

2

INBREEDING, FINITE POPULATION SIZE, AND GENE FREQUENCY CHANGE

Rough Draft Version 31 July 1998

©1998 B. Walsh and M. Lynch

Please report errors to jbwalsh@u.arizona.edu

Mating is rarely, if ever, completely random in natural populations. The propagules of many species settle down in close proximity to their site of birth, and when mature, tend to mate with other nearby, somewhat-related individuals. This is supported by a number of investigations that have used genetic markers to track pollen flow in plants. In most herbaceous species, the vast majority of successful pollinations occur within 10 meters of the donor plant (Figure 1). This is almost always a much smaller area than the total range of the population. Effective gene flow away from a parent plant can be facilitated by self-incompatibility systems and by elevated mortality of seeds and seedlings near their parental source (Levin and Kerster 1974, Levin 1984). Nevertheless, there is little question that the assumption of random mating often made in population-genetic theory is violated to some degree in all populations. This tendency to mate with individuals with similar genotypes, no matter how minor, is referred to as inbreeding.

The spatial structure of populations is only one of several factors influencing the extent of inbreeding. Even if mating is completely random, there will still be some tendency to mate with relatives since all populations are finite in size. The smaller the population size, the greater this tendency will be. Thus, in a dioecious population with a stable size of two, all matings must be between full-sibs, even though the reproductive pair itself may be random. It follows that the genetic consequences of finite population size must be similar to those of inbreeding.

Inbreeding and finite population size have important effects on gene and genotype frequencies. In small randomly mating populations, heterozygosity is lost as gamete sampling causes allele frequencies to drift towards zero or one. Such random genetic drift of gene frequencies is of negligible importance in very large populations, but the level of homozygosity can nevertheless be inflated by consanguineous mating. Subdivision of a population into partially isolated demes also results in an increase in the variation in gene frequency between subgroups due to sampling error. The greater the degree of isolation of the subgroups, the more pronounced this differentiation will be.

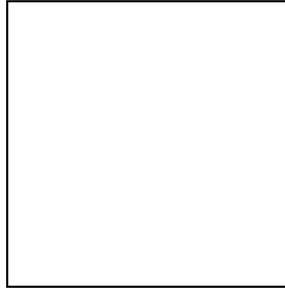


Figure 1. Effective gene flow via pollen for four species of herbaceous plants. In each study, visible or electrophoretic markers were scored in the progeny of a large number of plants surrounding a central (marked) individual. These plots do not completely describe “pollen shadows,” which are usually asymmetrical.

This chapter focuses on the aspects of population structure that influence the rate and magnitude of gene frequency change in finite populations. For the time being, it will be assumed that the allelic variation under consideration is not subject to selection. While this is clearly an overly simplistic view of the evolutionary process, there are important reasons for developing simple evolutionary models in which selection is not a force. First, selective forces that are very effective in large populations may be almost completely overwhelmed by random drift in very small populations (Chapter 27). In that case, models without selection may closely approximate reality. Second, neutral models for the behavior of gene frequencies provide null hypotheses for studies of natural selection. If one wishes to ascribe the observed variation within and between populations to adaptive evolutionary change, it is important to know what to expect in the absence of selection.

Much of the fundamental theory regarding the change in gene frequency distributions resulting from inbreeding and drift was derived by Wright in the first half of this century. Provine (1986) provides an excellent historical overview of this work and its relationship to Wright’s shifting-balance theory of evolution.

Increase in Homozygosity with Inbreeding

In Chapter 11, the inbreeding coefficient f was defined as the probability that two alleles at a locus in an individual are identical by descent. There is another useful way of thinking about the inbreeding coefficient. Consider a single locus with two alleles with frequencies p and q in the base population. In the absence of inbreeding and random drift, the expected frequency of heterozygotes in all generations is $2pq$ under the Hardy-Weinberg law. Suppose, however, that an individual has an inbreeding coefficient f due to the fact that some of its ancestors were related. What is the probability that this individual is a heterozygote at the locus under consideration? The individual can only be a heterozygote if it carries alleles that are not identical by descent (ignoring mutations), the probability of which is $(1 - f)$. If the two alleles are not identical by descent, they must have been acquired independently, so the probability that the genotype is a heterozygote is $2pq$, again by the Hardy-Weinberg law. The individual is therefore a heterozygote with probability $2pq(1 - f)$. Relative to the base population, it is only $(1 - f)$ as likely to be heterozygous. This argument applies regardless of the initial gene frequency and the initial number of alleles at the locus. Thus, f may be viewed as the reduction of an individual’s heterozygosity, averaged over all loci, relative to that expected had mating in all previous generations been random.

In natural populations, individuals vary in the degree of inbreeding due to the many

possible types of mating between relatives that may have occurred in the past. Nevertheless, one can envision (and implement) systems of mating that involve fixed relationships in which all members of the population have the same expected inbreeding coefficient. As first pointed out by Wright (1921b), such mating systems are analytically tractable, and a consideration of them helps clarify the rapidity with which inbreeding can lead to the loss of genetic variation within populations.

Consider first the most extreme form of inbreeding — obligate self-fertilization, a mode of reproduction in some plants and hermaphroditic animals. Since all self-fertilizing lines are reproductively isolated under this scheme of mating, a population consists of a single individual. If the individual is a heterozygote (**Bb**), it will produce equal frequencies of **BB**, **Bb**, **bB**, and **bb** progeny. Thus, an offspring of a self-fertilizing heterozygote has only a 50% chance of being a heterozygote. This applies to any heterozygous locus in any generation. Consequently, after t generations of obligate self-fertilization, the heterozygosity is reduced to a fraction

$$1 - f(t) = (1/2)^t \tag{1.1}$$

of its initial level. After $t = 10$ generations, only 0.1% of the initial heterozygosity remains (Figure 2).

