

3

THE GENETIC EFFECTIVE SIZE OF A POPULATION

Draft Version 30 March 2007

Throughout the preceding chapter, we assumed a population with an idealized set of Wright-Fisher features, including random mating among a homogeneous group of monoecious, self-compatible individuals with equal expected family sizes, discrete generations, and an absence of density fluctuations. Thus, 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 seem in doubt. In fact, much of the theory of inbreeding and random genetic drift can be generalized to other types of population structure 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 (Example 2.3) is a case in point – a set of populations with actual size $N = 16$ exhibited allele-frequency dynamics closely approximating the expectations for an idealized Wright-Fisher population with an effective size of just ten individuals.

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 size. 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 can in some cases be 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, magnifying the stochastic fluctuations of allele frequencies resulting from chance associations with other loci under selection. Tight linkage to a deleterious mutation can result in loss from the population, whereas tight linkage to a beneficial mutation can result in hitchhiking to fixation.

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 $P = 1/N$. 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,

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

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

Under the preceding 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 2.12a 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), but unless stated otherwise, the following discussion will assume populations of constant size.

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 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). To ease the presentation of this rather technical area, we will consider the contributions of several broad classes of factors to the definition of N_e on an individual basis. This is necessarily an oversimplification because N_e will generally be influenced simultaneously by multiple interacting factors.

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. 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 have been extended to sex-linked loci (Caballero 1995; Nagylaki 1995; Wang 1997; Charlesworth 2001).

Although the magnitude of its strength can vary by orders of magnitude (Chapter 4), random genetic drift is a fundamental evolutionary property of all populations, resulting from two stochastic processes. First, variation in family size, even that generated by nonselective forces, encourages some lineages of alleles (those that happen to be contained within large families) to expand at the expense of others. As we will see in the following sections, this type of variation arises by numerous mechanisms, including simple sampling of gametes, population spatial structure, variation among the sexes, and trapping in various genetic backgrounds by linkage. Second, in sexual species, segregation during meiosis leads to the stochastic transmission of alleles from heterozygous parents, as there is a 50% chance that each alternative allele will be inherited by any given offspring.

MONOECY

To illustrate the mathematical approach to deriving expressions for N_e , we first generalize the concept of the monoecious, self-compatible population to allow arbitrary gamete production by different individuals. Let k_i be the number of gametes that the i th 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 in generation t containing gametes derived from the

same parent is

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

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,

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

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, and the mean and variance 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. Such variation is expected to be the rule in natural populations occupying microenvironments that are heterogeneous with respect to resource and/or mate availability. Thus, on this basis alone, we can generally expect the effective population size to be less than the number of reproductive adults.

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

$$N_e = \frac{4N - 2}{\sigma_k^2 + 2} \quad (3.4)$$

If we now add the assumption 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 (0 or 1) of gametes drawn from a particular parent 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 deviation resulting because we assume a fixed population size. Substituting into Equation 3.4, it is now found that $N_e = N$, i.e., the inbreeding effective size of an ideal random-mating population is 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 random variance in family size doubles the effective population size beyond the actual number of breeding adults.

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 can result

if vagaries in oceanographic conditions are such that only a small fraction of adults actually produce gametes at points in time and space that allow a contribution to recruitment in 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 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 $(N_b/N) \cdot 2N \cdot (1/N_b)[1 - (1/N_b)] = 2(1 - (1/N_b))$. Summing these two sources of family-size variance, $\sigma_k^2 = [(4N - 2)/N_b] - 2$, 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 .

The breeding systems of some monoecious species deviate from the assumptions made in the preceding derivations. For example, many hermaphroditic species are self-incompatible, in which case immediate 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 of drawing a pair of genes derived from the same grandparent is

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

Employing the same substitutions used for Equation 3.3,

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

and for constant N such that $\mu_k = 2$,

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

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. With the prohibition of selfing, N must be at least two.

