Review article

Developments in the prediction of effective population size

ARMANDO CABALLERO

Institute of Cell, Animal and Population Biology, University of Edinburgh, West Mains Road, Edinburgh EH9 3JT, UK

Effective population size is a key parameter in evolutionary and quantitative genetics because it measures the rate of genetic drift and inbreeding. Predictive equations of effective size under a range of circumstances and some of their implications are reviewed in this paper. Derivations are made for the simplest cases, and the inter-relations between different formulae and methods are discussed.

Keywords: heterozygosity, inbreeding, nonrandom mating, population numbers, random genetic drift.

Contents

1 The idealized population and effective population size

In an infinitely large population and in the absence of mutation, migration and selection, gene and genotype frequencies remain constant over generations. In finite populations, however, gene frequencies fluctuate randomly from generation to generation as a result of the finite sampling of gametes. These erratic changes constitute the so-called dispersive process or genetic drift, which is likely to be one of the main factors governing molecular evolution (Kimura, 1983) and has implications on the rate and time to fixation of selectively advantageous or deleterious genes (see Crow & Kimura, 1970, chap. 8).

Genetic drift, at least in unstructured populations, is quantified by a single parameter, *effective population size*, which can be computed or estimated from laboratory or field data, and predicted under a range of circumstances. This paper reviews predictive equations of the effective size and their inter-relations. For completeness and for a better understanding of the concepts, some of the basic theory is summarized first.

The simplest possible conditions under which the dispersive process can be studied are met in the *Wright–Fisher idealized population* (Fisher, 1930; Wright, 1931). This consists of an infinite, randomly mated base population subdivided into infinitely many subpopulations, each with a constant number, N, of breeding individuals per generation. In each subpopulation, parents produce an infinite number of male and female gametes into a large pool from which only $2N$ are sampled and united to produce the N zygotes of the following generation. All individuals survive from birth to adulthood. Both the sampling of gametes and their union (including self-fertilization) are random, so that all parents have an equal chance of producing offspring and the distribution of offspring number is multinomial. Systematic changes in allele frequencies are excluded in this idealized population, generations do not overlap, and only autosomal loci are considered.

The dispersive process can be studied as a sampling process or as an inbreeding process because both an increase in the variance of gene frequency among subpopulations and an increase in homozygosity occur as a result of the finite population size.

Under the simple conditions of the idealized population, sampling of gametes is binomial and the variance of the change in gene frequency is:

$$\sigma_{\Delta q}^2 = \frac{q(1-q)}{2N}, \tag{1}$$

where q is the allele frequency of a gene in the infinite base population. The coefficient of inbreeding at generation t, the probability that two gametes which unite to produce a zygote in generation t carry identical by descent copies of a gene (Wright, 1922; Malécot, 1948), is

$$F_t = \frac{1}{2N} + \left(1 - \frac{1}{2N}\right) F_{t-1},$$

where the first term denotes identity by descent from copies of a gene of an individual in generation $t-1$ and the second, that from copies of a gene of an individual in previous generations. The rate of increase in inbreeding per generation is thus:

$$\Delta F = \frac{1}{2N}, \tag{2}$$

where

$$\Delta F = \frac{F_t - F_{t-1}}{1 - F_{t-1}}.$$

The observable consequence of this increase in inbreeding is a reduction in the expected heterozygosity (H) each generation,

$$\lambda = \frac{H_t}{H_{t-1}} = 1 - \Delta F, \tag{3}$$

or, relative to that in the base population,

$$\frac{H_t}{H_0} = 1 - F_t = (1 - \Delta F)^t. \tag{4}$$

$\lambda = 1 - 1/2N$ is the largest nonunit eigenvalue of the transition matrix

$$\mathbf{P} = \{p_{ij}\}, \; p_{ij} = \binom{2N}{j}\left(\frac{i}{2N}\right)^j\left(1 - \frac{i}{2N}\right)^{2N-j},$$

which, for the various possible states (number of copies, i, of a given allele) in generation $t-1$, gives the probability of each state (j) in generation t.

The relationship between the variance of gene frequency over subpopulations and the coefficient of inbreeding is:

$$\sigma_{q,t-1}^2 = q(1-q)\left[1 - \left(1 - \frac{1}{2N}\right)^t\right] = q(1-q)F_t,$$

where the one generation delay between σ_q^2 and F is due to the fact that drift begins one generation earlier than inbreeding. Thus, in N individuals randomly chosen from an infinite population there is yet no inbreeding but there is drift.

It is obvious that real populations are very unlikely to meet the conditions of the idealized population defined above and, therefore, the number of breeding individuals does not describe appropriately the effects of inbreeding and gene frequency drift in most practical situations. The concept of effective population size (N_e) was introduced by Sewall Wright (1931, 1938, 1939) to overcome this problem and has been developed subsequently by others, mainly James F. Crow and coworkers (Crow & Kimura, 1970, pp. 345–364; Crow & Denniston, 1988). The effective size of a population is defined as *the size of an idealized population which would give rise to the variance of change in gene frequency or the rate of inbreeding observed in the actual population under consideration,* i.e. $N_e = q(1-q)/2\sigma_{\Delta q}^2$ or $N_e = 1/2\Delta F$ (from eqns (1) and (2)), which correspond to the so-called variance

and inbreeding effective sizes, respectively. Thus, the effective size gives a measure of the rate of genetic drift and inbreeding in the population. As neutral genetic variation depends directly on these parameters, effective size gives a prediction of the impact of management practices on genetic variation. Also, as the effective size becomes smaller, weakly selected alleles become effectively neutral. It is, therefore, a primary variable to biologists concerned with monitoring genetic variation in natural populations and, particularly, those concerned with the management of endangered or zoo species (Lande & Barrowclough, 1987; Nunney & Campbell, 1993). Effective size is also important in plant and animal breeding because its magnitude affects the response to artificial selection and its variance (see Hill (1985a, 1986) for reviews).

If the variance of change in gene frequency or the rate of increase in inbreeding are known, because the genotypes can be distinguished and hence the genotypic frequencies estimated or because pedigrees are available, the effective population size can be estimated or computed directly from the expressions above. For example, if we can trace an observable quantity such as the heterozygosity so that we know its rate of decay (H_t/H_{t-1}), we can use eqns (2) and (3) to estimate the asymptotic N_e. This is what is called effective size 'for random extinction' (Crow (1954) after a result from Haldane (1939)) or 'eigenvalue' effective size, because it is the result arising when the largest nonunit eigenvalue of the transition matrix of a Markov Chain is obtained (see Ewens, 1979, 1982). The actual N_e can also be computed from the inbreeding coefficients obtained from pedigrees. Estimations can be made even when there are individuals with uncertain parentage in the pedigree, by including information external to the records (Pérez-Enciso *et al.*, 1992).

When information on genotypic frequencies or pedigrees is not available, effective size can still be predicted under certain circumstances (in which one or more assumptions of the idealized population are removed) when demographic data such as census numbers and variances and covariances of the number of progeny per parent are available. Effective size can be derived following a sampling drift approach or an inbreeding approach when the matter of interest is in the gene frequency drift or the increase in homozygosity, respectively (Crow, 1954). Estimates from these two approaches are the same under random mating and constant population size over generations but can be different under other circumstances, and also different from the eigenvalue effective size (Ewens, 1982). In what follows some of these predictions are revised. In general, diploid populations are considered, but haploid and polyploid populations will also be mentioned. The practical problems in the estimation of

the demographical parameters necessary to estimate N_e in natural populations are not dealt with in this paper. Some of these problems and a comparison between different methods of estimation are given by Begon *et al.* (1980) (see also Nunney & Elam (1994)). A recent account of the practical and theoretical considerations of the estimation of N_e from temporal changes in gene frequency of allozyme markers is given by Waples (1989).

2 Self-fertilization not allowed

The simplest deviation from the idealized population which can be considered is the exclusion of self-fertilization in monoecious species. Under this situation, the probability that two gametes which unite to produce a zygote in generation t carry identical copies of a gene of an individual in generation $t-2$ is $(1 + F_{t-2})/2N$, and the probability that they carry identical copies of a gene from different individuals is $(1 - 1/N)F_{t-1}$. Thus, the inbreeding in generation t is

$$F_t = \frac{1 + F_{t-2}}{2N} + \left(1 - \frac{1}{N}\right) F_{t-1} . \tag{5}$$

In this case, genetic drift begins two generations earlier than inbreeding. By using eqns (3) and (4), expression (5) leads to $N\lambda^2 - \lambda(N-1) - 1/2 = 0$, with solution $\lambda = (N - 1 + \sqrt{N^2 + 1})/2N \approx 1 - 1/(2N + 1)$ (Wright, 1969, p. 195). Thus, from eqn (3):

$$\Delta F \approx \frac{1}{2N + 1} , \tag{6}$$

and, by the definition of N_e and using eqn (2), $N_e \approx N + 0.5$, which is a very small difference if N is large. Note that this value is the eigenvalue effective size because it is computed from the largest nonunit eigenvalue (λ) of the transition matrix corresponding to this model (see Ewens, 1979, p. 107). Henceforth, for populations without selfing, effective size will be computed with reference to an idealized population in which selfing is not allowed, so that in the idealized conditions, $N_e = N$. A discussion on this point is made later.

Avoidance of other types of mating between relatives also has a very small effect when N is large (see, for example, Wright, 1969, p. 198).

3 Different numbers of male and female parents

Assume that the number of male parents (N_m) is different from that of female parents (N_f), and that these are constant over generations. As half the genes at any

generation (say $t-1$) come from male parents and half from female parents, the probability that two genes in $t-1$, which unite to produce an offspring in t, both come from males in generation $t-2$ is $1/4$, and the probability that they come from the same male is $1/4N_m$, and analogously for females. Therefore, the probability that two genes which unite at random to produce a zygote in generation t come from the same individual in generation $t-2$ (regardless of sex) is $1/4N_m + 1/4N_f$. In the idealized population, this probability is $1/N$, so by equating $1/N_e$ to the former,

$$\frac{1}{N_e} = \frac{1}{4N_m} + \frac{1}{4N_f} \tag{7}$$

(Wright, 1931).

Defining $N = N_m + N_f$, $m = N_m/N$ and $f = N_f/N$, then $N_e = 4mfN$, which shows that N_e is maximal (and equal to N) when $m = f = 1/2$, otherwise $N_e < N$. It also shows that the size of the less numerous sex has the larger impact on N_e. For example, if $m = 0.01$, i.e. $N_m = 0.01N$ and $N_f = 0.99N$, then $N_e \approx 0.04N = 4N_m$.

4 Variable population size over generations

In the idealized population the number of breeding individuals (N) is constant over generations. Let us consider a situation where the population size varies over generations, with size N_i in generation i. From eqns (3) and (4), the expected heterozygosity in generation t relative to that in the base population is $H_t/H_0 = \Pi_{i=1}^{t}(1 - 1/2N_i)$. This can be equated to the relative heterozygosity in the idealized population (eqns (2) and (4), replacing N by N_e), i.e. $(1 - 1/2N_e)^t$. When population sizes are large and t much smaller than any of these, this latter equality can be approximated by $1 - \Sigma_{i=1}^{t}1/2N_i \approx 1 - t/2N_e$, from where:

$$\frac{1}{N_e} \approx \frac{1}{t}\sum_{i=1}^{t}\frac{1}{N_i} \tag{8}$$

(Wright, 1938; Crow & Kimura, 1970, pp. 109–110), i.e. using the harmonic rather than the geometric mean. The extension to separate sexes is $1/N_e = (1/t)\Sigma_{i=1}^{t}(1/4N_{m,i} + 1/4N_{f,i})$ (Chia & Pollak, 1974). Because N_e is a harmonic mean, two important points similar to those explained for eqn (7) appear. Firstly, the maximal N_e given a total $\Sigma_{i=1}^{t}N_i$, is achieved with constant population size over generations. Secondly, N_e is most strongly affected by periods of reduced population size. In other words, if a bottleneck occurs, causing an increase in inbreeding, this is not restored by a later expansion of the population size. Nagylaki (1992, pp. 239–242) discusses the effects of population growth on genetic heterogeneity and Motro & Thompson (1982)

have investigated the robustness of eqn (8) in situations which take into account the existence of mutations.

As was mentioned before, genetic drift begins one generation (if selfing is allowed) or two (if not) earlier than inbreeding. If population size varies over time, for a given generation, genetic drift depends on the number of individuals in that generation whereas inbreeding depends on the number of their parents (if selfing is allowed) or their grandparents (if not).

5 Nonrandom contribution from parents

In the idealized population all parents have an equal chance of leaving offspring which become parents in the next generation and differences in the contributions of parents are exclusively due to sampling. Under these circumstances the number of successful gametes (those which will produce zygotes) contributed by parents is binomially distributed with mean $\mu_k = 2$ and variance:

$$\sigma_k^2 = 2(1 - 1/N). \tag{9}$$

In real populations, however, parents may have different probabilities of contributing offspring because of differences in their fertility or in the viability of their offspring or, perhaps, because of impositions by the breeder. Thus, parents will vary in their contributions more than expected just by chance. As a result, σ_k^2 will be generally higher than in the idealized population (eqn (9)) and the effective population size will be smaller than the number of breeding individuals, as will be shown below.

With non-age-structured populations, the most appropriate stage to which these means and variances of numbers of offspring may refer is at sexual maturity rather than at some earlier stage, so that the variances are least (Fisher, 1939). Crow & Morton (1955) have developed formulae for converting, under different models of survival, measurement made at earlier stages.