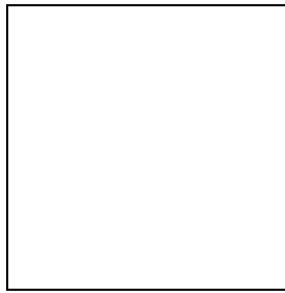


Figure 2. The fraction of initial heterozygosity remaining after enforced self-fertilization, brother-sister mating, and double first-cousin mating.

The next most intense system of inbreeding is 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

$$1 - f(t) \simeq (0.81)^t \tag{1.2a}$$

Thus, starting from a non-inbred base population, 12 generations of full-sib mating are sufficient for the loss of 90% heterozygosity. The exact recursion formula for $f(t)$ under full-sib mating is

$$f(t) = \frac{1}{4}[1 + 2f(t - 1) + f(t - 2)] \tag{1.2b}$$

With a constant population size of four breeding adults, the minimum relationship between individuals is that of double-first cousins (Figure 3). Starting with four unrelated individuals, it takes three generations for the appearance of alleles identical by descent in the same individual, and thereafter

$$1 - f(t) \simeq (0.92)^t \tag{1.3}$$

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

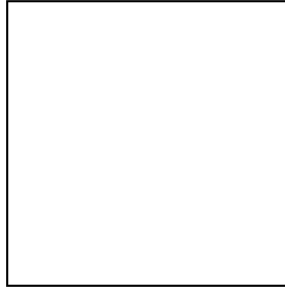


Figure 3. Mating schemes under continual double first-cousin mating (left) and under circular mating with four individuals (right). Genes identical-by-descent do not appear in the same individual for three and four generations respectively.

The main point of these three examples is that the rate of loss of heterozygosity caused by inbreeding can be greatly reduced by decreasing the relatedness between members of mating pairs: from 50% to 19% to 8% loss/generation for obligate self-fertilization, full-sib mating, and double first-cousin mating. The expansion of a random mating population has similar consequences (below).

For large populations with complex pedigrees that extend back several generations, the inbreeding coefficient should be computed for each individual. Computer algorithms developed for these purposes are used extensively by animal breeders (Boyce 1983, MacCluer et al. 1983). Unfortunately, the investigator of natural populations usually has little information on individual pedigrees. We shall see below that rough estimates of the average degree of inbreeding are still attainable provided one has information on the population size and mating system.

Before proceeding, a small clarification needs to be made about the inbreeding coefficient. When f is defined as the depression of heterozygosity relative to the Hardy-Weinberg expectation, it may sometimes take on negative values, which implies an inflation of heterozygosity. This can occur if there is disassortative mating with respect to genotype. When the sexes are separate, a slight inflation of heterozygosity is expected even under random mating due to differences in gene frequency between the sexes that result from gamete sampling (Robertson 1965), but this is of negligible significance unless the population size is very small (< 50). These types of results are consistent with the definition of f as the correlation between gametes — $f < 0$ implies a negative correlation, i.e., negative assortative mating.

Finite Population Size and Random Genetic Drift

Populations produce a finite number of gametes each generation, and a finite number of these are drawn to produce the next generation's zygotes. Thus, sampling error usually will insure that the gene frequency in the offspring generation deviates somewhat from that in the parent generation. There are two ways to conceptualize the sampling variance of gene frequency. On the one hand, it can be viewed as the variance in gene frequency among replicate offspring populations of the same size, derived from the same parent population. Alternatively, it describes the variance in gene frequency in a specific offspring

population among all loci with gene frequency p in the parental generation. The implication of both viewpoints is the same. In the absence of any counteracting evolutionary forces, the dispersive effects of genetic drift will continue each generation until all loci have become fixed for a single allele. The consequences of random drift are therefore two-fold: 1) the genetic variation within populations is gradually lost as gene frequencies drift to zero or one, and 2) isolated populations diverge as they become fixed for alternate alleles.

To put things on a more formal basis, consider a large pool of gametes, a fraction p of which contain the **B** allele. $2N$ gametes are randomly drawn to produce a new generation of size N . The expected frequencies of genotypes **BB**, **Bb**, and **bb** are, from the Hardy-Weinberg expectation, p^2 , $2p(1-p)$, and $(1-p)^2$. The expected number of **B** alleles contained in any offspring is simply $2 \times p^2 + 1 \times 2p(1-p) = 2p$, but the expected square of the number of **B** alleles carried is $2^2 \times p^2 + 1^2 \times 2p(1-p) = 2p(1+p)$. The variance in the number of **B** alleles carried by each individual is therefore $2p(1+p) - (2p)^2 = 2p(1-p)$, while the variance for the total number of **B** alleles carried by the N offspring is $2Np(1-p)$. Since $2N$ genes have been drawn, the variance of the frequency of allele **B** is $2Np(1-p)/(2N)^2 = p(1-p)/2N$. Thus, the sampling variance of gene frequency is directly proportional to the heterozygosity and the reciprocal of population size.

The expression $p(1-p)/2N$ only defines the dispersion that results from a single generation of gamete sampling, conditional on gene frequency p in the parent population. A fuller account of the long-term consequences of genetic drift requires an alternate approach. The theory will first be described for an idealized situation: 1) a monoecious, randomly mating population that produces an effectively infinite number of gametes, 2) with non-overlapping generations, 3) a complete absence of evolutionary forces such as selection, migration, and mutation, and 4) constant population size. Later in the chapter, it will be shown how the idealized model can be generalized to incorporate different systems of mating.