In most hermaphroditic species, there is also 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,

$$N_e = \frac{N_{t-1}}{(4\sigma_{ep}/\mu_k^2) + 1} \quad (3.8)$$

where σ_{ep} is the covariance of the numbers of successful male (p) and female (e) gametes per parent (Crow and Denniston 1988). If σ_{ep} 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 σ_{ep} 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 μ_e for seed (e) production = μ_p for pollen (p) production = 1, and $\mu_k = \mu_e + \mu_p = 2$, 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 $\sigma_{kp} = 3\sigma_{ke}$, and a perfect correlation between seed and pollen production. Because the correlation between the number of female and male gametes, ke and kp , produced per individual is defined to be $\sigma_{ep}/(\sigma_{ke}\sigma_{kp})$, the latter assumption implies $1 = \sigma_{ep}/[\sigma_{ke} \cdot 3\sigma_{ke}]$. Assuming random mating, what is N_e ? Substituting $\sigma_{ep} = 3\sigma_{ke}^2$ into Equation 3.8, we obtain $N_e = N/[(12\sigma_{ke}^2/\mu_k^2) + 1]$. Thus, σ_{ke}^2/μ_k^2 in the range of 1 to 4, N_e is between 8% and 2% 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. Such a condition is 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

$$P = \frac{1}{N_e} = \frac{1}{4N_{em}} + \frac{1}{4N_{ef}} \quad (3.9)$$

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 acquire Equation 3.3. Skipping the intermediate steps, we simply note that

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

where s denotes the sex (m or f), and μ_{sk} and σ_{sk}^2 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,

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

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

where for male parents, σ_{mm}^2 is the variance of male progeny number, σ_{mf}^2 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, $\sigma_{mm,mf}$ and $\sigma_{fm,ff}$ 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,

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

where ϕ is the sex ratio (proportion of females). 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 $\phi \rightarrow 1$, $N_e \rightarrow 4(1-\phi)N_{t-1} \simeq 4N_{m,t-1}$.

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 grow up in different microenvironments that influence the availability of resources and mates, it is reasonable to expect that σ_k^2 will usually exceed the mean, 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 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 effective female 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 that are 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 that the N_e/N ratio for single generations is typically ≤ 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/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 schedule of age-specific survival and reproduction. For an ideal monoecious population,

$$T = \frac{\sum_{i=1}^{\tau} i l_i b_i}{\sum_{i=1}^{\tau} l_i b_i} \quad (3.13)$$

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),

$$T = \frac{T_{mm} + T_{mf} + T_{fm} + T_{ff}}{4} \quad (3.14)$$

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 $N = N_{eb}T$, where ϕ_b is the sex ratio of newborns and $N_{eb} = 4\phi_b(1 - \phi_b)N_b$ is the effective size of the newborn age class, 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 schedules for survival (l_i) and reproduction (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

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

where $\sum_s^2 = \left(\sum_{j \geq i+1}^{\tau_s} l_j^s b_j^s \right)^2$, 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 statistical 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 less than N , but generally above $N/2$, 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 in ascertaining paternity. 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 summations in the denominator of Equation 3.15 reflect the variation in lifetime reproductive success of females and males. As outlined in the following 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, $N_e \simeq 0.15 \times 270 \times 9.32 = 378$.

For comparative purposes, it is sometimes useful to convert the effective size of an age-structured population to an annual effective size, $N_y = TN_e$, which is the size of an ideal monoecious population with a generation time of one year corresponding to the annual increment in inbreeding in the observed population. For the red deer example, $N_y = 9.32 \times 378 = 3523$.

The age-specific survival rates, l_i , 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 $(\sum_{j>i+1}^{T_s} l_j^s b_j^s)^2$, and column (3) is the product of (1) and (2).

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 by the projection of $e^{-t/2N_e^*}$ depending upon the temporal changes in N_{e_i}

Example 3.4. Population bottlenecks have especially pronounced effects on N_e^* . To see this, 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. This is much closer to the expectation for a constant population size of 10 than 100. Frankham (1995) and Vucetich et al. (1997) have shown for a diversity of natural populations of animals that N_e^* is frequently in the range 10 to 20% of N_e .