Differences among parents in their contributions other than by sampling might be due to noninherited or inherited causes, and the impact of the two situations on the effective size is different. Noninherited differences of contribution of parents can arise, for example, from environmental causes, randomly allocated among parents each generation, or to differences in fertility of the parents that are uncorrelated with the fertility of their offspring. When differences in contributions are due to inherited causes (by genetic or cultural transmission), however, this means that the offspring and later descendants of a parent with a high contribution partially inherit this property. We will first revise predictions of effective size when the nonrandom contribution of parents is due to noninherited causes.

5.1 Variation due to noninherited causes

5.1.1 Monoecious diploid species with selfing allowed

As stated above, effective size can be derived following a sampling drift approach or an inbreeding approach. The former is considered in what follows, with a derivation from Caballero & Hill (1992a) similar to those of Latter (1959) and Hill (1979). Alternative derivations are given by Pollak (1977), Crow & Denniston (1988) and Nagylaki (1992, pp. 261–267). Let us assume that in generation $t-1$ there are N_{t-1} individuals and individual i produces s_i selfed offspring and n_i nonselfed offspring, giving a total number of successful gametes contributed of $k_i = 2s_i + n_i$. Let x_i be the frequency of a given allele at an arbitrary neutral locus in individual i (thus, x_i is 0, 1/2 or 1 if it carries zero, one or two copies of the allele, respectively). The allele frequency in the population in generation $t-1$ is:

$$q_{t-1} = \frac{1}{N_{t-1}} \sum_{i=1}^{N_{t-1}} x_i = \frac{1}{2N_t} \mu_k \sum_{i=1}^{N_{t-1}} x_i,$$

by noting that $2N_t = N_{t-1}\mu_k$, where μ_k is the mean number of gametes contributed per individual. Analogously, the gene frequency in generation t is:

$$q_t = \frac{1}{2N_t} \sum_{i=1}^{N_{t-1}} \left[\sum_{j=1}^{s_i} (2x_i + \delta_{ij1} + \delta_{ij2}) + \sum_{j=1}^{n_i} (x_i + \delta_{ij}) \right],$$

where δ_{ij} is the difference in gene frequency between the jth sampled gene and its parental value x_i, i.e. δ_{ij} is zero if the parent is a homozygote or $\pm 1/2$ if a heterozygote (subscripts 1 and 2 denote the two genes passed to each selfed offspring). If s_i or n_i are zero, the corresponding terms drop out. The variance of change in gene frequency is then:

$$V_{\Delta q} = V[q_t - q_{t-1}] = \frac{1}{4N_t^2} \left\{ V\left[\sum_{i=1}^{N_{t-1}} (2s_i x_i + n_i x_i - \mu_k x_i) \right] \right.$$
$$\left. + V\left[\sum_{i=1}^{N_{t-1}} \left(\sum_{j=1}^{s_i} (\delta_{ij1} + \delta_{ij2}) + \sum_{j=1}^{n_i} \delta_{ij} \right) \right] \right\}.$$

Gene frequencies and numbers of gametes per parent are assumed to be uncorrelated, and Mendelian sampling terms are also uncorrelated and have equal variance ($V[\delta_{ij}]$). We take a retrospective approach (Crow & Kimura, 1970, p. 353) in defining the effective size from the observed distribution of the number of gametes contributed per parent to the next generation, such that the value of $k_i = 2s_i + n_i$ for parent i is fixed, describing the actual number of gametes contributed by parent i, and not a random variable. We defined however, the variance of family size as $V(k_i) = [\sum_{i=1}^{N_{t-1}} (k_i - \mu_k)^2]/N_{t-1}$. Finally, we ignore sec-

ond order terms introduced by correlations among the x_i, because their sum is fixed. Therefore,

$$V_{\Delta q} = \frac{1}{4N_t^2} [N_{t-1} V(x_i) V(k_i) + N_{t-1}\mu_k V(\delta_{ij})]$$

$$= \frac{1}{2N_t\mu_k} [V(x_i) V(k_i) + \mu_k V(\delta_{ij})].$$

Now, we take expectations for the different quantities over an infinite number of conceptual replicates of the population. Thus, $V(x_i) = q(1-q)(1 + \alpha_{I,t-1})/2$ where $\alpha_{I,t-1}$ is the deviation from Hardy–Weinberg proportions in generation $t-1$, approximately equivalent to Wright's F_{IS} statistic (Wright, 1969, pp. 294–295). The expectation of $V(\delta_{ij})$ equals the expected frequency of heterozygotes $2q(1-q)(1 - \alpha_{I,t-1})$ times the variance generated from them $(1/4)$. Finally we substitute $V(k_i)$ by the variance of family size in an infinite population (i.e. with the Gaussian correction), $S_k^2 = V(k_i)N_{t-1}/(N_{t-1} - 1)$. Hence,

$$V_{\Delta q} = \frac{1}{2N_t\mu_k} \left[S_k^2 \frac{q(1-q)}{2} (1 + \alpha_{I,t-1}) + \mu_k \frac{q(1-q)}{2} (1 - \alpha_{I,t-1}) \right].$$

Equating this to the variance of drift per generation in the idealized population (eqn (1)) and rearranging we obtain:

$$N_e = \frac{2N_t}{(1 - \alpha_{I,t-1}) + \frac{S_k^2}{\mu_k}(1 + \alpha_{I,t-1})} \tag{10}$$

(Crow & Morton, 1955; Kimura & Crow, 1963a; Crow & Denniston, 1988). When N is constant over generations, $N_t = N$, $\mu_k = 2$ and eqn (10) reduces to:

$$N_e = \frac{4N}{2(1 - \alpha_I) + S_k^2(1 + \alpha_I)}. \tag{11}$$

Even if mating of gametes including random selfing is at random, α_I is not exactly zero due to the discreteness of the possible number of genotypes in a finite population, and the appropriate value is obtained by Kimura & Crow (1963a) (see also Robertson, 1965). Consider the $2N$ gametes from which the N individuals in the next generation are derived. If p and q are the frequencies of the two alleles under consideration in the sample, the proportion of heterozygotes is obtained by sampling without replacement:

$$\frac{(2Np)(2Nq)}{\binom{2N}{2}} = 2pq \left(\frac{2N}{2N-1} \right),$$

and equating to $2pq(1 - \alpha_I)$

$$\alpha_I = \frac{-1}{2N - 1}.$$ (12)

In the idealized population, σ_k^2 is given by eqn (9), $S_k^2 = \sigma_k^2 N/(N-1) = 2$ and from eqn (11), $N_e = N$, as expected from Wright's concept of effective population size. In this situation we note that about half the variance in gene frequency drift is due to variable progeny numbers, and half is due to segregation from heterozygotes.

Let us now consider the general situation where a given proportion of offspring (β) is produced each generation by selfing and the remainder $(1 - \beta)$ by random mating, but the self-fertilization habit is not inherited. Then, it can be shown that α_I quickly asymptotes to $\beta/(2 - \beta)$ (Haldane, 1924; Li, 1976, pp. 243–244). This is the equilibrium inbreeding of an infinite population when the increase in inbreeding due to selfing is counterbalanced by its breakdown due to outcrossing. In the particular case where the numbers of selfed and nonselfed progeny are independently Poisson distributed, $S_k^2 = 4\beta + 2(1 - \beta) = 2 + 2\beta$ and eqn (11) reduces to:

$$N_e = \frac{N}{1 + \alpha_I}$$ (13)

(Li, 1976, p. 562; Pollak, 1987; Caballero & Hill, 1992a).

The above derivations were made following a sampling drift approach. Effective population size can also be derived from an inbreeding approach, and inbreeding and variance effective sizes can be different under certain circumstances. Let us assume again that the population consists of N_{t-1} monoecious individuals in generation $t - 1$ and that mating is at random, for simplicity. Individual i will contribute k_i successful gametes to the next generation, so the average number of successful gametes per individual is $\mu_k = \Sigma k_i / N_{t-1}$, where the summation in this and the following equations is for $i = 1$ to N_{t-1}. Thus, the probability that two uniting gametes in generation t come from the same individual in generation $t - 1$ is:

$$\frac{\Sigma k_i(k_i - 1)/2}{N_{t-1}\mu_k(N_{t-1}\mu_k - 1)/2} = \frac{\Sigma k_i^2 - \Sigma k_i}{N_{t-1}\mu_k(N_{t-1}\mu_k - 1)}.$$

Now, $\sigma_k^2 = (\Sigma k_i^2/N_{t-1}) - \mu_k^2$ and, therefore, $\Sigma k_i^2 = N_{t-1}(\sigma_k^2 + \mu_k^2)$, and $\Sigma k_i = N_{t-1}\mu_k$. Substituting these into the above expression and rearranging, the probability that two gametes in generation t come from the

same individual in $t - 1$ is:

$$\frac{\sigma_k^2 + \mu_k^2 - \mu_k}{\mu_k(N_{t-1}\mu_k - 1)}.$$

In the idealized population the corresponding probability is $1/N$. Hence, replacing N by N_e and equating both we obtain:

$$N_e = \frac{N_{t-1}\mu_k - 1}{\mu_k - 1 + \sigma_k^2/\mu_k}$$ (14)

(Crow, 1954; Kimura & Crow, 1963a; Pollak, 1977; Crow & Denniston, 1988). Again, substituting μ_k by 2 and σ_k^2 by its value in the idealized population (eqn (9)), we obtain $N_e = N$, as expected.

A comparison between eqns (10) and (14) shows the conceptual difference between inbreeding and variance effective numbers. Both equations were derived to account for the change in gene frequency or the inbreeding in generation t. Thus, we note that the variance effective number (eqn (10)) depends on the number of progeny (N_t), while the inbreeding effective number (eqn (14)) depends on the number of parents (N_{t-1}). This is in accordance with the one generation gap between inbreeding and gene frequency variance stated before for populations with selfing allowed. Thus, the inbreeding effective size is smaller than the variance effective size when the population size is growing and vice versa. Both numbers measure different properties and the choice of which to measure depends on whether interest is in the gene identity or in the amount of gene frequency drift.

Templeton (1980) has suggested that the most favourable condition for rapid speciation caused by a founder event occurs when there is a rapid accumulation of inbreeding without a severe reduction in genetic variability or, in other words, when the inbreeding effective size of the founder population is low but the variance effective size is high. However, as discussed by Barton & Charlesworth (1984), Templeton only considers a single generation involved in the process. If a sequence of generations with restricted numbers of individuals is considered, the ultimate reduction in heterozygosity and variance of gene frequency both depend on the sequence of numbers, as illustrated in eqn (8), and inbreeding and variance effective size are the same over the whole period.

When N is constant over generations, eqn (14) reduces to:

$$N_e = \frac{4N - 2}{2 + \sigma_k^2}$$ (15)

(Wright, 1938, 1939). Haldane (1939) also arrived at an approximate form of eqn (15), $N_e = 4N/(2 + \sigma_k^2)$, when deriving the asymptotic rate of decay of segregating loci in a population. Eqn (15) is the same as the variance effective size (eqn (11)) when S_k^2 is replaced by σ_k^2 with the Gaussian correction, i.e. $S_k^2 = \sigma_k^2 N/(N-1)$, and α_I from eqn(12). Thus, inbreeding and variance effective numbers are the same when population size is constant over generations, and we will keep this assumption henceforth for simplicity.

5.1.2 Separate sexes With separate sexes a similar derivation to that above in terms of sampling drift can be followed regarding couples instead of individuals. Alternative derivations are given by Moran & Watterson (1959) and Pollak (1977). Let us assume that the population consists of a constant number $N/2$ of families (couples) and family i contributes k_i offspring to the next generation. The mean frequency of a given allele in family i is $(x_{im} + x_{if})/2$, where m and f denote male and female parents, respectively. Thus, the allele frequency in generation $t-1$ is:

$$q_{t-1} = \frac{1}{N} \sum_{i=1}^{N/2} (x_{im} + x_{if}),$$

and in the following generation:

$$q_t = \frac{1}{N} \sum_{i=1}^{N/2} \left[k_i \left(\frac{x_{im} + x_{if}}{2} \right) + \frac{1}{2} \sum_{j=1}^{k_i} (\delta_{ijm} + \delta_{ijf}) \right].$$

Proceeding as before, the variance of genetic drift is:

$$V_{\Delta q} = V[q_t - q_{t-1}] = \frac{1}{N^2} \left\{ \left[\frac{N}{8} V(x_{im} + x_{if}) V(k_i) \right] \right.$$
$$\left. + \frac{N}{4} [V(\delta_{ijm}) + V(\delta_{ijf})] \right\}.$$

Now, the expectations of the different variances are $V(x_{im}) = V(x_{if}) = q(1-q)(1+\alpha_I)/2$, where α_I is the deviation from Hardy–Weinberg proportions in male or female parents and q is the average allele frequency in males and females, $cov(x_{im}, x_{if}) = q(1-q)\alpha_o$, where α_O is the correlation between genes in pairs of parents (or the correlation between genes within individuals in their offspring), $V(\delta_{ijm}) = V(\delta_{ijf}) = q(1-q)(1-\alpha_I)/2$, and finally, we substitute $V(k_i)$ by the variance of family size with the Gaussian correction, $S_k^2 = V(k_i)N/(N-2)$. Thus, substituting above and equating to the variance of drift per generation in the idealized population and rearranging we obtain:

$$N_e = \frac{4N}{2(1 - \alpha_I) + S_k^2(1 + \alpha_I + 2\alpha_o)} \tag{16}$$

(Caballero and Hill, 1992a).