Long-term loss of genetic variation within populations. Even when mating is completely random, there is always a small chance that uniting gametes will derive from related individuals. For example, in a randomly mating, monoecious population containing only two individuals, there are only four possible genes at each locus. Therefore, the probability that a gamete will randomly unite with another containing a direct copy of the same gene is $1/4$. With four individuals, there are eight genes, and this probability becomes $1/8$. Thus, in the idealized situation, the probability that two direct copies of any parental gene will randomly unite in an offspring is $1/2N$.

Although the quantity $1/2N$ may be thought of as the new inbreeding that is incurred each generation, this does not fully describe the build-up of homozygosity in a population. Even if uniting gametes do not carry genes that are direct copies of a parental gene, they may still be identical-by-descent because of inbreeding in a previous generation. Under random mating, the probability of this event is simply the inbreeding coefficient of the parental generation. Thus, since the probability of drawing genes that are not direct copies of the same parental gene is $(1 - 1/2N)$, the expected inbreeding coefficient in generation t is

$$f(t) = \frac{1}{2N} + \left(1 - \frac{1}{2N}\right) f(t-1) \tag{1.4}$$

Subtracting both sides from 1, this simplifies to the recursion formula

$$1 - f(t) = \left(1 - \frac{1}{2N}\right) [1 - f(t-1)] \tag{1.5a}$$

which generalizes to

$$1 - f(t) = \left(1 - \frac{1}{2N}\right)^t [1 - f(0)] \tag{1.5b}$$

Note that as $t \rightarrow \infty$, $[1 - f(t)] \rightarrow 0$ at a rate that is inversely proportional to population size. Individuals are expected to become completely inbred at loci that are unmodified by selection, mutation, and migration.

It is important to bear in mind that the equations given above predict the *expected* changes in f resulting from inbreeding. Because of variation in pedigree structure, the actual degree of inbreeding generally will vary between loci within individuals as well as between individuals within the population. For any locus, identity by descent is binomially distributed with mean f and variance $\sigma_f^2 = f(1 - f)$. With completely linked loci, this is also the total variance in f since there is no variation in f between loci. However, for unlinked loci, the realized inbreeding at each locus need not be the same. The coefficient of variation of $(1 - f)$ 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 (Weir et al. 1980). 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. Further information on this subject may be found in Jacquard (1975), Franklin (1977), and Cockerham and Weir (1983b).

Recall that $[1 - f(t)]$ is the expected heterozygosity (H_t) relative to that in the base population (H_0). From Equation (1.5) we see that after t generations at a constant population size N ,

$$H_t = \left(1 - \frac{1}{2N}\right)^t H_0 \quad (1.6)$$

where H_0 is the initial heterozygosity. See Bulmer (1980; p. 220) for an expression for the variance of H_t . This rate of decay of heterozygosity of $1/2N$ was first obtained by Wright (1931) using a rather different approach. It may be of interest to the non-mathematically inclined that Fisher, an excellent mathematician, obtained the wrong answer in 1922.

The time course for the loss of heterozygosity can be clarified by considering the exponential approximation to Equation 1.6. Since $(1 - x)^t \simeq \exp(-xt)$ for $|x| \ll 1$, for $N > 10$,

$$H_t \simeq H_0 \exp(-t/2N) \quad (1.7)$$

This can be solved to show that the heterozygosity is reduced to half of H_0 in about $1.4N$ generations and to 5% of H_0 in about $6N$ generations. If the population size is variable, Equation 1.6 becomes

$$H_t = H_0 \prod_{i=1}^t \left(1 - \frac{1}{2N_i}\right) \quad (1.8)$$

where the \prod sign denotes a product of terms. This expression illustrates an important point. Each of the terms, $[1 - (1/2N_i)]$, is necessarily less than one. Thus, an expansion of population size can only reduce the rate of erosion of heterozygosity; it cannot eliminate it.

Development of between-population variance. A natural consequence of gene frequency drift within populations is the divergence of isolated replicate populations. Suppose a monoecious base population with gene frequency p_0 is suddenly split into several completely isolated subpopulations each of size N with random mating within each subpopulation and an absence of selection, migration, and mutation. The variance in gene frequency in generation t is

$$\sigma_p^2(t) = E(p_t^2) - E^2(p_t)$$

Adding and subtracting $E(p_t)$,

$$\begin{aligned}\sigma_p^2(t) &= [E(p_t) - E^2(p_t)] + [E(p_t^2) - E(p_t)] \\ &= E(p_t)[1 - E(p_t)] - E[p_t(1 - p_t)]\end{aligned}$$

Because there are no systematic forces causing the gene frequency to increase or decrease, $E(p_t) = p_0$, and the first quantity on the right is $p_0(1 - p_0)$. The quantity $E[p_t(1 - p_t)]$ is half the expected heterozygosity in a population in generation t . Substituting Equation (1.6),

$$\sigma_p^2(t) = p_0(1 - p_0) \left[1 - \left(1 - \frac{1}{2N} \right)^t \right] \quad (1.9a)$$

which is well approximated by

$$\sigma_p^2(t) \simeq p_0(1 - p_0)[1 - \exp(-t/2N)] \quad (1.9b)$$

for $N > 10$. Thus, the between-population variance asymptotically approaches $p_0(1 - p_0)$, which for a diallelic locus is half the heterozygosity in the base population. As the process of random drift proceeds, gene frequencies may go through many phases of increase and decline, but eventually each gene will be totally fixed or lost. Ignoring new mutations, the probability that an allele is ultimately fixed in a replicate population is its initial frequency p_0 , while its probability of loss is $(1 - p_0)$.