PARTIAL INBREEDING

In most of the previous formulations, the assumption was made that the union of gametes is random. However, 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 and there is no overall loss of diversity (Wright 1951, 1969; Hedrick 1986; Hedrick and Cockerham 1986). Such an equilibrium does not exist for finite populations, because 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 that the population is otherwise randomly mating. Further details can be found in 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,

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

(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

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

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 $N_e = N/2$. Complete inbreeding results in a 50% reduction in N_e because the nonindependence imposed by inbreeding reduces the effective number alleles per locus to just one per individual.

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

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

(Wang 1995). Three special cases are of interest. For a population derived entirely by full-sib mating, $\beta = 1$ and $N_e \simeq (N/2) + 1$. With complete avoidance of sib-mating, $\beta = 0$ and $N_e \simeq N + 1$. Under random mating, because there are $N/2$ families, the probability of full-sib mating is $\beta = 2/N$, implying that $N_e \simeq N + 0.5$.

SPECIAL SYSTEMS OF MATING

It is important to bear in mind that, strictly speaking, the equations for N_e given above predict the average 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 average 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 $(3N)^{-1/2}$ for randomly mating monoecious populations, $(6N)^{-1/2}$ for randomly mating but monogamous, dioecious populations, and $(12N)^{-1/2}$ for monoecy with selfing excluded and for dioecy with random mating. These asymptotic values are reached after only a few generations of inbreeding. 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).

Nevertheless, as first pointed out by Wright (1921), one can envision (and implement) systems of mating that involve 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. Because all self-fertilizing 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. After t generations of selfing, the expected fraction of heterozygotes at any locus is reduced to

$$H_t = H_0(0.50)^t \quad (3.22)$$

where H_0 denotes the initial level of variation. More than 90% of the initial heterozygosity is lost after just $t = 4$ generations (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

$$H_t \simeq H_0(0.81)^t \quad (3.23a)$$

Thus, starting from a non-inbred base population, twelve generations of full-sib mating results 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

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

The minimum relationship between individuals in a constant population of four breeding adults 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

$$H_t \simeq H_0(0.92)^t \quad (3.24)$$

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

These types of results are of special interest to managers attempting to preserve small, captive populations of endangered species and/or breeding stocks. Given a limited number of founders and an upper ceiling on the number of individuals that can be maintained, 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** 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, and an added advantage of MAI is that for a population of size $N = 2^m$, m generations pass before any inbreeding occurs at all (Figure 3.2). 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 MAI 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.

–Insert Figure 3.2 Here–

When N is large, such that $m \gg N$, the asymptotic rate of loss of heterozygosity is $\simeq 1/(4N)$ per generation under MAI, showing that MAI has the same effect as

doubling the effective size of a random mating population. This result is not strictly a consequence of the avoidance of inbreeding, but rather the outcome of all families producing equal numbers of offspring. In fact, as noted above in the context of randomly mating populations, if family sizes are equilibrated, provided $N \geq 4$, the erosion of heterozygosity is

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

As emphasized earlier, variance in allele-frequency change arises from variance in the number of progeny left by each individual and from segregational variance resulting from the sampling of alleles within individuals. For populations of even moderate size, about half of the total sampling variance of allele frequency arises from each source. Thus, by ensuring that every individual produces exactly one son and one daughter, equilibration of family size eliminates the first source of sampling error.

Kimura and Crow (1963a) subsequently noted that Wright's intuition that MAI minimizes the long-term loss of genetic variation is not strictly 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. 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 utility of circular mating is 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 outcrossing population will ultimately become homozygous at every locus, and Robertson (1964) arrived at the counterintuitive conclusion that the rate of loss of overall genetic variation 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. For example, for the case of complete selfing with each individual contributing exactly one offspring to the next generation, or for complete full-sib mating with each family contributing exactly one son and one daughter to the next generation, permanent population subdivision is established from the outset and the entire collection of inbred lines will asymptotically approach a state of permanent allele preservation. With all lines eventually becoming homozygous, the overall level of preservation of genetic diversity will be defined by the number of lines, as each family will preserve the equivalent of one gamete from the base population. For a locus with initial allele frequency p , assuming a large number of families, the allele frequency in the total collection of lines would remain close to p , so subsequent random mating of the lines would render the heterozygosity close to its original state, $2p(1-p)$. It must be emphasized, however, that these arguments assume that intense inbreeding has no consequences that might endanger line survival. In reality, very small lines are likely to die out occasionally just by accident, extreme inbreeding often has serious deleterious effects on fitness (LW Chapter 10), and the gradual replacement of extinct lines by members of surviving lines will lead to further loss of variation.

