigma_w^2 = 4\sigma_w^2$, which (as above) is then multiplied by 2^2 to translate relative fitness into the genetic variance in family size. Although this result ignores the fact that the stochastic effects of selection will dissipate over time as favorable alleles at the loci under selection go to fixation, this additional layer of complexity is readily incorporated. Letting L denote the per-generation fractional loss of additive genetic variance at selected loci, the preceding series simply has terms in powers of $(1-L)/2$ instead of $1/2$, and again after converting the variance in relative fitness to the absolute scale, the long-term effective population size becomes

$$N_e \simeq \frac{4N}{\sigma_k^2 + 2 + 4[2/(1+L)]^2\sigma_w^2} \tag{3.29b}$$

(Santiago and Caballero 1995). If we further assume that baseline variation in family sizes unassociated with selection simply reflects random sampling, then from above $\sigma_k^2 \simeq 2$, and under Robertson's assumption of no reduction in variance by selection ($L = 0$), the long-term effective size becomes

$$N_e = \frac{N}{1 + 4\sigma_w^2} \tag{3.29c}$$

Equations 3.29b,c are quite general in the sense that they apply to any scheme of selection. However, they are also a bit opaque in that the mechanistic determinants

of σ_w^2 and L are not defined, and the substantial effects of linkage are omitted. The remainder of this chapter is focused on the removal of these limitations.

Example 3.5. The genetic variance for relative mean-family fitness is a function of the intensity of selection and the heritability of the selected traits. For example, in the case of truncation selection on a single trait (Chapter 12), $\sigma_w^2 = i^2 t_{FS}$, where i is the standardized selection differential (the change in mean phenotype imposed by selection in units of phenotypic standard deviations), and t_{FS} is the phenotypic correlation among full sibs (Milkman 1978), which is equivalent to half the heritability for an ideal trait with an additive genetic basis (LW Chapter 18). For situations in which the most extreme 1 to 10% of the phenotypic distribution is selected, i is in the range of 2.7 to 1.8 (LW Chapter 2), and t_{FS} takes on a maximum value of 0.5 when the heritability of the trait is equal to 1.0. Thus, recalling Equations 3.29a-c, with very strong truncation selection on a highly heritable trait, σ_w^2 may take on high enough values to reduce N_e severalfold relative to the expectation in the absence of selection, even when the selected loci are unlinked. Chapter 23 examines the reduction in N_e due to truncation selection on a trait in more detail.

Selective Sweeps and Genetic Draft

The effects of linked loci on N_e are substantially greater than those from unlinked loci for the simple reason that chromosomally juxtaposed sites are necessarily mutually influenced by each other's fitness attributes for extended periods. For example, a neutral allele linked to a site under positive selection can **hitch-hike** to high frequencies (and even fixation) if the force of selection is strong relative to the recombination rate between the sites (Maynard Smith and Haigh 1974). The alleles at such linked, neutral loci are expected to have a much more recent common ancestor, and therefore a smaller effective population size, than those for loci in regions not influenced by selection. One direct consequence of this reduction in N_e is a depressed amount of molecular variation at neutral sites in regions of low recombination, an expectation that is in agreement with many empirical observations (Chapter 8). However, as will be emphasized in the next section, the periodic fixation of favorable alleles is just one potential explanation for this kind of observation, an alternative hypothesis being **background selection**, the constant purging of new deleterious mutations (Charlesworth 2012).