Equation 1.9 does not fully describe the pattern of development of differences between lines. For example, it yields little insight into the actual form of the distribution of population gene frequencies, and it does not specify the probability of fixation of alleles. However, the idea that many allelic variants at enzymatic loci are selectively neutral (Kimura 1983) generated much interest in these issues, and the theory is now well developed.

Kimura (1955) first derived the full probability distribution for the gene frequency among isolated randomly mating populations. The mathematical details are very complex and will not be given here (some of the necessary machinery is discussed in Appendix 29A). Figure 4 captures the essential features by illustrating the temporal dynamics of population differentiation starting with base populations with gene frequencies 0.5 and 0.1. There are two noteworthy properties of these distributions.

First, the total area under the curves declines and the distributions become flatter with increasing time. The flattening of the distribution arises as the gene frequencies drift apart in the isolated populations. The reduction in total probability mass arises as the gene becomes fixed or extinct in local populations. The proportions of populations that have experienced gene fixation or extinction are not illustrated in the figure. Thus, the area under the curve is proportional to the frequency of populations that still contain both alleles.

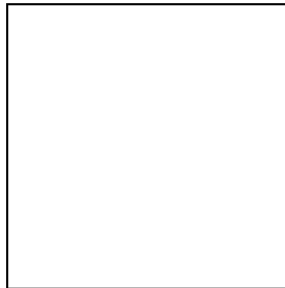


Figure 4. Expected probability distributions for the frequency of neutral alleles in replicate, randomly mating populations of size N after t generations of divergence. The initial gene frequency in the base population is 0.5 on the left and 0.1 on the right. The abscissa is the population gene frequency, while the ordinate is proportional to the probability of occurrence of that frequency. From Kimura (1955).

Second, the expected distribution is a function of t/N generations. That is, the differentiation among populations of size 100 after 10 generations is the same as that among populations of size 1000 after 100 generations or populations of size 10,000 after 1000 generations. This should not be too surprising. The scaling of the temporal dynamics of random genetic drift to the reciprocal of population size has already arisen three times in previous paragraphs: the development of the inbreeding (Equation 1.5) and loss of heterozygosity (Equation 1.6) within populations, and the development of the variance of gene frequency between populations (Equation 1.9).

Buri's Experiment

Since all populations are finite in size, the theory of random genetic drift is of central significance to population genetics. It has played a particularly influential role in the quantitative development of the neutral theory of molecular evolution (Kimura 1983), which purports that the major cause of allelic divergence between populations and species is the random introduction of neutral mutations and their subsequent chance fixation. It may therefore come as a surprise that direct empirical support for the simple drift formulations is almost completely lacking. Highly replicated experiments, which are required to evaluate the statistical aspects of within- and between-population shifts in gene frequency, are very rare.

There is, however, one massive experiment whose results are remarkably consistent with the theory outlined above. Starting with two homozygous lines of *Drosophila melanogaster*, one of which was fixed for allele bw^{75} and the other for allele bw at the brown locus, Buri (1956) established 212 F_1 hybrid populations. For the following 19 generations, he randomly mated 8 males and 8 females within each population and monitored the changes in allele frequency. This could be done visually because the genotype at the brown locus determines the eye color: $\text{bw}^{75}\text{bw}^{75}$ = bright red-orange, bw^{75}bw = deep red-brown, and bwbw = white. (Two separate experiments, one with 107 and the other with 105 populations, were actually performed, but the results are so similar that they have been pooled in the following analysis).

Buri's observations indicate that the bw^{75} and bw alleles were effectively neutral with respect to each other (Figure 5). In the absence of selection, the expected frequency of the bw^{75} allele averaged over all 212 populations is 0.50 in all generations. Nevertheless, just as the frequency within any population is expected to deviate from 0.5 because of drift, so will the mean gene frequency in the total aggregate of populations. The sampling variance of the mean frequency is equal to the sum of the expected within- and between-population variances divided by 212. The latter quantity has already been defined as Equation 1.9 while the former is half the expected binomial sampling variance divided by the sample size ($2N$), or $p_0(1 - p_0)[1 - (1/2N)]^t/2N$. The figure shows that although the frequency of the bw^{75} allele, averaged over all populations, increased to 0.525, it generally remained within two standard errors of the expectation under pure drift. The overall pattern in the change in mean frequency is therefore compatible with the expectations for a neutral locus subject to random genetic drift.

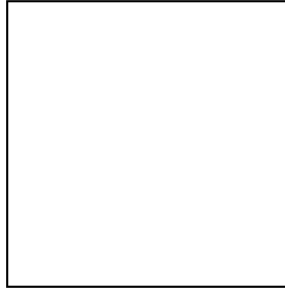


Figure 5. Temporal change in the frequency of the \mathbf{bw}^{75} allele, averaged over 212 isolated populations of *Drosophila melanogaster*, each consisting of 8 breeding males and 8 breeding females. The dashed line denotes the expectation for a neutral allele plus two standard errors. From Buri (1956).

The dynamics of the between-population divergence (Figure 6) are qualitatively very similar to the expected pattern illustrated in the left panel of Figure 4. As the population gene frequencies diverge, the initially bell-shaped distribution becomes flatter and then begins to acquire a U-shape as populations that are fixed for the \mathbf{bw}^{75} or \mathbf{bw} alleles begin to accumulate. Eventually, the distribution would have consisted of only two classes, those fixed for \mathbf{bw}^{75} and those fixed for \mathbf{bw} , in roughly equal frequency.