POPULATION SUBDIVISION

In contrast to the single closed-population setting assumed up to now, many species occupy spatially structured environments, with limited migration occurring among a series of interconnected **demes**, collectively referred to as a **metapopulation**. Such structure causes local inbreeding, as mates tend to be more related than random members of the entire population. An expansive series of different population structures can be imagined, but we will focus most of our discussion on an **island model**, wherein each deme derives random immigrants from a general migrant pool drawn from the entire metapopulation. Migration will be assumed to be conservative, in the sense that local deme sizes remain constant through time, and generations will be assumed to be discrete. Even this scheme admits to many possibilities, as both deme sizes and contributions to the migrant pool can vary. But this latter point leads to a central conclusion – like variation in family size, variation in deme productivity often results in a reduction in the effective size of a metapopulation relative to the total number of breeding adults.

This conclusion was not reached in the simplest finite island model first analyzed by Wright (1951), where the metapopulation consisted of d demes, each containing a fixed number N of ideal randomly mating individuals. Under the simplest version of this model, each deme contributes a fraction m of its genes to the migrant pool, yielding an expected migration rate from any deme to any other of $m/(d-1)$. If these rates are kept constant, an equilibrium is eventually reached, at which point the increase in population divergence resulting from within-deme genetic drift is balanced by the exchange of alleles by migration. A remarkable feature of this model is that the equilibrium amount of within-population genetic variation is independent of the migration rate, a feature known as the **geographic invariance principle**. Provided there is some potential migratory route between all demes, regardless of the level of migration, the mean coalescence time between random pairs of genes within demes is $2Nd$ generations, i.e., twice the sum of the demic effective population sizes (Li 1976; Strobeck 1987; Nagylaki 2000). On the other hand, the mean coalescence time for two genes randomly drawn from the entire metapopulation is

$$\bar{t}_c = 2Nd + \frac{(d-1)^2}{2dm} \quad (3.26)$$

(Li 1976; Slatkin 1991; Nei and Takahata 1993). Recalling from Chapter 2 that the mean coalescence time for an ideal population is $2N$ generations, Equation 3.26 implies

$$N_e = Nd + \frac{(d-1)^2}{4dm} \quad (3.27)$$

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 $N < 1/(4m)$, N_e exceeds $4Nd$. This confirms Robertson's (1964) supposition that population subdivision can reduce the rate of loss of overall variation by drift as unique alleles are sequestered within individual demes.

Numerous authors have pointed out that the 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 recolonization involves 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**), and that the newly colonized deme is immediately restored to size N in a single generation. This simple modification results in a reduction in N_e for the metapopulation by inducing variation in deme productivity, owing to the fact that demes contributing to a colonization event experience a burst of productivity relative to demes that do not.

Using the logic noted above, Pannell and Charlesworth (1999) derived approximate expressions for mean coalescence times for metapopulations with this type of extinction/colonization behavior. The metapopulation N_e depends on the relative rates of extinction and migration and the size and type of the 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 ($e \ll k/N$),

$$N_e \simeq Nd \frac{[4m + (1/N)]}{4(e + m)} \quad (3.28a)$$

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 $e \rightarrow 0$ unless N is tiny. On the other hand, if the extinction rate is relatively high, such that $e \gg m$ and $e \gg k/N$,

$$N_e \simeq \frac{d}{4e} \quad (3.28b)$$

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 approximation,

$$N_e \simeq \frac{d(1 + 4Nm)}{4m + 2\sigma_K^2(1 + 2m)} \quad (3.29)$$

where σ_K^2 denotes the among-deme variance in the number of gametes contributing to the next generation. Comparing this expression to Equation 3.27, 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 Nd , 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 $1/(2N)$. 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 the allele frequencies at the locus under consideration as well as those elsewhere in the genome are immune to selective processes. This is of course unrealistic, and we will now see that selection will generally cause a further reduction in N_e by inflating the among-family variance in offspring production. Evaluating the magnitude of such effects is rendered difficult by the fact that unlike family-size variation induced by environmental heterogeneity, which can be erased in a single generation, variation associated with selection will be sustained across generations if it has a genetic basis. Such heritable transmission will elevate the genetic representation of some individuals in future generations relative to chance expectations. The effects of selection 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 to elimination or fixation.