A simple way of evaluating the effects of **selective sweeps** of beneficial mutations on variation at completely linked neutral loci was presented by Gillespie (2000). Recall from Chapter 2 that the variance of neutral allele-frequency change from generation to generation in a Wright-Fisher diploid population is equal to $p(1-p)/(2N)$, where p is the current allele frequency. Now imagine that this locus is completely linked to other genomic sites incurring beneficial mutations that collectively cause rapid fixations at an average rate of δ per generation. Because such mutations arise independently of the allele at the linked neutral locus, such selective sweeps will result in the fixation of neutral alleles with probabilities proportional

to their current frequencies, in this case p and $(1-p)$. If we assume that selective sweeps cleanse a population of linked variation essentially instantaneously (or at least rapidly relative to the usual rate of genetic drift), then conditional on a sweep occurring, the variance in allele-frequency change will be $p(1-p)$. Thus, for a neutral locus in an ideal randomly mating population subject to periodic selective sweeps, the variance in allele-frequency change is $\sim p(1-p)\{[(1-\delta)/(2N)]+\delta\}$. Because this expression applies to all initial allele frequencies, equating the term in brackets to $1/(2N_e)$ yields

$$N_e \simeq \frac{N}{1+2N\delta} \tag{3.30a}$$

a result also obtained by Maruyama and Birky (1991) by a different method. Here and below, it is appropriate to view N as the effective size of a population based solely on the demographic considerations noted earlier in this chapter.

When selective sweeps are rare relative to the strength of random genetic drift, such that $\delta \ll 1/(2N)$, $N_e \simeq N$, but as $N \to \infty$, $N_e \to 1/\delta$, showing that even populations with enormous numbers of reproductive adults may approach an asymptotic upper limit to N_e defined not by genetic drift but by **genetic draft** (Gillespie 2000) – the stochastic result of hitch-hiking effects that inevitably arise in linked genomes. That is, when N is very large, the effective size of a population can be more strongly influenced by the physical (i.e., linkage) features of the genome than by demographic factors. In principle, the frequency of selective sweeps might increase with N, as larger populations provide more opportunities for rare beneficial mutations, so strong linkage may even lead to the potential situation in which N_e eventually scales negatively with absolute population size (Lynch 2007).

The preceding result applies to the extreme case of complete linkage. If, instead, a significant amount of recombination occurs between a neutral marker and the selected locus while the latter is proceeding toward fixation, then a selective sweep is not expected to completely remove the variation at the marker locus. As detailed in Chapter 8, the extent to which a neutral locus can free itself of stochastic associations with newly arising beneficial mutations depends on the rate of the sweep (which in turn is a function of the relative power of selection and drift, $s/[1/(2N)] = 2Ns$) as well as on the relative power of recombination and selection (r/s, where r is the rate of recombination between the two loci).

Incorporation of these technical issues, worked out by Wiehe and Stephan (1993), lead to an expression identical in form to Equation 3.30a, with $2N\delta$ being replaced by a term that is smaller in absolute value. Gillespie (2000) expressed this influence of recombination as

$$N_e \simeq \frac{N}{1+2Nf_s\delta} \tag{3.30b}$$

where f_s is the probability that no recombination occurs between the selected site and the marker locus under consideration during the sweep. Following the completion of the sweep, f_s is equivalent to the probability that a random individual will contain two identical-by-descent copies of the original neutral allele on the gamete in which the new favorable mutation arose. With free recombination $f_s \simeq 0$, but with complete linkage, $f_s = 1$, returning us to Equation 3.30a. The general form for f_s is

$$f_s \simeq (4Ns)^{-c/s} \tag{3.31}$$

as further discussed in Chapter 8.

Background Selection

We now turn to the influence of selection against recurrently appearing deleterious mutations, which cause a still further reduction in N_e as a consequence of induced variation in family size. Contrary to the situation with selective sweeps, whose effects are sporadic and chromosomally restricted in scope, the effects of recurrent deleterious mutations are expected to be persistent across the entire genome for the simple reason that the vast majority of mutations are deleterious (LW Chapter 10). Here we attempt to provide a heuristic understanding of the effects of such **background selection** by considering separately the effects of unlinked and linked deleterious mutations, relying on a simple model in which interfering loci harbor two alternative allelic types (beneficial and deleterious).