With different numbers of males and females, the general equation for the effective number of individuals of sex s is:

$$N_{es} = 4N_s \left/ \left\{ \left[\frac{1}{\mu_{sm}} + \frac{1}{\mu_{sf}} \right] (1 - \alpha_{I,s}) \right. \right.$$

$$\left. + \left[\left(\frac{S_{sm}^2}{\mu_{sm}^2} + \frac{2S_{sm,sf}}{\mu_{sm}\mu_{sf}} + \frac{S_{sf}^2}{\mu_{sf}^2} \right) \right] (1 + \alpha_{I,s} + 2\alpha_o) \right\} \tag{17}$$

(Crow & Denniston, 1988, except the term $2\alpha_o$), where N_s is the number of parents of sex s, $\mu_{sm} = N_m/N_s$ ($\mu_{sf} = N_f/N_s$) and S_{sm}^2 (S_{sf}^2) are the mean and variance of male (female) offspring from parents of sex s, respectively, $S_{sm,sf}$ is the covariance between the numbers of male and female offspring from parents of sex s, and $\alpha_{I,s}$ is the deviation from Hardy–Weinberg proportions in parents of sex s.

As the total number of progeny of each sex is assumed to be independently fixed, the changes in gene frequency in males and females are also independent, and we can combine the effects of the two sexes by expression (7), replacing N_s by N_{es},

$$\frac{1}{N_e} = \frac{1}{4N_{em}} + \frac{1}{4N_{ef}} \tag{18}$$

(Crow & Denniston, 1988). This is an approximation implying two assumptions: firstly, the changes in gene frequency in both sexes are weighted equally, as half of the genes come from each sex; secondly, it is assumed that gene frequencies are the same in both sexes, which is reasonable at the steady state. Note that when $N_m = N_f$, eqns (17) and (18) reduce to eqn (16).

When mating is at random $\alpha_I = -1/(2N_s - 1)$ (from eqn (12)) as the numbers of males and females are independently fixed, and $\alpha_O = 0$. Then eqn (16) is the same as for the monoecious case (eqn (11)).

When population size is large, α_I can be neglected and rearranging eqns (17) and (18) we arrive at:

$$\frac{1}{N_e} = \frac{1}{16N_m} \left[2 + S_{mm}^2 + 2\left(\frac{N_m}{N_f} \right) S_{mm,mf} + \left(\frac{N_m}{N_f} \right)^2 S_{mf}^2 \right]$$

$$+ \frac{1}{16N_f} \left[2 + S_{ff}^2 + 2\left(\frac{N_f}{N_m} \right) S_{fm,ff} + \left(\frac{N_f}{N_m} \right)^2 S_{fm}^2 \right], \tag{19}$$

as obtained by Hill (1972b, 1979).

Now consider possible deviations from random mating. With a mix of full-sib mating and random mating, α_I quickly asymptotes to $\beta/(4 - 3\beta)$ (Ghai, 1969; Li, 1976, p. 245), where β is now the average proportion of full-sib matings per generation. In this case there is a positive correlation between gene frequencies in male and female parents (α_O) and, neglecting second order terms, this is approximately equal to α_I, because if the mating system is continued for several generations, correlations of genes within individuals approximately equal correlations of genes between mated individuals in the steady state. In other words, coefficients of inbreeding of individual parents (α_I) are approximately equal to the coancestry between mates, i.e inbreeding of their offspring (α_O). Thus, eqn (16) gives:

$$N_e = \frac{4N}{2(1 - \alpha_I) + S_k^2(1 + 3\alpha_I)}. \tag{20}$$

When the numbers of male and female offspring are independently Poisson distributed, $S_k^2 = 2$ and eqn (20) reduces to the same equation as with partial selfing $(N_e = N/(1 + \alpha_I)$, eqn (13)). However, if the numbers of full-sib and non-full-sib matings are independently Poisson distributed, $S_k^2 = 2 + 2\beta$, and eqn (20) reduces to $N_e = N/(1 + 3\alpha_I)$ (Pollak, 1987, 1988).

Equations (16) and (17) can be applied analogously to other systems of partial inbreeding by using the appropriate α_I. The general expression for the equilibrium value of α_I is $\alpha_I = \beta_j/[2^j - \beta(2^j - 1)]$, where β_j is the proportion of a mating system of jth − 1 degree (Hedrick & Cockerham, 1986). For example, for partial selfing, $j = 1$, and for partial half-sib mating where mothers of the half-sibs are again half-sibs, $j = 3$ (second degree mating). Approximations when more than one type of partial inbreeding occurs can be obtained by $\alpha_I = (\Sigma_j \beta_j/2^j)/[1 - \Sigma_j \beta_j(1 - 0.5^j)]$ (Hedrick & Cockerham, 1986).

An analogous derivation for the effective size with separate sexes and random mating from an inbreeding point of view can be given. Consider the first generation of a certain mating structure and assume there are $N_{s,0}$ individuals of sex s in generation 0 (the initial generation). The probability that two gametes uniting to form a zygote in generation 2, and coming from an individual of sex s, came from the same individual of sex s in generation 0 is:

$$\frac{\Sigma k_{smi}k_{sfi}}{\Sigma k_{smi}\Sigma k_{sfi}} = \frac{\mu_{sm}\mu_{sf} + \sigma_{sm,sf}}{N_{s,0}\mu_{sm}\mu_{sf}},$$

where k_{smi} (k_{sfi}) is the number of male (female) progeny from the ith individual of sex s in generation 0, and $\sigma_{sm,sf}$ is the covariance of the numbers of male and

female offspring from parents of sex s. Equating this probability to that in the idealized population, the effective number of individuals of sex s in generation 0 is:

$$N_{es,0} = \frac{N_{s,0}\mu_{sm}\mu_{sf}}{\mu_{sm}\mu_{sf} + \sigma_{sm,sf}}$$

(Crow & Denniston, 1988), which can be combined for the two sexes by eqn (18). If $N_m = N_f = N/2$ and these are constant over generations,

$$N_{e,0} = \frac{N}{1 + \sigma_{m,f}}, \tag{21}$$

where $\sigma_{m,f}$ is the covariance of the numbers of male and female offspring from parents of any sex.

This equation is strictly valid for the effective size in the initial generation. When more generations of the mating structure are considered the equation holds only for some situations because it accounts only for the possibility of identity in generation t from copies of a gene in generation $t - 2$ passed through a male and a female in generation $t - 1$. If one more generation is considered, copies of that gene in generation $t - 2$, carried by two males (or two females) in generation $t - 1$, may combine in their grandoffspring in generation $t + 1$ (Caballero & Hill, 1992b). Thus we can assume that whether two individuals in generation $t - 1$ are both male, both female, or one of each, is independent of the probability of their having common ancestry in generation $t - 2$. Hence, we can use eqn (14) separately for each parental sex, except that N refers to generation $t - 2$,

$$N_{es} = \frac{N_{s,t-2}\mu_s - 1}{\mu_s - 1 + \sigma_s^2/\mu_s} \tag{22}$$

(Crow & Denniston, 1988), where μ_s and σ_s^2 are the mean and variance of the number of offspring produced by parents of sex s, respectively. Now, substituting expression (22) for males and females into eqn (18), we obtain:

$$N_e = \frac{N_{t-2}\mu_k - 2}{\mu_k - 1 + \sigma_k^2/\mu_k}, \tag{23}$$

where $N_{t-2} = N_{m,t-2} + N_{f,t-2}$, $\mu_k = 2m\mu_m = 2f\mu_f$ and $\sigma_k^2 = m\sigma_m^2 + f\sigma_f^2 + mf(\mu_m - \mu_f)^2$, and m (f) is the proportion of male (female) parents (Kimura & Crow, 1963a; Crow & Denniston, 1988). If N is constant over generations, $\mu_k = 2$ and eqn (23) reduces to:

$$N_e = \frac{4N - 4}{2 + \sigma_k^2}. \tag{24}$$

A different form of eqn (24),

$$N_e = \frac{4N(N-1)}{N_m(\sigma_m^2 + \mu_m^2 - \mu_m) + N_f(\sigma_f^2 + \mu_f^2 - \mu_f)},$$ (25)

was obtained by Pollak (1977) and can also be derived by substituting eqn (22) (assuming constant numbers over generations) into eqn (18). Malécot (1951) and Moran & Watterson (1959) also obtained eqn (25) neglecting second order terms (leaving $4N^2$ in the numerator). An alternative derivation of eqn (24), neglecting second order terms, is also given by Caballero & Hill (1992b). In that paper, numbers of male and female parents were assumed to be the same for simplicity. If different numbers of male and female parents are considered, the derivation leads straightforwardly to eqn (19).

Eqn (24) applied to a case with different numbers of males and females illustrates the fact that this is a particular case of increased variance of family size, when individuals of the less numerous sex in general produce more offspring than those of the more numerous sex. For equal number of males and females and random mating, eqn (24) gives the same result as eqn (16). For different numbers of the two sexes, however, eqns (24) and (25) give the same result as eqns (17) and (18) or eqn (19) with random mating and Poisson distribution of offspring numbers, in which case all expressions reduce to eqn (7), but not in general. This is because eqns (24) and (25) are based on the assumption that there is no distinction between sexes in the offspring (Crow & Denniston, 1988; Caballero, 1994). Thus, when variances and covariances for the four pathways (σ_{mm}^2, σ_{mf}^2, etc.) can be estimated, eqns (17) and (18) or eqn (19) should be used instead of eqns (24) and (25).

Note that, if N is constant, eqn (24) applies for any generation, except the initial one under certain situations, for which eqn (21) should be applied. The following examples illustrate this point. For a multinomial distribution of family size $\sigma_k^2 = 2(1 - 2/N)$ and $\sigma_{m,f} = 0$. Thus, both eqns (21) and (24) predict $N_e = N$. However, if, for example, each couple contributes exactly one male and one female to the next generation, $\sigma_k^2 = \sigma_{m,f} = 0$ and from eqn (21) $N_{e,0} = N$ in the initial generation, because there is a one generation delay in the appearance of the effects of inbreeding (Ugarte *et al.*, 1990), whereas from eqn (24) $N_e = 2N - 2$, in subsequent generations.

A derivation of the effective size with partial full-sib and random mating from an inbreeding point of view has been made by Caballero & Hill (1992a),

$$N_e = \frac{N(4 - 3\beta)}{S_k^2 + 2 - 2\beta}.$$

Because $\alpha_I = \beta/(4 - 3\beta)$, on rearrangement $\beta = 4\alpha_I/(1 + 3\alpha_I)$, and substituting into the above, we obtain eqn (20), as expected.

As was discussed at the beginning of this section, the variance of the number of offspring contributed per parent is, in general, larger than expected by random sampling (eqn (9)), and the effective size is smaller than the number of breeding individuals. Thus, Crow & Morton (1955) have summarized observations in natural and laboratory populations (including *Drosophila*, snails and humans) revealing a ratio $N_e/N \approx 0.70$ (although for *Drosophila* males, values are somewhat smaller, about 0.4). Nei & Tajima (1981) have suggested much lower values (< 0.1) as typical of small organisms and Nunney & Campbell (1993) suggest values below 0.25 for captive populations. Briscoe *et al.* (1992) have estimated and reviewed values of N_e/N for captive populations of *Drosophila* below 0.1 in most cases, and Frankham (1994) summarizes estimates for 20 species. Low ratios are prevalent in highly fecund species like insects and fish, with ratios frequently below 0.1. Species with lower fecundity, like birds and mammals, have larger ratios but generally below 0.5, and particularly low in polygynous species. Finally, Nunney & Elam (1994) summarize and recalculate 14 estimates of N_e/N obtained by ecological methods in a variety of species with overlapping generations. Values range from 0.56 to 1.27, with an average of 0.73 but, as the authors stress, N is defined in these estimates as the number of adults of reproductive age and, therefore, the ratio of N_e to census number is likely to be smaller.

It is also possible, however, that the variance of family size is smaller than the value in eqn (9) and $N_e > N$. As an extreme, if, as mentioned above, each parent contributes exactly two gametes to successful breeders in the next generation, $\sigma_k^2 = 0$ and, from eqn (24), $N_e = 2N - 2 \approx 2N$. This situation is sometimes called 'minimal inbreeding' because the rate of change in inbreeding is about half what it would be in an idealized population with the same number of parents, and this has been of great value in the design of unselected control populations of domestic animals. In this case, random genetic drift is entirely attributed to Mendelian segregation in heterozygotes.

When the number of males is smaller than that of females and each male is mated to N_f/N_m females, 'minimal inbreeding' can be achieved by choosing as parents one male and N_f/N_m females from each male's progeny and one female and N_m/N_f males from each female's progeny. In this situation, $S_{fm}^2 = (N_m/N_f)[1 - (N_m/N_f)]$ and $S_{mm}^2 = S_{mf}^2 = S_{ff}^2 = S_{mm,mf} = S_{fm,ff} = 0$. Substituting these into eqns (17) and

(18) (neglecting α_I and α_O, for simplicity), we obtain:

$$\frac{1}{N_e} \approx \frac{3}{16N_m} + \frac{1}{16N_f},$$

(see eqn (7)), as found by Gowe et al. (1959).