Before proceeding, it must be emphasized that because N_e is defined in the context of stochastic processes operating on hypothetically neutral loci, which serve as benchmarks for the pure drift process, our concern here is not so much with the loci under selection, but with the effects of such selection on allele-frequency dynamics elsewhere in the genome. The issues here are quite technical, as 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. Our goal is to simply provide some heuristic understanding of why selection is expected to almost always lead to a substantial reduction in N_e , well beyond that expected on the basis of demographic arguments alone.

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 (our previous σ_k^2), in the first generation of selection there will also be an among-family variance in relative fitness associated with selection itself, σ_w^2 . Here the relative fitness of the i th family (w_i) is simply the expected contribution to the next generation relative to the average in the population, such that $\sum_{i=1}^{N/2} w_i/(N/2) = 1$, where we assume a balanced sex ratio and $N/2$ families. Assuming large N and features other than selection in accordance with the standard Wright-Fisher model, a single generation of selection will reduce the effective population size to

$$N_e \simeq \frac{4N}{\sigma_k^2 + 2 + 4\sigma_w^2} \quad (3.30a)$$

which is identical in form to Equation 3.4, except for the additional variance associated with selection in the denominator. The form of this expression demonstrates

that the stochastic associations of neutral alleles with families with different genetic endowments initially has the same qualitative effect as stochastic associations with microenvironmental differences.

Robertson (1961) had the key insight that with subsequent generations of selection, new stochastic associations will arise while old associations will be lost at rate 0.5 with free recombination, leading to a long-term cumulative variance in the contribution to future generations proportional to $[1+(1/2)+(1/4)+(1/8)+\dots]^2\sigma_w^2 = 4\sigma_w^2$. However, this result ignores the fact that the effects of selection will dissipate over time as the favorable alleles at the loci under selection spread through the population, thereby leading to lower background variation. Letting L denote the per-generation fractional loss of additive genetic variance at selected loci, the long-term effective population size becomes

$$N_e \simeq \frac{4N}{\sigma_k^2 + 2 + 4[2/(1+L)]^2\sigma_w^2} \quad (3.30b)$$

(Santiago and Caballero 1995). If we further assume that variation in family sizes unassociated with selection simply reflects binomial sampling, then from above $\sigma_k^2 \simeq 2$, and under Robertson's assumption of $L = 0$, the long-term effective size becomes

$$N_e = \frac{N}{1 + 4\sigma_w^2} \quad (3.30c)$$

Equations 3.30b,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. The remainder of this chapter is focused on the removal of this limitation.

Example 3.5. The genetic variance for the relative fitness of family means is a function of the intensity of selection and the heritability of the selected traits, and in the case of truncation selection on a single trait, $\sigma_w^2 = i^2 t_{FS}$, 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 and no maternal effects (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, 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.

Strong linkage will necessarily increase the influence of selection on N_e by maintaining stochastic associations of alleles at unselected loci for longer periods. The issues have been explored most closely in theoretical studies of the influence of selection on standing levels of molecular heterozygosity at linked neutral markers.

The consequences of two forms of selection have been considered: the fixation of beneficial mutations, and the purging of deleterious mutations.

A simple way of evaluating the consequences of the fixation of stochastically arising 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 diploid Wright-Fisher 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 a stochastic rate δ per generation. Because such mutations arise independently of the allele at the linked neutral locus, the resultant **selective sweeps** will result in the fixation of neutral alleles with probabilities proportional to their current frequencies, in this case p and $(1-p)$ for the two alternative allelic types. 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 at completely linked loci, 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 $1/(2N_e)$, we obtain

$$N_e \simeq \frac{N}{1 + 2N\delta} \quad (3.31a)$$

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 on demographic considerations alone.