If the beneficial allele at a locus mutates to a defective type at rate u per generation, with the latter causing a fractional reduction in heterozygote fitness equal to s, the equilibrium frequency of the deleterious allele is equal to u/s (provided s is substantially stronger than the power of drift and mutation) (Chapter 7). The genetic variance in relative fitness associated with this locus then has an expected value close to $2us$. This result can be obtained by noting that the additive genetic variance for a single locus is equal to $2a^2p(1-p)$ (LW Chapter 4), where a is the difference in phenotype between adjacent genotypic classes, and p is the allele frequency. In this case, $a = s$, and because u/s is small, $p(1-p) \simeq p = u/s$. Summing over all n loci capable of mutating to deleterious alleles, the total genetic variance in fitness among individuals is $2nus = Us$, where $U = 2nu$ is the diploid deleterious mutation rate.

This result can be used to evaluate the overall effect of unlinked background selection by noting that unless the number of chromosomes is very tiny, the vast majority of pairs of genes within genomes are unlinked (with x chromosomes of equal size, the fraction of linked pairs will be $< 1/x^2$ because genes located on opposite ends of chromosomes are effectively unlinked). Thus, we can make use of Robertson's equation (3.29c), noting that the variance of mean family fitness is $Us/2$ after discounting by averaging over both parents. This shows that in the absence of any linkage, deleterious mutations are expected to cause a relatively small reduction in N_e,

$$N_e \simeq \frac{N}{1+2Us} \simeq Ne^{-2Us} \tag{3.32}$$

with the exponential approximation applying under the assumption that $Us \ll 1$. This assumption is justified by numerous observations suggesting that U is on the order of 0.1 to 1.0 and s is on the order of 0.01 (LW Chapter 12).

Some impression of the impact of linkage follows from the logic used to obtain Equations 3.29b,c. As noted above, for unlinked loci, the initial stochastic associations of neutral alleles and selected loci last for an average of $\sum_{i=0}^{\infty}(1/2)^i = 2$ generations, giving a total contribution to the variance proportional to 2^2. Letting the recombination rate between loci be c, this expression generalizes to $\sum_{i=0}^{\infty}(1-c)^i = 1/c$ generations, and hence a contribution of $1/c^2$ to the variance. Thus, a single selected

locus is expected to reduce the variation at a linked neutral locus, which in effect causes a local depression in the effective population size to

$$N_e \simeq Ne^{-us/c^2} \tag{3.33}$$

(Barton 1995). With $c \ll 0.5$, the absolute value of the exponential term can be considerably larger than that for unlinked loci, $4us$. The challenge is to determine the joint effects of the full spectrum of all linked and unlinked loci surrounding the neutral reference locus.

Insight into the overall power of selection on linked deleterious mutations can be gleaned by considering the extreme case of a completely nonrecombining, but otherwise sexual, genome, i.e., allowing only for segregation during gamete production. Assuming a total of n selected loci, for which the mutant alleles have identical and multiplicative effects on fitness, the average number of deleterious alleles in a gamete is $nu/s = U/(2s)$, where U is again the deleterious mutation rate per diploid genome. Assuming large enough N_e that deleterious mutations do not go to fixation, with this type of genomic architecture, only those gametes that are free of deleterious mutations can contribute to the future genetic constitution of the population. Because the number of deleterious mutations per gamete is Poisson distributed in sufficiently large populations with multiplicative selection (Kimura and Maruyama 1966), the frequency of such mutation-free gametes is simply $e^{-U/(2s)}$, leading to the conclusion that with complete linkage, selection against segregating deleterious mutations leads to

$$N_e = Ne^{-U/(2s)} \tag{3.34}$$

(Charlesworth et al. 1993). This shows that background selection has the potential to cause a dramatic reduction in N_e in a nonrecombining (but segregating) population. For example, with $U = 0.1$ and $s = 0.01$, $e^{-U/(2s)} = 0.000045$. This expression also applies to a nonrecombining chromosomal region if U is redefined to be the deleterious mutation rate for the region under consideration.