It is now worth stressing the difference between the eigenvalue effective size, defined at the beginning of this paper, and the variance and inbreeding effective sizes. The difference becomes apparent in the model of separate sexes discussed above. When N is constant over generations, the distribution of family size is binomial and $N_m = N_f = N/2$, the inbreeding and variance effective sizes from eqns (16) or (24) are N and not $N + 0.5$, as would be predicted approximately by the eigenvalue effective size. This is because all the expressions for the inbreeding and variance effective sizes for separate sexes in this section approximately predict $(1/2\Delta F) - (1/2)$ (see eqn (6)), as self-fertilization is considered to be absent in the derivations. In fact, eqns (16) and (24) with σ_k^2 as the variance of successful gametes contributed by parents apply equally to monoecious populations when self-fertilization is prohibited (Crow & Denniston, 1988; Caballero & Hill, 1992b; Nagylaki, 1992, p. 247). In all these cases the inbreeding coefficient is predicted by eqn (5), and the reduction in heterozygotes approximated by eqns (3) and (6) (in both cases replacing N by N_e). For example, for $N = 2$, under the idealized conditions, $N_e = 2$ from eqn (16) or eqn (24) and $\Delta F \approx 0.20$ from eqn (6) (more exactly, $\Delta F = 0.191$). Thus, for each model under consideration (in this section and later) we specify a particular idealized population. For example, for separate sexes the idealized population has selfing excluded or, for haploid populations (see later) the idealized population is haploid, so that in all these circumstances $N_e = N$ under the appropriate idealized conditions. Alternatively, we can define a unique idealized population (as above), with N_e always obtained from $1/2\Delta F$ (eqn (2)) and 0.5 added to the expressions of N_e above without selfing (e.g. Wright, 1969, eqn (8.19)), so that inbreeding and variance effective sizes give about the same result as the eigenvalue effective size.

When the population size is expanding or contracting we have shown that inbreeding and variance effective sizes can be different. In this case the eigenvalue effective size is closer to the variance effective size, because it is also defined 'by looking to the future rather than by looking to the past' (Ewens, 1982), as does the inbreeding effective size. For a more formal account and a deeper understanding of the concepts of effective size and idealized populations, consult Ewens (1979, pp. 104–112; 1982), Crow & Denniston (1988, discussion) and Gregorius (1991).

At first sight it is surprising that eqn (15), applying to populations with selfing permitted, has a larger numerator $(4N - 2)$ than eqn (24) $(4N - 4)$, which applies to populations where selfing is not allowed, suggesting that N_e is larger when selfing is allowed. This counterintuitive fact arises because σ_k^2 is not the same under both circumstances. For binomial sampling of gametes, $\sigma_k^2 = 2(1 - 1/N)$ with selfing allowed (eqn (9)) but $\sigma_k^2 = 2(1 - 2/N)$ without selfing and, in both cases, $N_e = N$.

5.1.3 Haploid and polyploid species For haploid populations, assume that there is a constant number of individuals, N, each generation, and that individual i contributes k_i offspring. The gene frequency in individual i is $x_i = 0$ or 1, and that of the population in two consecutive generations, $q_{t-1} = \Sigma_{i=1}^{N} x_i/N$ and $q_t = \Sigma_{i=1}^{N} x_i k_i/N$. Thus, proceeding as before, the variance of the change in gene frequency is $V_{\Delta q} = V(x_i)V(k_i)/N$. Now, the expected values of $V(x_i)$ and $V(k_i)$ in an infinite population are $q(1-q)$ and S_k^2, respectively. Thus, equating $V_{\Delta q}$ to the variance of change in gene frequency in a haploid idealized population, $q(1-q)/N_e$, we obtain:

$$N_e = \frac{N}{S_k^2}. \tag{26}$$

For a multinomial or Poisson distribution of offspring number $S_k^2 = 1$ and $N_e = N$. When $S_k^2 = 0$, $N_e = \infty$, because there is no genetic drift. Substituting S_k^2 by $\sigma_k^2 N/(N-1)$, eqn (26) gives:

$$N_e = \frac{N-1}{\sigma_k^2}, \tag{27}$$

where σ_k^2 is the actual variance of k_i in the population.

S. Wright and J. F. Crow (unpublished data) have derived an expression for the variance of change in gene frequency for an n-ploid population (Crow & Morton, 1955, eq. 11) which, when equated to eqn (1), gives:

$$N_e = \frac{4N}{4\left(\dfrac{n-1}{n}\right)(1 - \alpha_I) + \dfrac{4S_k^2}{\mu_k}\dfrac{1 + (n-1)\alpha_I}{n}} \tag{28}$$

For $n = 1$, $\mu_k = 1$ with constant census size over generations and eqn (28) reduces to eqn (26). For $n = 2$, $\mu_k = 2$ and eqn (28) gives eqn (16) under random mating.

5.1.4 X-linked genes Effective population size for X-linked loci can be derived with an argument analo-

gous to that leading to eqn (16) for autosomal loci. It is assumed that the male is the heterogametic sex. The results also apply, of course, to all loci in haplo–diploid species. Let us consider equal numbers of males and females and random mating, for simplicity. Mean gene frequencies of an X-linked gene in a population of constant size N in two consecutive generations are:

$$q_{t-1} = \frac{2}{N} \sum_{i=1}^{N/2} \left(\frac{x_{im}}{3} + \frac{2x_{if}}{3} \right),$$

and

$$q_t = \frac{2}{N} \sum_{i=1}^{N/2} \left[\frac{x_{im}k_{im(f)} + x_{if}k_{if}}{3} + \frac{1}{3} \sum_{j=1}^{k_{if}} \delta_{if(j)} \right],$$

where x_{im} ($= 0$ or 1) is the gene frequency in male parents, x_{if} ($= 0$, $1/2$ or 1) is the gene frequency in female parents, k_{if} is the number of offspring produced by female parents, $k_{im(f)}$ is the number of female offspring produced by male parents and $\delta_{if(j)}$ ($= +1/2$ or $-1/2$) is the deviation in gene frequency due to segregation of a heterozygous female parent.

Proceeding as before, and noting that under random mating x_{im} and x_{if} are uncorrelated, we arrive at:

$$V_{\Delta q} = V[q_t - q_{t-1}]$$

$$= \frac{1}{9N} \{ 2[V(x_{im})V(k_{im(f)}) + V(x_{if})V(k_{if})] + 4V(\delta_{if(j)}) \},$$

for which approximate expectations are $V(x_{im}) = q(1-q)$, $V(x_{if}) = q(1-q)/2$, $V(\delta_{if(j)}) = q(1-q)/2$, $V(k_{if}) = S_k^2$ (if monogamy is assumed) and $V(k_{im(f)}) = S_{mf}^2$. Substituting these into the above expression and equating to the variance of drift per generation for an autosomal gene in the idealized population we obtain:

$$N_e = \frac{9N}{4 + 4S_{mf}^2 + 2S_k^2} \tag{29}$$

(see eqn (16)). For a Poisson distribution of family size, $S_{mf}^2 = 1$, $S_k^2 = 2$ and $N_e = 3N/4$.

Pollak (1980, 1990) has derived a general expression for different numbers of male and female parents and overlapping generations. In the discrete-generation case the expression is:

$$\frac{1}{N_e} = \frac{1}{9N_m} \left[1 + 2\left(\frac{N_m}{N_f}\right)^2 S_{mf}^2 \right]$$

$$+ \frac{1}{9N_f} \left[1 + S_{ff}^2 + 2\left(\frac{N_f}{N_m}\right) S_{fm,ff} + \left(\frac{N_f}{N_m}\right)^2 S_{fm}^2 \right] \tag{30}$$

(see eqn (19)), which reduces to eqn (29) when $N_m = N_f = N/2$. An alternative derivation of eqn (30) following an inbreeding approach is given by Caballero (1994). For a binomially distributed family size and large N, when, for example $S_{mf}^2 \approx N_f/N_m$, eqn (30) reduces to:

$$N_e = \frac{9N_m N_f}{4N_m + 2N_f} \tag{31}$$

(Wright, 1933, 1939; see also Malécot, 1951; Kimura, 1963a; Nagylaki, 1981 and Avery, 1984 for alternative derivations). An easy deduction of eqn (31) is made by noting that the frequency of an X-linked gene is $q = q_m/3 + 2q_f/3$, where q_s is the frequency in sex s. Thus, $V(q) = V(q_m)/9 + 4V(q_f)/9$, where $V(q_m) = q_m(1-q_m)/N_m$ and $V(q_f) = q_f(1-q_f)/2N_f$. At the steady state $q_m = q_f = q$ and, substituting and rearrangeing, $V(q) = q(1-q)(2N_m + N_f)/(9N_m N_f)$, which equated to $q(1-q)/2N_e$ gives (31).

Moran & Watterson (1959), Ethier & Nagylaki (1980) and Nagylaki (1981), following different approaches, have arrived at an expression different from eqn (30) and with similar structure to eqn (25) for autosomal genes. This expression is correct for Poisson or multinomial distributions of family size but not in general because, as for eqn (25), it is based on the assumption that the sex of the offspring cannot be identified (Caballero, 1994).

It is interesting to note that with 'minimal inbreeding' and different numbers of the sexes, i.e. choosing as parents one male and N_f/N_m females from each male's progeny and one female and N_m/N_f males from each female's progeny, so that $S_{fm}^2 = (N_m/N_f)[1 - (N_m/N_f)]$ and $S_{ff}^2 = S_{ff}^2 = S_{fm,ff} = 0$, eqn (30) reduces to $9N_m/2$, i.e. it is independent of the number of females. Furthermore, for species where females are the heterogametic sex, the same mating procedure gives $N_e = 9N_m N_f/(3N_f - N_m)$, which indicates that, for minimal inbreeding to be attained in sex-linked genes, as few females as possible should be used.

We can compare eqn (31) with the corresponding expression (7) for autosomal loci, which yields $N_e = 4N_m N_f/(N_m + N_f)$. Equating eqn (31) to this it is found that N_e for an X-linked locus is smaller than that for an autosomal locus unless $N_f > 7N_m$. Effective size can be very small in social haplo–diploid insects for which there is often only one breeding female (the queen). In this case, eqn (31) reduces to $N_e = 9N_m/(4N_m + 2)$.

5.1.5 Systems of mating In all the above equations which apply for random mating it has been assumed that mating is made by random union of gametes. Different systems of mating have also an impact on

effective population size. Nunney (1993) has investigated some systems of mating which are likely to occur in natural populations. These include lottery polygyny, in which males attempt to mate with each of the females they encounter, but females mate only once or a few times, and other systems of mating with more extreme polygyny, for example those in which harems are formed. We will mention here some of his simplest results for lottery polygyny, which might be a likely system of mating in many insects. Further insight will be given later, in the section on overlapping generations.

Under lottery polygyny, we can assume, for simplicity, that all females mate at least once and that the average number of matings per female is n. Thus, defining $N = N_m + N_f$, $m = N_m/N$ and $f = N_f/N$, the total number of matings is nfN and the probability that a male is chosen to breed is the same for all males, $1/mN$. With this model, the effective size can be expressed as:

$$N_e = \frac{4mfN}{1 + (m/n)}.$$

Note that, if females mate many times (large n), the above expression tends to the result for random union of gametes (eqn (7)). With equal numbers of male and female parents,

$$N_e = \frac{N}{1 + (1/2n)}$$

and, if females mate only once ($n = 1$), $N_e = 2N/3$, a result observed in some *Drosophila* experiments (see Nunney, 1993 for references).

For X-linked genes and assuming $n = 1$,

$$N_e \approx \frac{9mfN}{2 + 4m},$$

which reduces to $N_e = 9N/16$ for a 1:1 sex ratio, instead of $3N/4$ obtained under random union of gametes.

5.2 Variation due to inherited causes

Expressions for the effective size displayed above apply when the number of offspring contributed per parent varies due to noninherited causes. Under these circumstances, changes in gene frequencies are uncorrelated in consecutive generations. When causes of variation are inherited, however, changes in gene frequencies are correlated over generations and the above expressions overpredict the effective size. These correlations occur because a fraction of the selective

advantage of individuals remains in descendants over generations. An individual with a high selective value tends to produce a large number of offspring which tend, in turn, to produce a large number of grand-offspring, and so on. Thus, the frequency and the probability of homozygosity of a neutral allele randomly associated with that individual will tend to increase until the average selective advantage of descendants is diluted by segregation and recombination. Effective size is computed from the variance of change in gene frequency or the rate of inbreeding of the neutral gene, which will generally be assumed to be unlinked to the selected genes.

Effective sizes for populations under selection have been derived following two approaches, from the variance of change in gene frequency and from the long-term contributions of ancestors. These are summarized in what follows, first for selection on the individual phenotypes and, later, for selection on indices based on individual phenotypes and family information.

5.2.1 Effective size from the variance of change in gene frequency over generations

The problem was first addressed by Robertson (1961) in the context of populations comprising full-sib families under truncation selection. He introduced the idea of the accumulation of selective advantages of individuals over generations. Thus, the selective advantage of an individual is expected to be reduced by one half each generation in its descendants (the average selective advantage of the offspring is half that of the parents), and the total selective advantage over generations increases in a series $1 + 1/2 + 1/4 + 1/8 + \ldots$, up to a limiting value of twice the selective advantage present in the initial generation. In what follows, a derivation of effective size will be made based on this idea.