When selective sweeps are rare relative to the power of random genetic drift, such that $\delta \ll 1/(2N)$, $N_e \simeq N$, but as $N \rightarrow \infty$, $N_e \rightarrow 1/\delta$, 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** (the stochastic result of hitch-hiking effects that inevitably arise in linked genomes). That is, when N is large, the effective size of a population can be more strongly influenced by the physical 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, in which case strong linkage could lead to a potential situation in which N_e eventually scales negatively with absolute population size (Lynch 2007).

If a significant amount of recombination occurs between a neutral marker and the selected locus while the latter is proceeding toward fixation, the selective sweep is not expected to completely remove the variation at the marker locus. The extent to which a neutral locus can free itself of stochastic associations with newly arising beneficial mutations will depend 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). The technical issues, worked out by Wiehe and Stephan (1993), lead to an expression identical in form to Equation 3.31a, with $2N\delta$ being replaced by a term that is smaller in absolute value. However, Gillespie (2000) also

derived a simple expression showing that recombination modifies Equation 3.31a to

$$N_e \simeq \frac{N}{1 + 2NE(p_s^2)\delta} \quad (3.31b)$$

where $E(p_s^2)$ is the expected homozygosity of the descendant alleles derived from the single neutral marker initially associated with the beneficial mutation following the completion of the selective sweep. The initial frequency of the associated marker is always $1/(2N)$, as we are referring here to the specific gamete in which the beneficial mutation occurred. Thus, with free recombination $E(p_s^2) \simeq 0$, but with complete linkage, $E(p_s^2) = 1$, returning us to Equation 3.31a. Although the precise mechanistic determinants of p_s are not made clear, Equation 3.31b has an appealingly simple interpretation, and Santiago and Caballero (2005) showed that

$$E(p_s^2) \simeq (2.6Ns)^{-2(r/s)} \quad (3.32)$$

We now turn to selection against recurrently appearing deleterious mutations, which again cause a reduction in N_e as a consequence of induced variation in family size. However, contrary to the situation with selective sweeps, whose effects are expected to be sporadic and chromosomally restricted in scope, the effects of recurrent deleterious mutations will be persistent across the entire genome. This is because deleterious mutations are distributed over large numbers of loci (LW Chapter 10), with the overall load of such mutations being maintained by a balance between introduction by mutation and removal by selection and drift. Here we attempt to provide a heuristic understanding of the effects of such **background selection**, again by considering separately the effects of unlinked and linked deleterious mutations, relying on a simple model in which loci undergoing selection 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 fitness in heterozygotes equal to s , the equilibrium allele frequency of the deleterious allele is equal to u/s (provided s is substantially stronger than the power of drift and mutation) (LW Chapter 10). The genetic variance in relative fitness associated with this locus then has an expected value close to $2us$. This result can be obtained by assuming a negligible frequency of mutant homozygotes (owing to the fact that u/s is small), so that the population effectively consists of just two classes of individuals: homozygotes and heterozygotes having relative fitnesses close to $(1 + 2u)$ and $(1 - s + 2u)$, respectively, using the fact that mean absolute fitness is $\sim (1 - 2u)$ and $2u \ll 1$. Summing over all n loci capable of mutating to deleterious alleles, the total genetic variance in fitness among individuals is $2nus = Us$, where U 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 unlinked pairs will be $< 1/x$ because genes located on opposite ends of chromosomes are effectively unlinked). Thus, we can make use of Robertson's equation (3.30c), 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 would cause a relatively small reduction in N_e ,

as

$$N_e \simeq \frac{N}{1 + 2Us} \simeq Ne^{-2Us} \quad (3.33)$$

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). Equation 3.33 provides an inadequate description of the overall effect of background selection as it does not account for the fact that although most pairs of loci are unlinked, every locus is nevertheless embedded within a region of restricted recombination. Some impression of the impact of linkage can be acquired by returning to the logic used to obtain Equation 3.30b,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. Letting the recombination rate between loci be r , this expression generalizes to $\sum_{i=0}^{\infty} r^i = 1/r$ generations. Thus, a single selected locus is expected to reduce the variation at a linked neutral locus such that

$$N_e \simeq Ne^{-us/r^2} \quad (3.34)$$

(Barton 1995). The challenge is to determine the joint effects of the full spectrum of linked and unlinked loci by accounting for the summed effects of terms of the form e^{-us/r^2} over the full range of selected loci surrounding the reference neutral locus.