Hudson and Kaplan (1994) extended this result to allow for recombination, assuming that the latter operates at uniform rates per physical distance over chromosomal regions. Their results show that

$$N_e \simeq Ne^{-U/(2s+C)} \tag{3.35a}$$

where C denotes the rate of recombination between the ends of the region. Because $1/(2s+C) \ll s$, the absolute value of the exponential term is much smaller than that obtained for freely recombining loci ($2Us$ from Equation 3.32), showing that the total contribution from interference from unlinked loci (which is embedded in Equation 3.35a) is relatively minor relative to that from loci in the immediate vicinity of the neutral locus. Moreover, because s is expected to be $\ll 1$, and for an entire chromosome, C is of order 1.0, Equation 3.35a can be roughly approximated as

$$N_e \simeq Ne^{-U/C} \tag{3.35b}$$

where U/C is equivalent to the diploid deleterious mutation rate per unit of recombination (Hudson and Kaplan 1994, 1995). This result, which has been obtained

by several different methods (Barton 1995; Nordborg et al. 1996; Santiago and Caballero 1998), shows that the impact of segregrating deleterious mutations on N_e is largely independent of the mutational effect s. As will be seen in Chapter 4, the ratio of mutation and recombination rates can be estimated from molecular polymorphism data, so if the fraction of mutations that are deleterious is known, U/C is also estimable.

Finally, it is worth noting that some conditions exist under which selection may actually promote an increase in N_e, the most obvious being increases in the coalescence times for linked alleles in a chromosomal region within which sites under balancing selection are embedded. Pálsson and Pamilo (1999) also found that with very strong linkage and a low efficiency of selection ($2Ns < 1$), repulsion disequilibrium can build up between simultaneously segregating deleterious mutations, leading to a form of associative overdominance (LW Chapter 10) and an elevation of N_e. Although this condition arises in the absence of direct balancing selection on any specific site, it remains unclear whether the special requirements necessary for such an outcome are very common. Santiago and Caballero (2005) also found that in a subdivided population, selective sweeps within demes can sometimes lead to an increase in N_e for the total metapopulation, as the migration of a sweeping chromosomal region drags new variation into a recipient deme, leading to an overall effect akin to balancing selection. The magnitude of this effect is defined by a special balance of recombination, selection, and migration rates.

With the above exceptions aside, there are two general lessons to be learned from all of the above discussion. First, although the individual demographic and genetic effects that influence N_e may appear to be only moderate in nature, their cumulative effects can easily depress N_e below the actual number of reproductive individuals by several orders of magnitude. Second, although population geneticists often develop analytical descriptions of various processes under the assumption of an effectively infinite population size, the physical linkage of the genome ensures that even populations with extraordinarily large N need not be immune to drift-like processes imposed by hitch-hiking effects. These points will be made clearer in Chapter 4 as we explore the direct manifestation of such effects on standing variation in natural populations.

Example 3.6. Because natural populations are subject to both positive and negative selective forces, the total influence of selection on N_e must reflect *both* background selection and selective sweeps. This necessarily raises even more technical issues than outlined above. Significant progress has been made by Kim and Stephan (2000), and here we simply outline the basic result. If background selection operates as an essentially continuous process, as a consequence of the recurrent introduction of deleterious alleles, the depressive effects of both forms of selection, as well as baseline demographic effects, may be treated as largely independent. The reduction in N_e resulting from background selection can then be obtained by use of one of the above expressions, e.g., Equation 3.35b as a first-order approximation for a sexual population (with N already taking into consideration demographic effects).