We first consider effective size after one generation of selection and, thereafter, long-term predictions will be addressed. Changes in gene frequency are the result of three independent processes acting each generation, random association between the neutral gene and families with a selective advantage or disadvantage, random sampling of individuals among families, and Mendelian sampling of the neutral gene in the families in which one or both parents are heterozygotes. The last two processes also occur when nonrandom contribution of parents results from noninherited causes, and are represented by the two terms in the denominator of, for instance, eqn (16). We consider now the first process, that caused by inherited variation. Assume that the population consists of a constant number $N/2$ of families (couples), and let $V(k_s)$ be the variance of family size in the first generation of selection. This will have two components, variance due to

the differences in fitnesses among families, and variance due to random sampling of individuals among families. As population size is constant, $\Sigma k_{si} = N$ (sums are for $i = 1$ to $N/2$), where k_{si} is the number of offspring contributed by family i. The k_{si} have mean 2 over families and variance:

$$V(k_s) = \frac{\Sigma k_{si}^2}{N/2} - 4 \ . \tag{32}$$

Let f_i be the relative fitness of family i (or $2f_i$, the expected number of offspring contributed by family i) with mean $\Sigma f_i/(N/2) = 1$ and variance $C^2 = \Sigma f_i^2/(N/2) - 1$. Let us consider, for simplicity, that the number of offspring available for selection is Poisson distributed. Thus, for family i, $E[k_{si}] = \sigma_{k_{si}}^2 = 2f_i$. Taking expectations in eqn (32),

$$V(k_s) = \frac{\Sigma E[k_{si}^2]}{N/2} - 4 = \frac{\Sigma(\sigma_{k_{si}}^2 + E[k_{si}]^2)}{N/2} - 4$$

$$= \frac{\Sigma(2f_i + 4f_i^2)}{N/2} - 4 \ .$$

Noting that $\Sigma f_i = N/2$ and $\Sigma f_i^2/(N/2) = C^2 + 1$, we obtain:

$$V(k_s) = 2 + 4C^2 = S_k^2 + 4C^2,$$

where S_k^2 is the variance of family size when selection is at random. Thus, including the selective term $(4C^2)$ into eqn (16),

$$N_{e,1} = \frac{4N}{2(1 - \alpha_I) + (S_k^2 + 4C^2)(1 + \alpha_I + 2\alpha_O)} \tag{33}$$

(Santiago & Caballero, in press). For a binomial or constant distribution of the number of offspring per family available for selection, $V(k_s)$ has an additional term on C^2/n (Santiago & Caballero, in press), where n is the average number of offspring evaluated per family, but this term is usually very small and can be neglected, so that eqn (33) can be used as a general approximation. For Poisson distributed sampling variation of family size $(S_k^2 = 2)$, random mating and large N $(\alpha_I \approx \alpha_O = 0)$, eqn (33) reduces to:

$$N_{e,1} = \frac{N}{1 + C^2} \tag{34}$$

(see Robertson, 1961).

Burrows (1984a,b) has also developed predictions of effective size after one generation of selection based on estimating probabilities of coselection of sibs. His equation for mass selection with equal numbers of

sexes and Poisson sampling variation of family size is the same as eqn (34) but higher order terms ignored by Robertson (1961) are included in the prediction of C^2. Wray et al. (1990) have assessed the predictive value of this method as well as other one- and two-generation methods (for example, an equation similar to eqn (19) but including also variances in family size from grandparents to grandoffspring) and shown, as expected, that they overpredict effective sizes of populations under continued selection.

Let us then consider predictions in the long-term, with the following argument from Santiago & Caballero (in press). Among the three random processes stated above occurring each generation, only that due to random associations between the neutral gene and families with a selective advantage or disadvantage is expected to persist over generations. The change in gene frequency in the first generation caused by this process can be represented as a covariance between the expected selective advantages of families (f_i) and their gene frequencies (p_i) as $cov_1(f_i, p_i)$. If differences among the selective advantages of the families are due to an additive component of variation, it can be shown (Santiago & Caballero, in press) that a fraction of this covariance will remain in generation 2, $cov_2(f_i, p_i) = [G(1 + r)/2] cov_1(f_i, p_i)$ where G is the remaining proportion of genetic variance after selection and r is the correlation between the expected values of f_i of male and female parents. Thus $cov_2(f_i, p_i)$ is the expected change of gene frequency in the second generation given an association in the first. The expected changes in the following generations will also be in the same direction, but the magnitude of these changes will decrease by a proportion $G(1 + r)/2$ every generation (under steady selection, G can be assumed to be approximately constant over generations). Therefore, the cumulative effect of selection can be represented as the sum of an infinite series:

$$Q = \sum_{i=0}^{\infty} \left[\frac{G}{2}(1 + r) \right]^i = \frac{2}{2 - G(1 + r)} \ . \tag{35}$$

If the reduction in variance each generation is neglected (i.e. $G = 1$) and r is ignored (assuming large N and random mating), $Q = 1 + 1/2 + 1/4 + 1/8 + \ldots = 2$, as was predicted by Robertson (1961).

Therefore, eqn (16), expressed as:

$$N_e = \frac{4N}{2(1 - \alpha_I) + (S_k^2 + 4Q^2C^2)(1 + \alpha_I + 2\alpha_O)} \tag{36}$$

(Santiago & Caballero, in press) with Q from eqn (35) (squared because it affects the variance), predicts the effective size under continued selection when the

asymptotic state of correlated changes in gene frequency is reached and assuming that population size and selective and nonselective components of variation are constant over generations. For a Poisson or multinomially distributed sampling variance of family size ($S_k^2 = 2$), random mating and large N ($\alpha_I \approx \alpha_O = 0$), eqn (36) reduces to:

$$N_e = \frac{N}{1 + Q^2 C^2} \tag{37}$$

(see Robertson, 1961).

The values of G and C^2 in eqns (35) to (37) depend on the genetic system and selective procedure. For a model of truncation selection on a normally distributed trait controlled by an infinitesimal model of gene effects (Fisher, 1918), $G = 1 - kh^2$ (Bulmer, 1980), where $k = i(i - x)$, i is the selection intensity (standardized selection differential), x is the truncation point in the standardized normal distribution, and h^2 is the heritability of the trait. Moreover, for the same model C^2 can be approximated by $i^2 \rho_{FS}$ (Robertson, 1961; Milkman, 1978), where ρ_{FS} is the phenotypic intraclass correlation of full-sib family members, which is a function of the heritability (under no selection, $\rho_{FS} = h^2/2$). Thus, the larger the intensity of selection and/or the heritability, the larger C^2 and, therefore, the smaller the effective size (see eqn (37)).

With random mating, Q reaches its asymptotic value in a few generations. However, under some systems of nonrandom mating, especially those which increase the frequency of mating between relatives, a longer period of time is needed to reach the asymptotic Q value. Laboratory experiments with *Drosophila* have shown that predictions of effective size with $Q = 2$ give reasonable approximations (Jones, 1969; Yoo, 1980; Gallego & García-Dorado, 1986; Gallego & Caballero, 1990). However, in these experiments selection was weak, the heritability of the selected trait low or the asymptotic stage had not been reached. Stochastic simulations have shown that under intense selection and high heritability, predictions with $Q = 2$ are severe underestimates (Hill, 1985b; Wray & Thompson, 1990) and the reduction in genetic variance (G) should be considered.

Note that the term ($S_k^2 + 4Q^2 C^2$) in eqn (36) can be partitioned into the variance of the family size after one generation of selection ($S_k^2 + 4C^2$, see eqn (33)), plus the cumulative effect of selection on an inherited trait ($4[Q^2 - 1]C^2$). Common environmental variation of sibs as well as nonadditive genetic components of the selective values can be included into the former because these effects are not cumulative.

When numbers of males and females are different, such that N_m males are mated to N_f/N_m females each,

effective size for sex s can be approximated by eqn (17) including a term $4Q^2 C_s^2$ analogously as for eqn (36) (Santiago & Caballero, in press), where $C_s^2 = [(C_{sm} + C_{sf})/2]^2$, C_{sm}^2 (C_{sf}^2) is the variance of the relative fitnesses of parents of sex s in their contributions to male (female) offspring, and Q is obtained from eqn (35) with $G = (G_m + G_f)/2$, G_s being the remaining proportion of genetic variance in selected individuals of sex s.

Under random mating of selected parents the expected value of r in eqn (35) is approximately zero. Under partial full-sib mating, however, there is a positive correlation between selective advantages of parents. When mating is made between full-sibs, the correlation between the expected selective values of both parents (r) is one, while for random mating it is approximately zero. Thus, if an average proportion β of the matings is between full-sibs each generation, the total correlation between the expected selective values of both parents in the new families is approximately $r = \beta$, and $Q = 2/(2 - G[1 + \beta])$.

The effective size of populations with assortative mating of selected parents can also be predicted by approximating r by $r_P \rho_{FS}$, where r_P is the phenotypic correlation among mates, i.e. approaching 1 or -1 for maximum positive or negative assortative mating, respectively.

5.2.2 Selection on fitness Based on the approach of Robertson (1961), Nei & Murata (1966) worked out an approximate equation for the effective size when the trait under consideration is fertility itself. The variance of progeny number, σ_k^2, is composed of an additive genetic component, σ_a^2, plus nonadditive, environmental and sampling components, σ_e^2. The cumulative effect of selection on fertility is equivalent to multiplying σ_a^2 by four, i.e. $Q^2 \approx 4$ from eqn (35), which would be a reasonable figure as the heritability of fertility is expected to be low. Thus, at the limit, the variance of progeny number is $4\sigma_a^2 + \sigma_e^2 = 3\sigma_a^2 + \sigma_k^2 = (3h^2 + 1)\sigma_k^2$, where h^2 is the heritability of fertility. Substituting this into eqn (15) for a monoecious population,

$$N_e = \frac{4N - 2}{2 + (3h^2 + 1)\sigma_k^2}.$$

They also extended their derivation to the case where different sets of genes are controlling fertility in males and females.

Following the approach of Nei & Murata (1966), N. H. Barton (personal communication) has investigated the effective size under situations where genetic variance for fitness is maintained by deleterious mutation–selection balance or fluctuation of the selective forces to produce a stable polymorphism, and

accounting for linkage. Data from *Drosophila* suggest that only deleterious mutation–selection balance leads to substantial reductions of the effective size. Under a Poisson distributed sampling variance of family size, N_e can be approximated by:

$$N_e = \frac{N}{1 + (2\mu/R)},$$

where u is the mutation rate per map length and generation, and R is the map length in Morgans. For example, the mutation rate for fitness for the second chromosome in *Drosophila* is considered to be about 0.4 mutants per haploid chromosome per generation. As this chromosome is about one Morgan long ($R = 1$), this gives $N_e \approx N/2$.

It is worth noting that, under selection, the inbreeding effective size from genealogies can be different from that obtained from changes in gene or genotype frequency of neutral markers if these are linked to the selected genes. Furthermore, under selection, effective size can also be estimated from the probability of fixation of advantageous genes. For example, for large Ns, where s is the selection coefficient of a mutant gene, the fixation probability of that mutant is $2s(N_e/N)$ (Crow & Kimura, 1970, p. 426) with additive gene action. This value of N_e can be different from the effective size obtained from the variance in neutral allele frequency under certain circumstances, for example if extinction and colonization occurs in subdivided populations (Barton, 1993).

5.2.3 Effective size from long-term contributions of ancestors   The arguments followed above are based on the cumulative change of gene frequency of a neutral gene over generations. An alternative viewpoint for predicting effective sizes under selection is in terms of the long-term genetic contributions made by individuals (Wray & Thompson, 1990). The effective size can be approximately expressed as a function of the mean (μ_c) and variance (S_c^2) of the contributions of ancestors in the first generation (when selection starts) to descendants in the limit,

$$N_e = \frac{2N}{\mu_c^2 + S_c^2}. \tag{38}$$

After several generations the long-term contributions of genes from an ancestor stabilize and are the same for all individuals in the population, with the values differing between ancestors. This dispersal of genes is that analysed in retrospect from pedigree data by James & McBride (1958), using the 'percentage of genes' technique (see James, 1962).

First consider the situation where there is no selection, or this occurs on a noninherited trait. The mean and variance of the contributions from ancestors to descendants in the first generation are $\mu_c = 1$ (if population size is constant over generations), and $S_c^2 = S_k^2/4$, because each parent contributes half the genes to their offspring. This variance is increased over generations to twice its initial value, i.e. $S_c^2 = S_k^2/2$ in the limit, because even under no selection, some ancestors will have larger contributions than others. Substituting this limiting expectation into eqn (38) and for $S_k^2 = 2$, $N_e = N$, as expected.

Under selection on an inherited trait, the variance of the contributions will be increased by a larger amount than without selection, because some ancestors will tend to contribute more genes to descendants than others. Equating expression (37) to (38), we arrive at the approximate relation $S_c^2 = S_k^2/2 + 2Q^2C^2$ for a multinomially or Poisson distributed sampling variance of family size and constant N each generation, which illustrates the above point.

Wray & Thompson (1990) developed a recursive algorithm to approximate variances of long-term contributions under selection and random mating of selected parents. The method, however, requires some recurrence computing for predictions to be made. Woolliams *et al.* (1993) and Wray *et al.* (1994), working on the infinitesimal model, have derived equations to predict means and variances of the contributions of ancestors to descendants under random mating of selected parents.