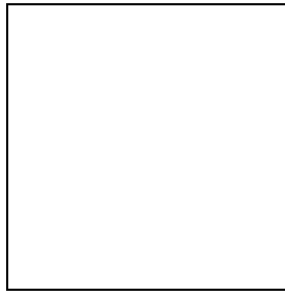


Figure 6. Distribution of the number of \mathbf{bw}^{75} genes in 212 populations of *Drosophila melanogaster* each initiated with a frequency of 0.5. From Buri (1956).

The actual rate of divergence illustrated in Figure 6 is somewhat greater than that expected for randomly mating populations of 16 individuals. However, this does not necessarily invalidate the theory outlined above. It is possible, for example, that not all 16 potential parents reproduced each generation, and/or that the distribution of family sizes deviated from randomness. From the standpoint of genetic drift, either of these conditions would cause the populations to behave *effectively* as though they were smaller than the actual size.

With the massive amount of data in Buri's experiment, it was possible to obtain an empirical estimate of this effective population size. Recall that the sampling variance of gene frequency from one generation to the next is $p(1-p)/2N$. Not including fixed classes, there are 31 possible gene frequencies in Buri's populations ($1/32$ to $31/32$). Each of these 31 classes was observed at various times in one or more of the 212 populations. Focusing on any one class, the sampling variance conditional on p could then be calculated from the gene frequencies observed in the subsequent generation. The 31 points shown in Figure 7 provide an empirical description of the function $p(1-p)/2N$. An excellent fit is obtained if it is assumed that the average effective population size was $N = 10.2$ rather than the idealized

16. In other words, the sampling variance of gene frequency is in very close accord with that expected for an average ideal population of 10.2 randomly mating individuals.

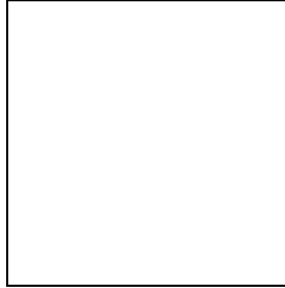


Figure 7. Observed sampling variances of gene frequency for situations in which the donor population contained 1 to 31 *bw*⁷⁵ genes. The dashed line is the expected pattern, $p(1-p)/2N$, if the actual populations of 8 males and 8 females were randomly mating and all had an equal chance of contributing offspring. The solid line describes the pattern for an average effective population size of 10.2. From Buri (1956).

Figure 8 shows that when N is set equal to 10.2 in Equations 1.6 and 1.9, the erosion of average heterozygosity within populations and the build-up of between-population variance of gene frequency are quite consistent with the neutral model.

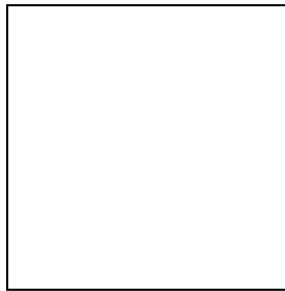


Figure 8. Reduction in the mean heterozygosity within populations and build-up of gene frequency variance between-populations for the brown locus in 212 replicate populations of *Drosophila melanogaster*. Fitted lines are Equations 1.6 and 1.9 with $p_0 = 0.5$ and $N = 10.2$. The expected heterozygosity is 0.5 in generations 1 and 2 because the base population (generation 0) consisted entirely of heterozygotes, and because with separate sexes, an additional generation is required for the unification of alleles that are identical by descent. From Buri (1956).

Avoidance of the Loss of Genetic Variation

The genetic consequences of inbreeding and random genetic drift are of particular concern to managers of small, captive populations of endangered species. Here we consider just one of the many practical questions that arise in this area. 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 variance? In his 1921b paper, Wright suggested that the best way to minimize the build-up of homozygosity in a

small population would be to restrict matings to pairs of individuals with the least degree of relatedness. Such a breeding scheme is known as *maximum avoidance of inbreeding* or MAI and is exemplified by the double first-cousin mating design in Figure 3. 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 $N = 64$ under a maximum avoidance scheme, $m + 1 = 7$ generations would pass before two copies of a founding gene could appear in the same individual. Once the inbreeding begins, the proportion of heterozygosity lost each generation is very nearly constant, as pointed out above for population sizes of 1, 2, and 4. More generally, for $N \geq 4$, the erosion of heterozygosity under MAI is closely approximated by

$$1 - f(t) \simeq \left(1 - \frac{1}{4N}\right)^t \quad (1.10)$$

where t is the number of generations after the onset of inbreeding (Wright 1951). Comparing this expression with Equation (1.5), it can be seen that MAI has the same effect as doubling the size of a random mating population.

Kimura and Crow (1963a) subsequently noted that Wright's intuition that MAI minimizes the long-term loss of genetic variation is not strictly correct. A *circular mating* (CM) scheme (Figure 3) leads ultimately 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." The last individual in the linear array is 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 or more generations before its superiority is realized. Since most of the genetic variation has been lost by this time, the practical utility of circular mating is questionable.

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. Robertson (1964) obtained the more general (and counterintuitive) result 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. Although the individual lines are all expected to become homozygous eventually, different lines are likely to become fixed for different sets of genes. Subsequent crossing of the lines would then restore the population to something close to its original state. It must be emphasized, however, that the preceding arguments assume that intense inbreeding in small lines has no serious selective consequences that might endanger the survival of the lines. In reality, very small lines are likely to die out occasionally just by accident, and extreme inbreeding often has serious deleterious effects on fitness (Chapter 22). Many of the technical details on the dynamics of inbreeding and random genetic drift in subdivided populations can be found in Cockerham (1970).

Effective Population Size

Up to now, we have assumed a randomly mating population, constant in size, and consisting of a homogeneous set of monoecious, self-compatible individuals. Since almost all populations deviate from this ideal structure in some way, the question arises as to the relevance of the preceding theory. 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. Instead of relying on the total number of individuals as a measure of population

size, we construct a surrogate index that takes into account the deviations from the ideal model. Such an index has become widely known as N_e , the *effective population size*, following the early and influential work of Wright (1931, 1938, 1939).