Consider a large monoecious population of constant breeding size and variance in

family size $\sigma_k^2 = 4$. Based on demographic considerations alone, from Equation 3.4, $N_e \simeq N/3$. Letting $U = 1$ and $C = 1$, Equation 3.35b implies that background deleterious mutations further reduce N_e to $(N/3)e^{-1} \simeq 0.12N$. The effective population size dictated by these demographic and deleterious-mutation processes further defines the background N_{eb} within which occasional beneficial mutations arise and sweep to fixation, so that the effective population size resulting from the joint operation of all three effects can be approximated by substituting N_{eb} for N in Equation 3.30b. Supposing a complete sweep occurs every 10,000 generations (so that $f_s = 1$ and $\delta = 0.0001$), then $N_e = 0.12N/[1 + (0.24N \cdot 0.0001)]$. With $N = 10^4$, 10^6, and 10^8, this implies $N_e \simeq 0.1N = 1000$, $0.005N = 5000$, and $0.00005N = 5000$, respectively.

In general, the joint operation of background selection and selective sweeps will reduce N_e more than either does alone, although it is at least in principle possible for background selection to reduce the influence of selective sweeps in regions of very low recombination by depressing the probability of fixation of beneficial mutations. The simultaneous operation of positive selection on multiple loci (which was ignored in the derivation of Equations 3.30a,b) can also slightly alleviate the overall effects of selection on N_e as simultaneously segregating mutations interfere with each others' fixation, thereby reducing the incidence of complete selective sweeps (Kim and Stephan 2003). These issues are examined in greater detail in Chapter 8.

Literature Cited

Barton, N. H. 1995. Linkage and the limits to natural selection. *Genetics* 140: 821–841. [3]

Briscoe, D. A., J. M. Malpica, A. Robertson, G. J. Smith, R. Frankham, R. G. Banks, and J. S. F. Barker. 1992. Rapid loss of genetic variation in large captive populations of *Drosophila* flies: implications for the genetic management of captive populations. *Cons. Biol.* 6: 416–425. [3]

Caballero, A. 1994. Developments in the prediction of effective population size. *Heredity* 73: 657–679. [3]

Caballero, A. 1995. On the effective size of populations with separate sexes, with particular reference to sex-linked genes. *Genetics* 139: 1007–1011. [3]

Caballero, A., and W. G. Hill. 1992. A note on the inbreeding effective population size. *Evolution* 46: 1969-1972. [3]

Charlesworth, B. 1994. *Evolution in age-structured populations.* Second Edition, Cambridge Univ. Press, Cambridge, UK. [3]

Charlesworth, B. 2001. The effect of life-history and mode of inheritance on neutral genetic variability. *Genet. Res.* 77: 153–166. [3]

Charlesworth, B. 2009. Effective population size and patterns of molecular evolution and variation. *Nature Rev. Genetics* 10: 195–205. [3]

Charlesworth, B. 2012. The effects of deleterious mutations on evolution at linked sites. *Genetics* 190: 5–22. [3]

Charlesworth, B., M. T. Morgan, and D. Charlesworth. 1993. The effect of deleterious mutations on neutral molecular variation. *Genetics* 134: 1289–1303. [3]

Choy, S. C., and B. S. Weir. 1978. Exact inbreeding coefficients in populations with overlapping generations. *Genetics* 89: 591–614. [3]

Clutton-Brock, T. H., F. E. Guinness, and S. D. Albon. 1982. *Red deer: behavior and ecology of the two sexes.* Univ. Chicago Press, Chicago, IL. [3]

Cockerham, C. C., and B. S. Weir. 1983. Variance of actual inbreeding. *Theor. Popul. Biol.* 23: 85–109. [3]

Crow, J. F. 1954. Breeding structure of populations. II. Effective population number. *In* O. Kempthorne, T. A. Baneroft, J. W. Gowen, and J. L. Lush (eds.), *Statistics and mathematics in biology,* pp. 543–556. Iowa State College Press, Ames, IA. [3]

Crow, J. F., and C. Denniston. 1988. Inbreeding and variance effective population numbers. *Evolution* 42: 482–495. [3]

Crow, J. F., and M. Kimura. 1970. *An introduction to population genetics theory.* Harper and Row, Publ., NY. [3]