The approximate expression found by Woolliams *et al.* (1994) is:

$$\frac{1}{N_e} = \frac{1 + i^2 \rho_{HS,m}}{4N_m} + \frac{1 + i^2 \rho_{HS,m} + 2i^2 \rho_{HS,f}}{4N_f}$$

$$+ 2i^2(Q^2 - 1)\left[\frac{\rho_{HS,m}}{8N_m} + \frac{\rho_{HS,f}}{8N_f}\right] - \frac{1}{4T}, \tag{39}$$

where i is the average selection intensity in males and females, $\rho_{HS,m}(\rho_{HS,f})$ is the correlation among half-sibs for males (females), Q is obtained as eqn (35) with $r = 0$, and T is the number of scored individuals of each sex. The last term in the equation is a correction for the hypergeometric sampling (when there is a constant number of individuals evaluated per family), and should be ignored for a multinomial distribution of family size.

When eqn (39) is reduced to the simplest case of equal numbers of males and females and multinomial sampling variance of family size, the result is slightly different from eqn (37). Thus, taking $N_m = N_f$, $2i^2 \rho_{HS} = i^2 \rho_{FS} = C^2$ and removing the last term in eqn (39), this reduces to $N_e = N/[1 + (1 + Q^2)C^2/2]$ (Wray *et*

al., 1994), i.e. it has the term $(1 + Q^2)/2$ instead of Q^2 as in eqn (37). This difference reflects the fact that approximations are based on different approaches. Woolliams *et al.* (1993) and Wray *et al.* (1994) made derivations in terms of the contributions from ancestors in the first generation to descendants in the limit and, therefore, they used the equation which would apply in the second generation of selection making corrections to get a limiting solution. Equation (37), however, is the asymptotic equation.

5.2.4 Index selection In animal breeding practice, selection on an index including family information is often carried out. Robertson (1961) showed an approximate result for the case where individual records (P) and means of full-sib families (F) are used in the index $I = (P - F) + w(F - \bar{P})$, where $\bar{P}$ is the mean of the population and w is the weight given to family information. For large family size, the optimum value of w (that which maximizes the correlation between breeding values and index values of individuals) is approximately $w = (1 - \rho_{FS})/\rho_{FS}$ for full-sib families (Lush, 1947). The variance between family means is scaled by w^2 and, hence, the intraclass correlation of full sibs for index values is $1 - \rho_{FS}$. Thus, a first order approximation of the effective size under index selection can be obtained from eqn (36), where $C^2 = i^2(1 - \rho_{FS})$. If the heritability is small (and then ρ_{FS} is small), a large weight will be given to family information, C^2 will be large and N_e severely reduced.

Wray *et al.* (1994) have derived a more precise and general equation for the case where selection is based on an index of individual records and the means of full-sib and half-sib records. The expression can be formulated to have the same structure as eqn (19) except that variances and covariances have a cumulative and a noncumulative term. For example, S^2_{mf} is replaced by $S^2_{e(mf)} + Q^2_{mf}S^2_{g(mf)}$, and $S_{mm,mf}$ by $S_{e(mm,mf)} + Q_{mm}Q_{mf}S_{g(mm,mf)}$. Noncumulative terms (with subscript e) are the sampling variances appropriate under no selection plus covariances due to coselection of sibs which are attributed to correlations arising from the mate of the parent of sex s or, for index selection, to shared estimation errors of family means. Cumulative terms (with subscript g) represent covariances of selection between sibs which are attributable to the parent of sex s. All these variances and covariances are functions of the number of individuals, selection intensities and intraclass correlations between indices of sibs. Terms Q_{sm} and Q_{sf} are similar to that expressed in eqn (35) with extra terms accounting for index selection (see Wray *et al.*, 1994 for details).

Under mass selection, this general equation reduces to eqn (39), approximately, and under no selection, it equals eqn (19).

6 Nonrandom mating with population subdivision

In this section we will discuss predictions of effective size in populations with some degree of partial subdivision into groups, following arguments by Kimura & Crow (1963b) and Robertson (1964) (see also Cockerham, 1969; Wright, 1969, pp. 324–335). Only situations in which population structure is under controlled conditions, as in laboratory populations, will be addressed. Structured natural population models, where variables such as mutation and migration rates among subpopulations have to be accounted for, will not be dealt with. A recent account for effective size in subdivided populations considering some of these factors is given by Chesser *et al.* (1993).

To understand nonrandom mating and population structure we have to distinguish between inbreeding due to nonrandom mating and inbreeding due to finite population size. If a very large population is subject to some sort of nonrandom mating, like mating between relatives, the decrease in heterozygosity produced is due to departure from Hardy–Weinberg proportions and can be restored approximately as soon as random mating occurs. However, in a small population with random mating, the decrease in heterozygosity is due to changes in the gene frequencies and cannot be restored except by crossing with other populations. To combine both effects, we define the coefficient F_{ST} as the probability that two genes taken at random from the population are identical by descent (i.e. the average coancestry of the population, Falconer, 1989, p. 88), and F_{IS} as the probability that the two homologous genes in an individual are identical by descent and coming from copies of a gene in a recent common ancestor. Thus, F_{IS} is defined relative to the population to which the individual belongs while F_{ST} is defined relative to infinitely many conceptual replicates of the population. Finally, the overall probability of identity for the two homologous genes in an individual (F_{IT}, the absolute coefficient of inbreeding relative also to many conceptual replicates of the population) is a composite of the other two,

$$(1 - F_{IT}) = (1 - F_{ST})(1 - F_{IS}) \tag{40}$$

(Crow & Kimura, 1970, p. 106). From the above definitions it is seen that F_{IS} is a measure of the decrease in heterozygosity due to nonrandom mating in the population and is approximately the same as the value α_I defined earlier in this paper. F_{ST} is a measure of the decrease in heterozygosity due to the finite size of the population and the effective size can be computed from its rate of increase. It can also be shown that F_{ST} gives a measure of the genetic drift among conceptual replicates of the population (see

Crow & Kimura, 1970, pp. 107–108). Finally, F_{IT} is a measure of the absolute decrease in heterozygosity in the population. These coefficients are known as Wright's F statistics (Wright, 1969, pp. 294–295) and were conceived as correlations of genes in the framework of subdivided populations. When defined as correlations, F_{IS} (and hence F_{IT}) can be negative when mating between relatives is avoided. F_{ST}, however, is always positive.

In an infinite nonrandom mating population, $F_{ST} = 0$ and $F_{IT} = F_{IS}$, rising toward a value which depends on the proportion of inbred matings (see asymptotic values of α_I above). In a finite random mating population, $F_{IS} \approx 0$ and F_{ST} in any generation will be equal to F_{IT} in the next, both rising toward unity. Thus, genetic drift (represented by F_{ST}) and decline in heterozygosity (represented by F_{IT}) will have the same rate of change (given by eqn (3)) except that the drift runs one generation ahead of the decline in heterozygosity. Finally, in a finite nonrandom mating population, F_{IS} will reach an asymptotic value and F_{IT} and F_{ST} will have the same rate of change, F_{IT} being larger (smaller) than F_{ST} if F_{IS} is positive (negative).

In this section we will address situations where the population has some degree of subdivision. In the lowest population level this requires equal family sizes ($S_k^2 = 0$), so that individuals (in monoecious populations) or couples (in dioecious) can be regarded as subpopulations. Under random mating, this has been shown to minimize the rate of increase in inbreeding ($N_e \approx 2N$). We now consider departures from random mating.

If the population is subdivided permanently in groups (independent sublines with completely different pedigrees, for example full-sib lines), genetic drift and heterozygosity will follow different fates because F_{IT} will rise toward unity whereas F_{ST} will not. In other words, the lines will become homozygous but genetic drift will be minimized as different alleles will be fixed in the different groups.

If population subdivision is not complete then heterozygosity and drift will have the same final rate but the decline in heterozygosity will precede the drift, and the greater the degree of subdivision, the lower the final rate (or the larger the effective size). This can be seen as follows. Substituting $S_k^2 = 0$ and $\alpha_I \approx F_{IS}$ in eqns (11), (16) or (20), this gives $N_e \approx 2N/(1 - F_{IS})$ and, using eqn (40),

$$N_e \approx \frac{2N(1 - F_{ST})}{(1 - F_{IT})} \qquad (41)$$

or, in terms of rates of decrease in heterozygosity (by eqns (2) and (3)), $1 - \lambda \approx (1/4N)(1 - F_{IT})/(1 - F_{ST})$, as

derived by Robertson (1964) in a different way, where λ is the factor by which $(1 - F_{IT})$ or $(1 - F_{ST})$ decline each generation. Expression (41) shows that any breeding system in which $F_{IT} > F_{ST}$ will increase N_e and, therefore, will decrease the final rate of decline in heterozygosity. This will be achieved when matings between relatives are more frequent than at random, i.e. when $F_{IS} > 0$. Thus, in these cases, the decline in heterozygosity will precede the genetic drift and there will be a larger decrease in heterozygosity in the earlier stages but a final slower rate than under random mating. Circular pair mating, for example, in which each individual is mated to its closest neighbour following a circular scheme, can be considered to be of this type and, in this case,

$$N_e \approx \frac{(N + 12)^2}{2\pi^2}$$

(Kimura and Crow, 1963b).

It should be stressed, however, that eqn (41) leads to $N_e > 2N$ for $F_{IS} > 0$ only when $S_k^2 = 0$ (equal family sizes). Otherwise, N_e will be generally smaller than N when mating between relatives is carried out (see for example eqn (13)). The ultimate reason for this is that, when $S_k^2 = 0$, all genetic drift comes from Mendelian segregation of heterozygotes, and this will be least when the frequency of heterozygotes is least.

Let us now consider the situation when mating between relatives is avoided. Assume that the population has size $N = 2^n$, where n is an integer. We have seen that, under random mating, $1 - F_{IT}$ lags one generation behind $1 - F_{ST}$. By making the appropriate mating strategy it is possible to avoid matings between individuals which have a common ancestor in the last n generations, what is known as 'maximum avoidance of inbreeding' (Wright, 1969, pp. 199–201), so that the decline in heterozygosity $(1 - F_{IT})$ lags $n + 1$ generations behind the genetic drift $(1 - F_{ST})$. Thus, $(1 - F_{ST})/(1 - F_{IT}) = \lambda^{n+1}$. Substituting into eqn (41), $N_e = 2N\lambda^{n+1}$ and, noting that, if N is large, $\lambda^{n+1} \approx (1 - 1/4N)^{n+1} \approx 1 - (n+1)/4N$, then:

$$N_e = 2N - \frac{n+1}{2},$$

as obtained by Robertson (1964). This indicates that by maximum avoidance of inbreeding, N_e will be smaller than under random mating and, therefore, although the decline in heterozygosity will be smaller in the initial generations than under random mating, the final rate will be faster. Again, it is worth stressing that this applies when $S_k^2 = 0$. Otherwise, avoidance of inbred matings tends to increase effective size, as illustrated, for instance, by eqn (6).

In summary, mating systems with equal family sizes where mating between relatives is more frequent than at random (implying partial subdivision of the population) will give larger N_e, and those where it is less frequent, smaller N_e, than under random mating.

7 Overlapping generations

The problem of overlapping generations in the computation of effective size has been addressed by a number of authors (see a list of references in Crow & Denniston, 1988). With discrete-generation populations, the rate of inbreeding soon reaches an asymptotic value of $\Delta F = 1/2N_e$, but with overlapping generations this takes longer until the age-structure stabilizes. Felsenstein (1971), Johnson (1977), Choy & Weir (1978) and Emigh & Pollak (1979) give exact recurrence equations for inbreeding in the early phases but we consider, however, the asymptotic stage.

Perhaps the simplest possible model comes from Moran (1962). He considered a haploid population of constant size N in which one individual randomly chosen dies in each time unit and is replaced by an offspring of an individual also randomly chosen (including the dead one). Moran showed that, in this case, the heterozygosity declines by a factor of $2/N^2$ per time unit. If time is scaled such that one generation has N time units, then heterozygosity declines by $2/N$ per generation and, equating this to the corresponding value in an ideal haploid population $(1/N_e)$, it gives $N_e = N/2$. This model, however, is unrealistic as it does not allow for ageing.

Felsenstein (1971), Johnson (1977) and Emigh & Pollak (1979) have derived precise equations for effective size in terms of the parental age distribution, assuming random mating, stable age structure, random deaths and births (i.e. no differences in fertility between individuals in the same age class and no correlation between birth and death rates in different age classes) and variation in family numbers due to noninherited causes. For haploid and monoecious diploid populations with N_c individuals born in each time period and n age classes,

$$N_e = \frac{N_c L}{1 + \sum_{i=1}^{n} z_{i+1}^2 \left(\frac{1}{l_{i+1}} - \frac{1}{l_i} \right)} \tag{42}$$

(Felsenstein, 1971; Johnson, 1977), where l_i is the probability of survival to age i, z_i is the probability that a newborn came from a parent at least of age i, i.e. $z_i = \sum_{j \geq i}^{n} {}_j p_j$, where p_j is the probability that a newborn came from a parent of age j. Finally, L is the generation interval (mean age of parents when their progeny are

born), i.e. $L = \sum_{i=1}^{n} i p_i = \sum_{i=1}^{n} z_i$. For Moran's model, we note that $N_c = 1$, $L = N$ and $z_i = l_i = (1 - 1/N)^{i-1}$. Substituting these into eqn (42) (taking $n = \infty$) and rearranging, we obtain $N_e = N/(2 - 1/N) \approx N/2$, as expected.