Some insight into the cumulative power of selection on linked deleterious mutations can be gleaned by first considering the extreme case of a completely nonrecombining, but otherwise sexual, genome, i.e., allowing 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 mutations in a gamete is $nq = U/(2s)$, where U is again the deleterious mutation rate per diploid genome. Under the assumptions of the model (large enough N_e that deleterious mutations do not go to fixation), only those gametes that are free of deleterious mutations can contribute to the future genetic constitution of the population, and 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 deleterious mutation-free gametes is simply $e^{-U/(2s)}$. Drawing from this observation, Charlesworth et al. (1993) reasoned that with complete linkage, background selection against segregating deleterious mutations will reduce the effective population size to

$$N_e = Ne^{-U/(2s)} \quad (3.35)$$

This shows that in a nonrecombining population, background selection has the potential to cause a dramatic reduction in N_e . 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 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

$$N_e \simeq Ne^{-U/(2s+R)} \quad (3.36a)$$

where R denotes the rate of recombination between the ends of the region. Because $(2s+R) < 1$, the absolute value of the exponent is much larger than that obtained for freely recombining loci ($2Us$ from Equation 3.33), showing that the total contribution from interference from unlinked loci (which is embedded in Equation 3.36a) is relatively minor compared to that from linked loci. Moreover, because s is expected to be quite small, and for an entire chromosome, $R \gg s$, Equation 3.36a can be roughly approximated as

$$N_e \simeq N e^{-U/R} \quad (3.36b)$$

where U/R 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 as in the case of the mutation load on the mean fitness of a sexually reproducing population (LW Chapter 10), the impact of segregating deleterious mutations on N_e is largely independent of the mutational effect s . As we will see 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/R is also estimable.

It should be appreciated that natural populations are subject to both positive and negative selective forces, so the total influence of selection on N_e needs to take into consideration 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 essentially continuously, as a consequence of the recurrent introduction of deleterious alleles, the depressive effects of both forms of selection may be treated as largely independent processes. The reduction in N_e resulting from background selection can then be obtained by use of one of the above expressions, e.g., Equation 3.36b as a first-order approximation for a sexual population (with N already taking into consideration demographic effects). The effective population size dictated by these background processes then defines the background N_{ed} within which occasional beneficial mutations arise, allowing the effective population size resulting from the joint operation of all effects to be approximated by substituting N_{ed} for N in Equation 3.31b. In general, it appears that the joint operation of background selection and selective sweeps will reduce N_e more than either does alone, although it is at least principle possible for background selection to reduce the influence of selective sweeps in regions of very low recombination, by diminishing the probability of fixation of beneficial mutations. The simultaneous operation of positive selection on multiple loci (which was ignored in the derivation of Equations 3.31a,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).

A few potential conditions exist under which selection may actually promote a mild increase in N_e . For example, Santiago and Caballero (2005) 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. The magnitude of this effect requires a special balance of recombination, selection, and migration rates. In addition, Pálsson and Pamilo (1999) found that with very strong linkage and a

low efficiency of selection ($2Ns < 0$), repulsion disequilibrium can build-up between simultaneously segregating deleterious mutations, leading to associative overdominance (LW Chapter 10) and an elevation of N_e . It remains unclear whether the special conditions necessary for these outcomes are very common.

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. For example, as noted above, demographic factors alone can depress N_e by 90% relative to N , with fluctuating population size causing a further 90% decline, and selection still more. 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. These points will be made clearer in the next chapter as we explore the manifestation drift and draft effects on standing variation in natural populations.