Crow, J. F., and N. E. Morton. 1955. Measurement of gene frequency drift in small populations. *Evolution* 9: 202–214. [3]

Emigh, T. H., and E. Pollak. 1979. Fixation probabilities and effective population numbers in diploid populations with overlapping generations. *Theor. Popul. Biol.* 15: 86–107. [3]

Felsenstein, J. 1971. Inbreeding and variance effective numbers in populations with overlapping generations. *Genetics* 68: 581–597. [3]

Frankham, R. 1995. Effective population size/adult population size ratios in wildlife: a review. *Genet. Res.* 66: 95–107. [3]

Franklin, I. R. 1977. The distribution of the proportion of the genome which is homozygous by descent in inbred individuals. *Theor. Popul. Biol.* 11: 60–80. [3]

Gillespie, J. H. 2000. Genetic drift in an infinite population. The pseudohitchhiking model. *Genetics* 155: 909–919. [3]

Gowe, R. S., A. Robertson, and B. D. H. Latter. 1959. Environment and poultry breeding problems. 5. The design of poultry control strains. *Poultry Sci.* 38: 462–471. [3]

Grant, P. R. 1990. Reproductive fitness in birds. *Trends Ecol. Evol.* 5: 379–380. [3]

Haldane, J. B. S. 1937. Some theoretical results of continued brother-sister mating. *J. Genetics* 34: 265–274. [3]

Hedgecock, D. 1994. Does variance in reproductive success limit effective population size of marine organisms? *In* A. Beaumont (ed.), *Genetics and evolution of aquatic organisms,* pp. 122–134. Chapman and Hall, London, UK. [3]

Hedrick, P. W. 1986. Average inbreeding or equilibrium inbreeding? *Amer. J. Hum. Genet.* 38: 965–970. [3]

Hedrick, P. W. 2005. Large variance in reproductive success and the N_e/N ratio. *Evolution* 59: 1596–1599. [3]

Hedrick, P. W., and C. C. Cockerham. 1986. Partial inbreeding: equilibrium heterozygosity and the heterozygosity paradox. *Evolution* 40: 856–861. [3]

Heywood, J. S. 1986. The effect of plant size variation on genetic drift in populations of annuals. *Amer. Natur.* 127: 851–861. [3]

Hill, W. G. 1972. Effective size of populations with overlapping generations. *Theor. Popul. Biol.* 3: 278–289. [3]

Hill, W. G. 1979. A note on effective population size with overlapping generations. *Genetics* 92: 317–322. [3]

Hudson, R. R., and N. L. Kaplan. 1994. Gene trees with background selection. *In* G. B. Golding (ed.), *Non-neutral evolution: theories and molecular data,* pp. 140–153. Chapman and Hall, New York. [3]

Hudson, R. R., and N. L. Kaplan. 1995. Deleterious background selection with recombination. *Genetics* 141: 1605–1617. [3]

Johnson, D. L. 1977. Inbreeding in populations with overlapping generations. *Genetics* 87: 581–591. [3]

Kim, Y., and W. Stephan. 2000. Joint effects of genetic hitchhiking and background selection on neutral variation. *Genetics* 155: 1415–1427. [3]

Kim, Y., and W. Stephan. 2003. Selective sweeps in the presence of interference among partially linked loci. *Genetics* 164: 389–398. [3]

Kimura, M., and J. F. Crow. 1963a. On the maximum avoidance of inbreeding. *Genet. Res.* 4: 399–415. [3]

Kimura, M., and J. F. Crow. 1963b. The measurement of effective population number. *Evolution* 17: 279–288. [3]

Kimura, M., and T. Maruyama. 1966. The mutational load with epistatic gene interactions in fitness. *Genetics* 54: 1337–1351. [3]

Lande, R., and G. F. Barrowclough. 1987. Effective population size, genetic variation, and their use in population management. *In* M. E. Soulé (ed.), *Viable populations for conservation,* pp. 87–124. Cambridge Univ. Press, Cambridge, UK. [3]