For separate sexes, the corresponding equation is:

$$N_e = 4mfN_c L \left/ \left[1 + m \sum_{i=1}^{n_m} z_{m,i+1}^2 \left(\frac{1}{l_{m,i+1}} - \frac{1}{l_{m,i}} \right) \right. \right.$$
$$\left. + f \sum_{i=1}^{n_f} z_{f,i+1}^2 \left(\frac{1}{l_{f,i+1}} - \frac{1}{l_{f,i}} \right) \right] \tag{43}$$

(Johnson, 1977; Emigh & Pollak, 1979), where m (f) is the proportion of males (females) among the newborn, and the other terms are equivalent to those in eqn (42) but referring to parents of sex male or female (according to the subscript m or f, respectively). L here is $(L_{mm} + L_{mf} + L_{fm} + L_{ff})/4$ where, for example, L_{mf} is the average age of male parents of daughters. To derive expression (43) random mating is assumed, but not necessarily random mating across age groups, because the effect of correlation between ages of mates is likely to be small (Hill, 1972b; Emigh & Pollak, 1979).

Note that, with nonoverlapping generations, $L = 1$, terms z in eqns (42) and (43) are zero and, N_e equals N and $4mfN$, respectively, as expected. Also, when $m = f = 1/2$, $n_m = n_f$, $l_{m,i} = l_{f,i}$ and $z_{m,i} = z_{f,i}$, eqn (43) reduces to eqn (42). An analogous expression to eqn (43) for X-linked genes is given by Pollak (1990).

Hill (1972b, 1979) has given a simpler and more general result than eqns (42) and (43). This is given in terms of variances and covariances of lifetime family sizes and, therefore, it allows for differential viabilities and fertilities in individuals of the same age class as well as correlations of fertilities of individuals at successive ages. As before, it is assumed that the population has constant size, a stable age structure, and variation in family numbers is due to noninherited causes. Consider the haploid model for simplicity. A cohort of N_c individuals is born each time period (say, each year), and the generation interval is L (defined as above). The gene frequency in individual i is $x_i = 0$ or 1, and the mean gene frequency of the cohort is $q = \sum_{i=1}^{N_c} x_i / N_c$. Individual i contributes k_i offspring over its entire lifetime and the mean contribution by the whole cohort is $Q = \sum_{i=1}^{N_c} x_i k_i / N_c L$. Now, the mean value of x_i is q and the mean value of k_i is 1, because the population has constant size. Thus, the expected value of Q is q/L so the variance of genetic drift is

$V[Q - q/L]$. Proceeding as before, we obtain:

$$N_e = \frac{N_c L}{S_k^2} \tag{44}$$

(see eqn (26)). This can be analogously extended to the other models. For example, for diploid models, the derivation would be similar but including also Mendelian segregation of heterozygotes. Thus, we replace N by $N_c L$ in eqns (11) and (16), N_s by $N_{cs}L$ in eqn (17), or N_s by $N_{cs}L$ (out of the brackets) and N_s by N_{cs} (within the brackets) in eqns (19) and (30), where N_{cs} is the number of individuals of sex s entering the population each time unit. Analogously, $N - 1$, $4N - 2$ and $4N - 4$ in eqns (27), (15) and (24), can be replaced by $(N_c - 1)L$, $(4N_c - 2)L$ and $(4N_c - 4)L$, respectively. Therefore, the effective size with overlapping generations is the same as that for discrete-generation populations having the same variance in lifetime progeny numbers and the same number of individuals entering the population each generation, provided the age-structure of the population is invariant. These expressions, however, might not have as much precision as for the nonoverlapping case, and should be used as approximations. This is because variances of progeny numbers refer to total lifespan, not to individuals born at a given time, like in the discrete case. It is worth noting that N_e is maximized by decreasing the variance of family size, but it is irrelevant whether a male has all his progeny, say when 2 years of age, or an equal proportion when 1, 2 or 3 years old, provided the total number is the same.

Under Moran's model, individuals have an exponential distribution of lifetime (age-independent survival) and $S_k^2 \approx 2$ (see Hill, 1972b). Thus, from eqn (44), $N_e \approx N_c L/2$, in agreement with Moran's result. Johnson (1977) has shown that, under the assumptions of random births and deaths, Hill's equation (expression (19) accounting for overlapping generations) equals eqn (43).

Hill (1972a) discusses breeding strategies with overlapping generations in the context of the design of control populations in animal breeding. To allow comparisons between populations with different generation intervals, he proposes an annual effective size (N_y) as the size of an idealized population with a generation interval of one year leading to the same variance of gene frequency drift observed per year as in the current population. Thus, $N_y = N_e L$ and, for example, a population with $N_e = 50$ and $L = 2$ years has the same N_y as another with $N_e = 100$ and $L = 1$ year.

Nunney (1991, 1993) has investigated the effects of fecundity, sex ratio and age-structure on effective population size. The model assumes a population with a stable age distribution and a constant number of N adults per generation (which equals $N_c L$ above) with a proportion of m males and f females. The sex ratio of the newborn is 1:1 and a possible adult biased sex ratio can be due to differences in recruitment of adults or to differences in survival (longevity) between sexes. For example, an adult sex ratio of 2:1 can occur because twice as many individuals of one sex breed than of the other, or because the same number of individuals of each sex breed but one sex lives twice as long as the other. It is generally assumed that adult fecundity (including mating success) is age-independent and, always, noninherited.

Nunney argued that variances and covariances of lifetime progeny numbers are difficult to obtain in practice, and rearranged Hill's expression (eqn (19) accounting for overlapping generations) in terms of means and variances of fecundity or death rate per breeding season. He considered several mating systems like random union of gametes, monogamy and some polygynous systems, and his results are summarized below.

For random union of gametes and biased sex ratio due to differences in survival between sexes,

$$N_e = \frac{N}{2 - \left(\dfrac{1}{2L_m} + \dfrac{1}{2L_f}\right)}, \tag{45}$$

where L_s is the generation interval in units of breeding seasons for sex s. However, for a biased sex ratio due to differences in recruitment between sexes,

$$N_e = \frac{4mfN}{2 - (1/L)}. \tag{46}$$

Note that the numerator in eqn (46) equals N_e from eqn (7), the well-known expression of Wright (1931) for different numbers of the sexes and, in fact, eqn (46) is a generalization of eqn (7). We can observe that a biased sex ratio lowers N_e irrespective of L in eqn (46) but has a negligible effect in eqn (45), once L is large. This shows that a biased sex ratio due to differences in recruitment (eqn (46), for which eqn (7) is a particular case with $L = 1$) lowers N_e more strongly than the same biased sex ratio due to differences in longevity between sexes (eqn (45)). This is interesting, as many sex ratio biases, at least in longer lived animals, are probably a result of the second reason (in general, of a higher mortality rate of adult males).

In all these expressions it is assumed that individuals are sexually mature by the next breeding season after birth and, therefore, they should be corrected if this is

not the case (see Nunney, 1993 and Nunney & Elam, 1994).

Expressions equivalent to eqns (45) and (46) for monogamy have an additional positive term added in the denominators because, unless the sex ratio is 1:1, some individuals are always unmated. Thus, with equal numbers of the sexes, random union of gametes and monogamy give the same N_e, otherwise, the second gives lower values. The results discussed above, however, are not altered when applied to monogamy.

When the sex ratio is 1:1, eqns (45) and (46) reduce to:

$$N_e = \frac{N}{2 - (1/L)} \qquad (47)$$

This can also be represented as $N_e = N/(1 + v)$, where v is the survival rate of adults between breeding seasons, because under age-independent survivorship, $L = 1/(1 - v)$. Eqns (45) to (47) illustrate that, except when L is very small, $N_e \rightarrow N/2$, which generalizes Moran's model, for which $L = N$ in eqn (47). This suggests that a value closer to $N/2$ should be assumed for many organisms in which generations overlap. This decline in N_e as L increases occurs because of the increased variance of lifespan progeny number generated by the age independent survivorship. If we equate the general equation of the effective size with overlapping generations (eqn (44) in its version for diploid populations), $N_e \approx 4N_c L/ (2 + S_k^2)$, to eqn (47) (in which N equals $N_c L$ and is assumed to be a fixed value), we obtain $S_k^2 \approx 6 - 4/L$. This is the variance of lifetime family size with age-independent survivorship and no differences in fertility among individuals. For $L = 1$, $S_k^2 = 2$ as expected, but as L increases, $S_k^2 \rightarrow 6$ and $N_e \rightarrow N/2$. A more realistic model in which death rate increases with age does not strongly affect the arguments discussed before, although increases the ratio N_e/N relative to the age-independent survivorship model.

Other polygynous systems reduce the effective size more. For example, for lottery polygyny, in which males attempt to mate with each of the females they encounter, but females mate only once or a few times,

$$N_e = \frac{N}{2 - \frac{1}{L}\left(1 - \frac{1}{2n}\right)}, \qquad (48)$$

where n is the number of times that females mate. Some results from this equation when generations do not overlap $(L = 1)$ were given before. With overlapping generations and large L, N_e tends again to $N/2$, as before. For $n = 1$, eqn (48) can also be

represented as $N_e = 2N/(3 + v)$, where v is defined as above.

Other types of mating systems with more extreme polygyny (for example, those in which harems are formed) severely reduce N_e, but in all cases, as L becomes large, N_e gets closer to $N/2$, regardless of the mating system. For random union of gametes, monogamy and lottery polygyny, N_e always decreases with increasing L, but with some extreme polygynous systems, N_e can sometimes increase towards $N/2$, as L increases. It is also interesting to note that these mating systems have asymmetrical effects on N_e depending on which is the more frequent sex, by contrast to what happens with random union of gametes or monogamy.

Finally, Orive (1993) has developed a method for determining effective size in organisms with complex life histories, like those with overlapping generations and both sexual and clonal reproduction. The method is based on finding the time at which two alleles had a common ancestor in the past (a 'coalescent' model), and equating it to that in the idealized population. Her results for populations with sexual reproduction agree with those reviewed above. The existence of clonal reproduction requires redefinition of the concept of generation interval, as clonal reproduction can be either accounted for or not, producing different results.

8 Prospects for future developments

This paper has reviewed predictive equations of the effective population size under a range of circumstances occurring in natural and experimental populations. Its main objective has been to integrate equations and methods into a similar framework, so that the inter-relations and significance of the different formulae and methods could be followed. Although these cover most important sources of genetic drift, needless to say we are far from covering all the complexities that occur even in the controlled conditions of a laboratory population.

Some of the equations of this review, like eqn (36) and, especially, its equivalent for different numbers of male and female parents, have a wide generality, and cover multiple combinations of deviations from the idealized conditions: nonrandom contribution from parents, different numbers of the sexes, partial inbreeding and other types of nonrandom mating, selection and different selective pressures in each sex. It would be desirable to continue investigating other combinations. For example, expression (36) and, in fact, all other equations in this review except that of N. H. Barton (personal communication) for situations like mutation-selection balance, refer to neutral genes unlinked to the selected system. E. Santiago (personal

communication) is developing equations in the line of eqn (36) with consideration of linkage to selected loci. Preliminary results show that N_e can be substantially reduced by selection in organisms with few chromosomes such as *Drosophila* for which it can be about 20–40 per cent smaller than predicted with the equations given above. For organisms with many chromosomes (say, more than 10), however, the effect of linkage can be neglected. Another field under current investigation is that of artificial selection when generations overlap (J. A. Woolliams, personal communication).

Complexities of structuring in natural populations where systematic forces of change in gene frequency, such as migration, play an important role, have not been the concern of this review because its aim has been to deal only with populations under more or less controlled conditions, like experimental populations. Practical problems in the estimation of effective size in natural populations have also been ignored. This is a major field which has received a great deal of attention but where there is still room for research.

It is clear that the theory of the prediction of effective size has evolved substantially since its first appearance in 1931, but it is also clear that there is still a long way to go before we can fully assess all factors affecting genetic drift, particularly, in natural populations.

Acknowledgements

References

AVERY, P. J. 1984. The population genetics of haplo-diploids and X-linked genes. *Genet. Res.*, **44**, 321–341.

BARTON, N. H. 1993. The probability of fixation of a favoured allele in a subdivided population. *Genet. Res.* **62**, 149–157.

BARTON, N. H. AND CHARLESWORTH, B. 1984. Genetic revolutions, founder effects and speciation. *Ann. Rev. Ecol. Syst.*, **15**, 133–164.

BEGON, M., KRIMBAS, C. B. AND LOUKAS, M. 1980. The genetics of *Drosophila subobscura* populations. XV. Effective size of a natural population estimated by three independent methods. *Heredity*, **45**, 335–350.

BRISCOE, D. A. MALPICA, J. M., ROBERTSON, A., SMITH, G. J., FRANKHAM, R., BANKS, R. G. AND BARKER, J. S. F. 1992. Rapid

loss of genetic variation in large captive populations of *Drosophila* flies. Implications for the genetic management of captive populations. *Conserv. Biol.*, **6**, 416–425.

BULMER, M. G. 1980. *The Mathematical Theory of Quantitative Genetics.* Clarendon Press, Oxford.

BURROWS, P. M. 1984a. Inbreeding under selection from unrelated families. *Biometrics*, **40**, 357–366.

BURROWS, P. M. 1984b. Inbreeding under selection from related families. *Biometrics*, **40**, 895–906.