In an ideal monoecious population, each gamete unites randomly with another gamete derived from the total population of N individuals. Thus, the probability that two randomly 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 the reciprocal of the actual population size. To account for this, P can be more generally regarded as the reciprocal of the effective rather than the actual population size.

To see the connection between N_e and the inbreeding in a population more clearly, recall the approach to predicting the future inbreeding coefficient in a 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 (identical by descent) 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 in which identity-by-descent can arise between uniting gametes,

$$\begin{aligned} f(t) &= \left(\frac{1}{N_e}\right) \left(\frac{1}{2}\right) + \left(\frac{1}{N_e}\right) \left(\frac{1}{2}\right) f(t-1) + \left(1 - \frac{1}{N_e}\right) f(t-1) \\ &= \frac{1}{2N_e} + \left(1 - \frac{1}{2N_e}\right) f(t-1) \end{aligned} \quad (1.11)$$

Notice that this expression is identical in form to Equation 1.4. Thus, the effective population size may be viewed as the size of an ideal population that would exhibit the same amount of inbreeding as the population under consideration. For reasons to be discussed below, the effective population size is almost always less than the actual size.

The effective population size is one of the most important parameters in population genetics. It enters almost all expressions for the dynamics and equilibrium states of gene frequencies in finite populations. While N_e is not as easily measured in natural populations as the total population size, it is at least in principle obtainable from observable demographic properties of populations (Latter 1959, Lande and Barrowclough 1987, Crow and Denniston 1988), as we now demonstrate.

Monoecy. In order to illustrate the mathematical approach to derive expressions for N_e , we first generalize the monoecious, self-compatible population to allow arbitrary numbers of gametes to be contributed by different members of the population. 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 are the mean and variance of successful gamete production per individual, and N_{t-1} is the number of reproducing parents. Assuming that mating is random and *isogamous* (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. Hence there are $\sum_{i=1}^{N_{t-1}} k_i(k_i - 1)$ total ways in which gametes can come from the same parent. In total, $N_{t-1}\mu_k$ gametes are produced, so the probability that zygotes in generation t contain gametes derived from the same parent is

$$P_t = \frac{1}{N_e} = \frac{\sum_{i=1}^{N_{t-1}} k_i(k_i - 1)}{N_{t-1}\mu_k(N_{t-1}\mu_k - 1)} \quad (1.12)$$

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 $N_{t-1}\mu_k = 2N_t$. Substituting into Equation 1.12 and inverting,

$$N_e = \frac{2N_t - 1}{(\sigma_k^2/\mu_k) + \mu_k - 1} \quad (1.13)$$

Thus, for a randomly mating monoecious population with discrete generations, the effective population size is a function of the actual population size and the mean and variance of successful gamete production. In principle, all three quantities can be estimated in natural populations.

The above expression simplifies greatly under a number of conditions. For example, for populations that are stable in size ($\mu_k = 2$),

$$N_e = \frac{4N_t - 2}{\sigma_k^2 + 2} \quad (1.14)$$

Consider also the situation in which each parent produces the same number of potential gametes. Since the variance in the number of gametes of a particular parent for each gamete drawn is $(1/N_{t-1})[1 - (1/N_{t-1})]$ (from the properties of a binomial distribution), and since a total of $\mu_k N_{t-1}$ gametes are drawn, $\sigma_k^2 = \mu_k[1 - (1/N_{t-1})]$. Substituting into Equation 1.14, it is found that $N_e = N_{t-1}$. This is to be expected since the conditions assumed are identical to those of the idealized model.

Many hermaphroditic species are self-incompatible 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 generation $t - 1$ (the 2 since each individual can serve as a mother or father). Since 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 since each parent carries two genes). Therefore, the probability of drawing a pair of genes from the same grandparent is

$$P_t = \frac{\sum_{i=1}^{N_{t-2}} k_i(k_i - 1)}{N_{t-2}\mu_k(N_{t-2}\mu_k - 2)} \quad (1.15)$$

assuming that the fertility of parents and offspring are uncorrelated. Employing the same substitutions used for Equation 1.13,

$$N_e = \frac{2(N_{t-1} - 1)}{(\sigma_k^2/\mu_k) + \mu_k - 1} \quad (1.16)$$

Note that for populations that are moderately large and stable in size, Equations 1.13 and 1.16 give essentially the same answer — the prohibition of selfing has a negligible influence on N_e . 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.

In many hermaphroditic species, there is a distinction between male and female gametes (*anisogamy*). When mating is random but selfing is prohibited, the effective population size is the same under isogamy and anisogamy, i.e., Equation 1.16 still applies (Crow and Denniston 1988). However, with selfing permitted,

$$N_e = \frac{N_{t-1}}{(4\sigma_{ep}/\mu_k^2) + 1} \quad (1.17)$$

where σ_{ep} is the covariance of the numbers of successful male and female 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 many more resources than others, the effective population size will be less than the observed size provided the population size is stable ($\mu_k = 2$). 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.