Latter, B. D. H. 1959. Genetic sampling in a random mating control population of constant size and sex-ratio. *Aust. J. Biol. Sci.* 12: 500–505. [3]

Lynch, M. 2007. *The origins of genome architecture.* Sinauer Assocs., Inc., Sunderland, MA. [3]

Maruyama, T., Birky, C. W., Jr. 1991. Effects of periodic selection on gene diversity in organelle genomes and other systems without recombination. *Genetics* 127: 449–451. [3]

Maruyama, T., and M. Kimura. 1980. Genetic variability and effective population size when local extinction and recolonization of subpopulations are frequent. *Proc. Natl. Acad. Sci. USA* 77: 6710–6714. [3]

Maynard-Smith, J., and J. Haigh. 1974. The hitch-hiking effect of a favourable gene. *Genet. Res.* 23: 23–35. [3]

Milkman, R. 1978. Selection differentials and selection coefficients. *Genetics* 88: 391–403. [3]

Morley, F. H. W. 1954. Selection for economic characters in Australian Merino sheep. IV. Further estimates of phenotypic and genetic parameters. *Aust. J. Agric. Res.* 6: 77–90. [3]

Nagylaki, T. 1995. The inbreeding effective population number in dioecious populations. *Genetics* 139: 473–485. [3]

Nordborg, M., B. Charlesworth, and D. Charlesworth. 1996. The effect of recombination on background selection. *Genet. Res.* 67: 159–174. [3]

Nunney, L. 1993. The influence of mating system and overlapping generations on effective population size. *Evolution* 47: 1329–1341. [3]

Nunney, L. 1999. The effective size of a hierarchically structured population. *Evolution* 53: 1–10. [3]

Nunney, L. 2002. The effective size of annual plant populations: the interaction of a seed bank with fluctuating population size in maintaining genetic variation. *Amer. Natur.* 160: 195-?204. [3]

Nunney, L., and D. R. Elam. 1994. Estimating the effective size of conserved populations. *Cons. Biol.* 8: 175–184. [3]

Pálsson, S., and P. Pamilo. 1999. The effects of deleterious mutations on linked, neutral variation in small populations. *Genetics* 153: 475-483. [3]

Pannell, J. R., and B. Charlesworth. 1999. Neutral genetic diversity in a metapopulation with recurrent local extinction and recolonization. *Evolution* 53: 664–676. [3]

Robertson, A. 1961. Inbreeding in artificial selection programmes. *Genet. Res.* 2: 189–194. [3]

Robertson, A. 1964. The effect of nonrandom mating within inbred lines on the rate of inbreeding. *Genet. Res.* 5: 164–167. [3]

Santiago, E., and A. Caballero. 1995. Effective size of populations under selection. *Genetics* 139: 1013–1030. [3]

Santiago, E., and A. Caballero. 1998. Effective size and polymorphism of linked neutral loci in populations under directional selection. *Genetics* 149: 2105–2117. [3]

Santiago, E., and A. Caballero. 2005. Variation after a selective sweep in a subdivided population. *Genetics* 169: 475-483. [3]

Slatkin, M. 1977. Gene flow and genetic drift in a species subject to frequent local extinctions. *Theor. Popul. Biol.* 12: 253–262. [3]

Templeton, A. R. 2006. *Population genetics and microevolutionary theory.* John Wiley & Sons, Inc., Hoboken, NJ. [3]

Vucetich, J. A., T. A. Waite, and L. Nunney. 1997. Fluctuating population size and the ratio of effective to census population size. *Evolution* 51: 2017–2021. [3]

Wang, J. 1995. Exact inbreeding coefficient and effective size of finite populations under partial sib mating. *Genetics* 140: 357–363. [3]

Wang, J. 1996. Inbreeding and variance effective sizes for nonrandom mating populations. *Evolution* 50: 1786–1794. [3]