CABALLERO, A. 1994. On the effective size of populations with separate sexes, with particular reference to sex-linked genes. *Genetics* (in press).

CABALLERO, A. AND HILL, W. G. 1992a. Effective size of non-random mating populations. *Genetics*, **130**, 909–916.

CABALLERO, A. AND HILL, W. G. 1992b. A note on the inbreeding effective population size. *Evolution*, **46**, 1969–1972.

CHESSER, R. K., RHODES, O. E., SUGG, D. W. AND SCHNABEL, A. 1993. Effective sizes for subdivided populations. *Genetics*, **135**, 1221–1232.

CHIA, A. B. AND POLLAK, E. 1974. The inbreeding effective number and the effective number of alleles in a population that varies in size. *Theor. Pop. Biol.*, **6**, 149–172.

CHOY, S. C. AND WEIR, B. S. 1978. Exact inbreeding coefficients in populations with overlapping generations. *Genetics*, **89**, 591–614.

COCKERHAM, C. C. 1969. Variance of gene frequencies. *Evolution*, **23**, 72–84.

CROW, J. F. 1954. Breeding structure of populations. II. Effective population number. In: Kempthorne, O., Bancroft, T. A., Gowen J. W., and Lush, J. L. (eds) *Statistics and Mathematics in Biology*, pp. 543–556. Iowa State College Press, Ames, IA.

CROW, J. F. AND DENNISTON, C. 1988. Inbreeding and variance effective population numbers. *Evolution*, **42**, 482–495.

CROW, J. F. AND KIMURA, M. 1970. *An Introduction to Population Genetics Theory.* Harper and Row, New York.

CROW, J. F. AND MORTON, N. E. 1955. Measurement of gene frequency drift in small populations. *Evolution*, **9**, 202–214.

EMIGH, T. H. AND POLLAK, E. 1979. Fixation probabilities and effective population numbers in diploid populations with overlapping generations. *Theor. Pop. Biol.*, **15**, 86–107.

ETHIER, S. N. AND NAGYLAKI, T. 1980. Diffusion approximations of Markov chains with two time scales and applications to population genetics. *Adv. Appl. Prob.*, **12**, 14–49.

EWENS, W. J. 1979. *Mathematical Population Genetics.* Springer, Berlin.

EWENS, W. J. 1982. On the concept of effective population size. *Theor. Pop. Biol.*, **21**, 373–378.

FALCONER, D. S. 1989. *Introduction to Quantitative Genetics*, 3rd edn. Longman, Harlow, Essex.

FELSENSTEIN, J. 1971. Inbreeding and variance effective numbers in populations with overlapping generations. *Genetics*, **68**, 581–597.

FISHER, R. A. 1918. The correlation between relatives on the supposition of Mendelian inheritance. *Trans. R. Soc. Edinb.*, **52**, 399–433.

FISHER, R. A. 1930. *The Genetical Theory of Natural Selection.* Clarendon Press, Oxford.

FISHER, R. A. 1939. Stage of development as a factor influencing the variance in the number of offspring, frequency of mutants and related quantities. *Ann. Eugen.,* **9**, 406–408.

FRANKHAM, R. 1994. Conservation of genetic diversity for animal improvement. *Proc. 5th World Cong. Genet. Applied to Lives. Prod.* **21**, 385–392.

GALLEGO, A. AND CABALLERO, A. 1990. The cumulative effect of artificial selection on the reduction of population effective size. Experimental and simulated data. *J. Anim. Breed. Genet.,* **107**, 180–187.

GALLEGO, A. AND GARCIA-DORADO, A. 1986. Evolution de l'effectif génétique de lignées de *Drosophila melanogaster* soumises à une sélection artificielle. *Génét. Sél. Évol.,* **18**, 249–260.

GHAI, G. L. 1969. Structure of populations under mixed random and sib mating. *Theor. Appl. Genet.,* **39**, 179–182.

GOWE, R. S., ROBERTSON, A. AND LATTER, B. D. H. 1959. Environment and poultry breeding problems. 5. The design of poultry control strains. *Poultry Sci.,* **38**, 462–471.

GREGORIUS, H.-R. 1991. On the concept of effective number. *Theor. Pop. Biol.,* **40**, 269–283.

HALDANE, J. B. S. 1924. A mathematical theory of natural and artificial selection. II. The influence of partial self-fertilisation, inbreeding, assortative mating, and selective fertilisation on the composition of Mendelian populations, and on natural selection. *Proc. Camb. Phil. Soc., (Biol. Sci.),* (later *Biol. Rev.*), **1**, 158–163.

HALDANE, J. B. S. 1939. The equilibrium between mutation and random extinction. *Ann. Eugen.,* **9**, 400–405.

HEDRICK, P. W. AND COCKERHAM, C. C. 1986. Partial inbreeding: equilibrium heterozygosity and the heterozygosity paradox. *Evolution,* **40**, 856–861.

HILL, W. G. 1972a. Estimation of genetic change. I. General theory and design of control populations. *Anim. Breed. Abstr.,* **40**, 1–15.

HILL, W. G. 1972b. Effective size of populations with overlapping generations. *Theor. Pop. Biol.,* **3**, 278–289.

HILL, W. G. 1979. A note on effective population size with overlapping generations. *Genetics,* **92**, 317–322.

HILL, W. G. 1985a. Effects of population size on response to short and long term selection. *J. Anim. Breed. Genet.,* **102**, 161–173.

HILL, W. G. 1985b. Fixation probabilities of mutant genes with artificial selection. *Génét. Sél. Évol.,* **17**, 351–358.

HILL, W. G. 1986. Population size and design of breeding programmes. *Proc. 3rd World Cong. Genet. Applied to Lives. Prod.,* **12**, 245–256.

JAMES, J. W. 1962. The spread of genes in random mating control populations. *Genet. Res.,* **3**, 1–19.

JAMES, J. W. AND McBRIDE, G. 1958. The spread of genes by natural and artificial selection in a closed poultry flock. *J. Genet.,* **56**, 55–62.

JOHNSON, D. L. 1977. Inbreeding in populations with overlapping generations. *Genetics,* **87**, 581–591.

JONES, L. P. 1969. Effect of artificial selection on rates of inbreeding in populations of *Drosophila melanogaster*. 1. Effect on early generations. *Aust. J. Biol. Sci.,* **22**, 143–155.

KIMURA, M. 1963. A probability method for treating inbreeding systems, especially with linked genes. *Biometrics,* **19**, 1–17.

KIMURA, M. 1983. *The Neutral Theory of Molecular Evolution.* Cambridge University Press, Cambridge.

KIMURA, M. AND CROW, J. F. 1963a. The measurement of effective population number. *Evolution,* **17**, 279–288.

KIMURA, M. AND CROW, J. F. 1963b. On the maximum avoidance of inbreeding. *Genet. Res.,* **4**, 399–415.

LATTER, B. D. H. 1959. Genetic sampling in a random mating population of constant size and sex ratio. *Aust. J. Biol. Sci.,* **12**, 500–505.

LANDE, R. AND BARROWCLOUGH, G. F. 1987. Effective population size, genetic variation, and their use in population management. In: Soulé, M. E. (ed.) *Viable Populations for Conservation,* pp. 87–123. Cambridge University Press, Cambridge.

LI, C. C. 1976. *First Course in Population Genetics.* Boxwood, Pacific Grove, CA.

LUSH, J. L. 1947. Family merit and individual merit as bases for selection. *Am. Nat.,* **81**, 241–261, 362–379.

MALÉCOT, G. 1948. *Les Mathématiques de l'Hérédité.* Masson et Cie, Paris.

MALÉCOT, G. 1951. Un traitement stochastique des problèmes linéaires (mutation, linkage, migration) en génétique de population. *Ann. Univ. Lyon, Sci., Sec. A,* **14**, 79–117.

MILKMAN, R. 1978. Selection differentials and selection coefficients. *Genetics,* **88**, 391–403.

MORAN, P. A. P. 1962. *The Statistical Processes of Evolutionary Theory.* Clarendon Press, Oxford.

MORAN, P. A. P. AND WATTERSON, G. A. 1959. The genetic effects of family structure in natural populations. *Aust. J. Biol. Sci.,* **12**, 1–15.

MOTRO, U. AND THOMPSON, G. 1982. On heterozygosity and the effective size of populations subject to size changes. *Evolution,* **36**, 1059–1066.

NAGYLAKI, T. 1981. The inbreeding effective population number and the expected homozygosity for an X-linked locus. *Genetics,* **97**, 731–737.

NAGYLAKI, T. 1992. *Theoretical Population Genetics.* Springer, Berlin.

NEI, M. AND MURATA, M. 1966. Effective population size when fertility is inherited. *Genet. Res.,* **8**, 257–260.

NEI, M. AND TAJIMA, F. 1981. Genetic drift and estimation of effective population size. *Genetics,* **110**, 495–511.

NUNNEY, L. 1991. The influence of age structure and fecundity on effective population size. *Proc. R. Soc. B,* **246**, 71–76.

NUNNEY, L. 1993. The influence of mating system and overlapping generations on effective population size. *Evolution,* **47**, 1329–1341.

NUNNEY, L. AND CAMPBELL, K. A. 1993. Assessing minimum viable population size: demography meets population genetics. *Trends Ecol. Evol.,* **8**, 234–239.

NUNNEY, L. AND ELAM, D. R. 1994. Estimating the effective size of conserved populations. *Conserv. Biol.,* **8**, 175–184.

ORIVE, M. E. 1993. Effective population size in organisms with complex life-histories. *Theor. Pop. Biol.,* **44**, 316–340.

PÉREZ-ENCISO, M., FERNANDO, R. L. AND GIANOLA, D. 1992. The use of the uncertain relationship matrix for computing inbreeding effective population size. *Proceedings of the*

43rd Meeting of the European Association of Animal Production, p. 264. Madrid.

POLLAK, E. 1977. Effective population numbers and their interrelations. *Proceedings of the Washington State University Conference on Biomathematics and Biostatistics,* pp. 115–144. Department of Pure and Applied Mathematics, Washington State University, Spokane, WA.

POLLAK, E. 1980. Effective population numbers and mean times to extinction in dioecious populations with overlapping generations. *Math. Biosci.,* **52,** 1–25.

POLLAK, E. 1987. On the theory of partially inbreeding finite populations. I. Partial selfing. *Genetics,* **117,** 353–360.

POLLAK, E. 1988. On the theory of partially inbreeding finite populations. II. Partial sib mating. *Genetics,* **120,** 303–311.

POLLAK, E. 1990. The effective population size of an age-structured population with a sex-linked locus. *Math. Biosci.,* **101,** 121–130.

ROBERTSON, A. 1961. Inbreeding in artificial selection programmes. *Genet. Res.,* **2,** 189–194.

ROBERTSON, A. 1964. The effect of nonrandom mating within inbred lines on the rate of inbreeding. *Genet. Res.,* **5,** 164–167.

ROBERTSON, A. 1965. The interpretation of genotypic ratios in domestic animal populations. *Anim. Prod.,* **7,** 319–324.

SANTIAGO, E. AND CABALLERO, A. 1994. Effective size of populations under selection. *Genetics* (in press).

TEMPLETON, A. R. 1980. The theory of speciation *via* the founder principle. *Genetics,* **94,** 1011–1038.

UGARTE, E., GALLEGO A. AND CABALLERO, A. 1990. The prediction of effective size in selected populations. *Investigación Agraria: Producción y Sanidad Animales,* **5,** 17–23.

WAPLES, R. S. 1989. A generalized approach for estimating effective population size from changes in allele frequency. *Genetics,* **121,** 379–391.

WOOLLIAMS, J. A., WRAY, N. R. AND THOMPSON, R. 1993. Prediction of long-term contributions and inbreeding in populations undergoing mass selection. *Genet. Res.,* **62,** 231–242.

WRAY, N. R. AND THOMPSON, R. 1990. Predictions of rates of inbreeding in selected populations. *Genet. Res.,* **55,** 41–54.

WRAY, N. R., WOOLLIAMS, J. A. AND THOMPSON, R. 1990. Methods for predicting rates of inbreeding in selected populations. *Theor. Appl. Genet.,* **80,** 503–512.

WRAY, N. R., WOOLLIAMS, J. A. AND THOMPSON, R. 1994. Prediction of rates of inbreeding in populations undergoing index selection. *Theor. Appl. Genet.,* **87,** 878–892.

WRIGHT, S. 1922. Coefficient of inbreeding and relationship. *Am. Nat.,* **56,** 330–338.

WRIGHT, S. 1931. Evolution in Mendelian populations. *Genetics,* **16,** 97–159.

WRIGHT, S. 1933. Inbreeding and homozygosis. *Proc. Natl. Acad. Sci. U.S.A.,* **19,** 411–420.

WRIGHT, S. 1938. Size of population and breeding structure in relation to evolution. *Science,* **87,** 430–431.

WRIGHT, S. 1939. Statistical genetics in relation to evolution. *Exposès de Biomètrie et de Statistique Biologique.* Herman & Cie, Paris, France.

WRIGHT, S. 1969. *Evolution and the Genetics of Populations,* vol. 2, *The Theory of Gene Frequencies.* The University of Chicago Press, Chicago, IL.

YOO, B. H. 1980. Long-term selection for a quantitative character on large replicate populations of *Drosophila melanogaster.* 5. The inbreeding effect of selection. *Aust. J. Biol. Sci.,* **33,** 713–723.