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., 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. Pop. Biol.* 23: 85–109. [3]
- Crow, J. F. 1954. Breeding structure of populations. II. Effective population number. In O. Kempthorne, T. A. Baneroff, 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. Pop. 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. Pop. Biol.* 11: 60–80. [3]
- Gillespie, J. H. 2000. Genetic drift in an infinite population. The pseudohitchhiking model. *Genetics* 155: 909–919. [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]
- Hedgcock, 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. Pop. 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.
- 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]
- Li, W.-H. 1976. Distribution of nucleotide differences between two randomly chosen cistrons in a subdivided population: the finite island model. *Theor. Popul. Biol.* 10: 303–308. [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]
- Milkman, R. 1978. Selection differentials and selection coefficients. *Genetics* 88: 391–403. [3]
- Nagylaki, T. 1995. The inbreeding effective population number in dioecious populations. *Genetics* 139: 473–485. [3]
- Nagylaki, T. 2000. Geographical invariance and the strong-migration limit in subdivided populations. *J. Math. Biol.* 41: 123–142. [3]
- Nei, M., and N. Takahata. 1993. Effective population size, genetic diversity, and coalescence time in subdivided populations. *J. Mol. Evol.* 37: 240–244. [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. Pop. Biol.* 12: 253–262. [3]
- Slatkin, M. 1991. Inbreeding coefficients and coalescence times. *Genet. Res.* 58: 167–175. [3]
- Strobeck, C. 1987. Average number of nucleotide differences in a sample from a single subpopulation: a test for population subdivision. *Genetics* 117: 149–153. [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. Pop. 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 Biométrie 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. 1951. The genetical structure of populations. *Ann. Eugenics* 15: 323–354. [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.

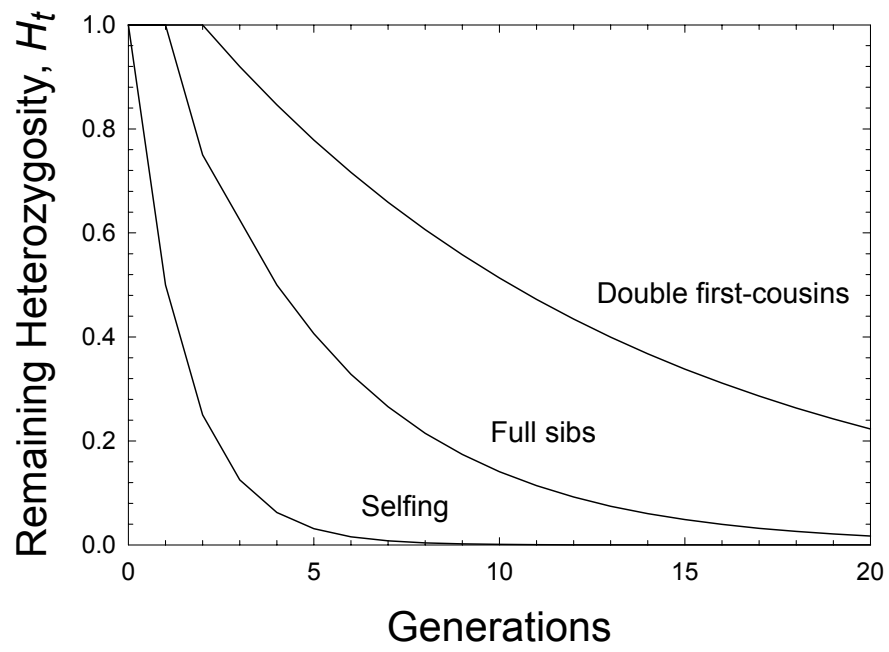
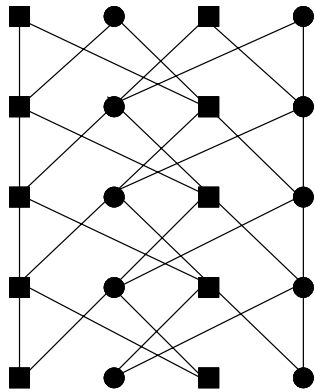


Figure 3.2. Mating schemes under continual double first-cousin mating (left) and circular mating (right), in both cases with four individuals per generation. Genes identical-by-descent do not appear in the same individual for three and four generations, respectively, under these two schemes. Males and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

Maximum Avoidance
of Inbreeding



Circular
Mating