Wang, J. 1997. More efficient breeding systems for controlling inbreeding and effective size in animal populations. *Heredity* 79: 591–599. [3]

Weir, B. S., P. J. Avery, and W. G. Hill. 1980. Effect of mating structure on variation in inbreeding. *Theor. Popul. Biol.* 18: 396–429. [3]

Wiehe, T. H., and W. Stephan. 1993. Analysis of a genetic hitchhiking model, and its application to DNA polymorphism data from *Drosophila melanogaster*. *Mol. Biol. Evol.* 10: 842–854. [3]

Whitlock, M. C., and N. H. Barton. 1997. The effective size of a subdivided population. *Genetics* 146: 427–441. [3]

Wright, S. 1921. Systems of mating. II. The effects of inbreeding on the genetic composition of a population. *Genetics* 6: 124-143. [3]

Wright, S. 1931. Evolution in Mendelian populations. *Genetics* 16: 97–159. [3]

Wright, S. 1938. Size of population and breeding structure in relation to evolution. *Science* 87: 430–431. [3]

Wright, S. 1939. Statistical genetics in relation to evolution. *Actualités Scientifiques et Industrielles, No. 802. Exposés de Biometrie et de Statistique Biologique. XIII.* Hermann and Cie, Paris. [3]

Wright, S. 1940. Breeding structure of populations in relation to speciation. *Amer. Natur.* 74: 232–248. [3]

Wright, S. 1969. *Evolution and the genetics of populations. II. The theory of gene frequencies.* Univ. Chicago Press, Chicago, IL. [3]

Figure 3.1. Erosion of expected heterozygosity under continuous breeding schemes involving self-fertilization, full-sib mating, and double first-cousin mating.

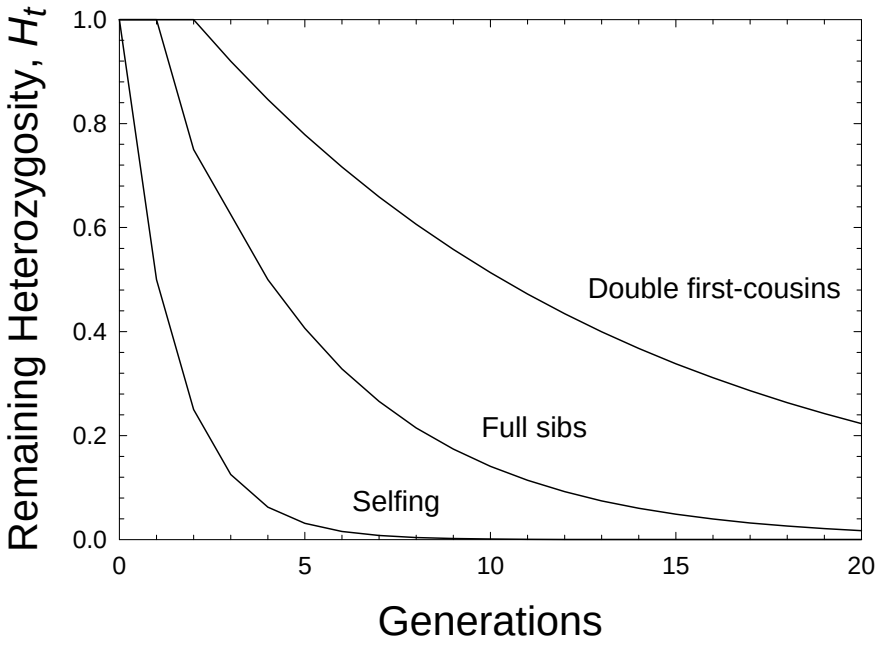

Figure 3.2. Mating schemes under continual double first-cousin mating (left) and under circular mating with four individuals (right). Genes identical-by-descent do not appear in the same individual for three and four generations, respectively, under these two schemes. Males and females are denoted by separate symbols (squares and circles), and lines connect parents (above) and offspring (below).