All of the preceding examples indicate that variance in successful gamete production is a major determinant of the effective population size. An increase in the variance of family size causes a reduction in N_e because of the enhanced likelihood of uniting gametes coming from the same prolific parent in subsequent generations. When the variance in family size is entirely a function of the random sampling of gametes (so that $\sigma_k^2 = \mu_k$) and the population size is stable ($\mu_k = 2$), the effective population size is essentially equivalent to the number of reproductive adults. It may be more surprising that in the opposite and extreme situation in which all parents produce an identical number of progeny ($\sigma_k^2 = 0$), $N_e = 2N_{t-1} - 1 \simeq 2N_{t-1}$, i.e., elimination of the variance in family size doubles the effective population size. It was shown earlier that the MAI scheme of mating results in a rate of loss of heterozygosity of about $1/4N$. This can now be seen to be a consequence of the MAI mating scheme causing every parent to produce the same number of offspring.

In natural populations where individuals grow up in different microenvironments, which influence the availability of resources and mates, σ_k^2 will usually exceed the mean. Thus, generally we can expect the effective population size to be less than the number of reproductive adults.

Example 1. Heywood (1986) has estimated σ_k^2/μ_k^2 for *seed* production to be on the order of 1 to 4 in a number of annual plants. Unfortunately, the value of σ_k^2 for *total* gamete production requires additional information on successful pollen production. Such information is extremely difficult to acquire due to problems in ascertaining paternity. For heuristic purposes, however, let us assume a stable monoecious population (μ_k for seed production = μ_k for pollen production = 1), a three-fold higher standard deviation for successful pollen production, and a perfect correlation between seed and pollen production, i.e., $1 = \sigma_{ep}/[\sigma_{k(\text{seed})} \cdot 3\sigma_{k(\text{seed})}]$. Assuming a stable population size, what is N_e ? Substituting $\sigma_{ep} = 3\sigma_{k(\text{seed})}^2$ and $\mu_k = 2$ into Equation 1.17, $N_e = N/(3\sigma_{k(\text{seed})}^2 + 1)$. For the cases in which $\sigma_{k(\text{seed})}^2 = 1$ and 4, N_e is 25% and 8% of the census number.

t

Dioecy. As in the case of monoecy with self-incompatibility, when the sexes are separate, the new inbreeding needs to be defined with reference to the grandparents since this is the earliest generation back to which the two genes of an individual can be traced to the same ancestor. An additional complication that arises with separate sexes is the possibility that the inbreeding through males and females may differ. This is expected, for example, in polygynous species in which most females mate with a relatively small segment of the male population. It is also expected in species with skewed sex ratios.

If O is the individual of interest and X and Y its mother and father, then there are two ways in which O may derive two genes from the same grandparent: 1) X and Y may share the same mother (probability $1/N_{ef}$ where $1/N_{ef}$ is the effective number of females), or 2) X and Y may share the same father (probability $1/N_{em}$). In each case, 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

$$P = \frac{1}{N_e} = \frac{1}{4N_{em}} + \frac{1}{4N_{ef}} \quad (1.18)$$

Assuming random mating (no prohibition of mating between sibs), the effective number of each sex can be derived by the same method used to acquire Equation 1.13,

$$N_{es} = \frac{2N_{s,t-1} - 1}{(\sigma_{sk}^2/\mu_{sk}) + \mu_{sk} - 1} \quad (1.19)$$

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 . 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), then the mean and variance of gamete production for the whole population are respectively $\mu_k = 2(1 - \phi)\mu_{mk} = 2\phi\mu_{fk}$ and $\sigma_k^2 = (1 - \phi)\sigma_{mk}^2 + \phi\sigma_{fk}^2 + \phi(1 - \phi)(\mu_{mk} - \mu_{fk})^2$, and Equation 1.18 yields Equation 1.16. Thus, 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 a recent 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, which is much greater than the ratio of 1 expected under random mating. Assuming a stable population size ($\mu_{fk} = 2$) and substituting into Equation 1.19, the effective female population size for these species is found to be 40-90% of the actual number of females.

Further simplification of Equation 1.19 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 1.18,

$$N_e = \frac{4N_{m,t-1}N_{f,t-1}}{N_{m,t-1} + N_{f,t-1}} = 4\phi(1 - \phi)N_{t-1} \quad (1.20)$$

N_e attains a maximum of N_{t-1} when the sex ratio is balanced ($\phi = 0.5$). With skewed sex ratios, the effective size of the population is influenced much more strongly by the density of the rarer sex than of the more abundant sex. For example, in a highly polygynous species as $\phi \rightarrow 1$, $N_e \rightarrow 4(1 - \phi)N_{t-1} \simeq 4N_{m,t-1}$.

Age structure. The previous formulae have been obtained under the assumption of discrete generations. Such expressions are reasonable for organisms such as annual plants (ignoring the problem of seed banks) or univoltine insects, but for species that reproduce at different ages, as in the case of most vertebrates, generations overlap in time. It turns out that this does not complicate things too much. The work of Hill (1972a, 1979) indicates a simple correspondence between the effective sizes of populations with and without age-structure.

In the previous formulations N was the number of reproductive individuals in the population in each generation. For age-structured populations, it is convenient to consider N_b , the number of births during each unit of time multiplied by the number of time units/generation. The latter quantity, known as the *generation time* (T), is the average age of parents giving birth. For an ideal monoecious population,

$$T = \frac{\sum_{i=1}^{\tau} il_i b_i}{\sum_{i=1}^{\tau} l_i b_i} \quad (1.21)$$

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. 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,

$$H_t = H_0 \prod_{i=0}^{t-1} \left(1 - \frac{1}{2N_{ei}}\right) \quad (1.24)$$

It is sometimes of interest to evaluate the size of an ideal population that would have the same expected heterozygosity after t generations of inbreeding as a population with variable size over the same period. An approximate answer can be obtained by noting that with moderately large N_{ei} , Equation 1.24 simplifies to

$$H_t \simeq H_0 \exp\left(-\sum_{i=0}^{t-1} \frac{1}{2N_{ei}}\right) \quad (1.25)$$

which may be compared to

$$H_t \simeq H_0 \exp(-t/2N_e)$$

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

$$N_e^* \simeq \frac{t}{\frac{1}{N_{e0}} + \frac{1}{N_{e1}} + \cdots + \frac{1}{N_{e,t-1}}} \quad (1.26)$$

The long-term effective size N_e^* is approximately equal to the harmonic mean of the generation-specific effective sizes. A star 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 projection of $\exp(-t/2N_e^*)$ depending upon the temporal changes in N_{ei} .

Example 2. 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 1.26, $N_e^* = 2/(0.1 + 0.01) = 18.2$. 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.

Some work has been done on the effective size of populations consisting of patches subject to periodic extinctions and recolonizations (Maruyama and Kimura 1980, Ewens et al. 1987, Ewens 1989a). Although it is highly technical, this work is of great relevance to natural populations. There is room for much further study here, since the current models make some rather unlikely assumptions about patterns of colonization.

Partial inbreeding. In all of the previous formulations, the assumption has been made that the union of gametes is random. More realistically, we might expect the frequency of mating between relatives to exceed that expected under random mating. Many plants, for example,

produce a high proportion of their offspring by self-fertilization. If the total population size is infinite, a fixed proportion of matings between relatives leads to an equilibrium at which 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). Such an equilibrium does not exist for finite populations since the gene frequency is subject to random genetic drift.

Wright (1951) gave limited attention to this matter, and there has been little work on the subject since then. Pollak (1987) has recently developed a general theory for partial selfing, showing that

$$N_e = \frac{4N_t \left(1 - \frac{\beta}{2}\right) - 2}{\left(1 + \frac{\beta}{2-\beta}\right) [\sigma_k^2 + 2(1 - \beta)]} \quad (1.27)$$

where β is the frequency of selfing. If the population size is moderately large and stable, gamete production is uniform among individuals, and outcrossed matings random, this simplifies to

$$N_e \simeq \frac{(2 - \beta)N_t}{2} \quad (1.28)$$

Partial selfing introduces further complications into the theory of inbreeding. The standard equation for the erosion of average heterozygosity (1.5) is replaced by

$$1 - f(t) \simeq \left[\left(1 - \frac{\beta}{2 - \beta}\right) \left(1 - \frac{1}{2N_e}\right)^t + \frac{\beta}{2 - \beta} \left(\frac{\beta}{2}\right)^t \right] [1 - f(0)] \quad (1.29)$$

(Pollak 1987). This is another area in which there is room for much more work.

Additional Considerations

Earlier in the chapter, it was emphasized that the effects of finite population size are twofold: the development of homozygosity within populations and the divergence of gene frequencies between replicate populations. Crow (1954) has emphasized that the former is most closely related to the number of parents (or grandparents) since it is based upon the probability of uniting gametes coming from the same ancestor. On the other hand, random gene frequency drift is primarily a function of the number of offspring produced since it is based on sampling error in the gamete stage. Thus, in an expanding or declining population, the rates of inbreeding and gene frequency drift depend on different population sizes. In order to clarify this distinction, Crow (1954) defined an inbreeding effective size (N_e^I) and a variance effective size (N_e^V).

In all of the previous applications of the concept of effective population size, we were concerned with N_e^I . The variance effective size is defined to be the size of an ideal monocious population that yields the binomial sampling variance of the gene frequency, $p(1 - p)/2N_e^V$, which matches that observed in a set of non-ideal replicate populations. General formulae for the variance effective size are available in Kimura and Crow (1963b), Crow and Kimura (1970), Crow and Morton (1955), and Crow and Denniston (1988).

The purpose of the preceding sections was to illustrate how the effective population size can be estimated if one can quantify the total size and various attributes of the mating system of the population. Under many circumstances, however, it is not possible to obtain all of the necessary demographic data. An alternative approach is the indirect method of monitoring temporal changes in gene frequencies, usually with allozyme markers, and

inferring the effective population size that is necessary to account for such change under the assumption that the markers are selectively neutral (Pollak 1983, Tajima and Nei 1984, Waples 1989, Weir 1990).

It should now be amply clear that the loss of heterozygosity from populations can result from two causes. The first of these, inbreeding due to consanguineous mating, is in a sense transient because it can be eliminated completely with one generation of random mating. The loss of heterozygosity resulting from finite population size is more permanent unless several isolated populations are available for crossing. Up to now, we have treated these two components of random genetic drift separately, but both can, and usually do, occur simultaneously in natural populations. Moreover, the effects of finite population size can often operate on several levels. Species are generally subdivided into several partially or completely isolated populations, which in turn may be fragmented into local demes, which may be further structured into family groups. Clearly, if one wants to characterize the mechanisms of random genetic drift in natural populations, a means for partitioning the causal sources is required.

Wright (1951, 1965, 1973) developed an ingenious framework for attacking the problem. The strategy is to take a hierarchical view of population structure. Consider the simple situation in which the total population is divided into several isolated subpopulations within which there may be nonrandom mating. Letting f_{IT} be the average inbreeding coefficient for members of the entire population, then $(1 - f_{IT})$ is the probability that a randomly chosen individual is not inbred at a particular locus. This must be equal to the product of the probability of not being inbred because of residence in a subpopulation of finite size $(1 - f_{ST})$ and the probability of not being inbred because of consanguineous mating within the subpopulation $(1 - f_{IS})$. Wright's f -statistics can be interpreted directly in terms of the observed and expected levels of heterozygosity at the total population and subpopulation levels. Consequently, several estimation procedures, which rely on observed measures of allozyme or nucleotide variation, are available (Nei 1987, Weir 1990).