

## Methods for predicting rates of inbreeding in selected populations

N. R. Wray\*, J. A. Woolliams<sup>1</sup> and R. Thompson<sup>1</sup>

University of Edinburgh, Institute of Animal Genetics, West Mains Road, Edinburgh EH9 3JN, UK

<sup>1</sup> Institute of Animal Physiology and Genetics, Edinburgh Research Station, Roslin, Midlothian, EH25 9PS, UK

Received November 15, 1989; Accepted February 15, 1990

Communicated by D. Van Vleck

**Summary.** In selected populations, families superior for the selected trait are likely to contribute more offspring to the next generation than inferior families and, as a consequence, the rate of inbreeding is likely to be higher in selected populations than in randomly mated populations of the same structure. Methods to predict rates of inbreeding in selected populations are discussed. The method of Burrows based on probabilities of coselection is reappraised in conjunction with the transition matrix method of Woolliams. The method of Latter based on variances and covariances of family size is also examined. These methods are one-generation approaches in the sense that they only account for selective advantage over a single generation, from parents to offspring. Two-generation methods are developed that account for selective advantage over two generations, from grandparent to grandoffspring as well as from parent to offspring. Predictions are compared to results from simulation. The best one-generation method was found to underpredict rates of inbreeding by 10–25%, and the two-generation methods were found to underpredict rates of inbreeding by 9–18%.

**Key words:** Inbreeding – Effective size – Selection

### Introduction

In a random mating population and in the absence of differences in viability and fecundity, all families have equal probabilities of contributing offspring to be parents of the next generation. In a population undergoing selection, families superior for the selected trait will contribute

more offspring to the next generation than inferior families and, as a consequence, the rate of inbreeding is higher in selected populations than in randomly mated populations. The parents of superior families are said to confer a selective advantage. The mean level of inbreeding in a given generation  $t$  ( $F_t$ ) and the rate of inbreeding [ $\Delta F$ , defined as  $(F_t - F_{t-1}) / (1 - F_{t-1})$ ] can easily be calculated from pedigree information after selection has occurred, but prediction of inbreeding rate in the planning stage of a breeding programme has proved to be difficult. Frequently, advantages of new breeding schemes are discussed solely in terms of responses to selection, with little regard to the effect of selection on inbreeding, and the assumed rate of inbreeding is appropriate only for random mating populations, therefore making objective assessment of innovations difficult.

Robertson (1961) was the first to discuss the prediction of rates of inbreeding in selected populations of full-sib families (equal numbers of males and females). Although his prediction has been found to severely overpredict rate of inbreeding in simulated populations undergoing intense selection or (mass) selection for traits with high heritabilities (Hill 1985), his approach was pioneering in that it attempted to account for the complete consequences of selective advantage from ancestors to all their descendants. Wray and Thompson (1990) developed the approach of Robertson (1961) to provide formulae of good predictive value. Their approach is appropriate for populations in which the numbers of males and females differ, although the resulting method is recursive and complex. Therefore, a need exists to examine potentially simpler approaches.

Burrows (1984a) presented a prediction for rate of inbreeding by estimating the probability of pairwise coancestry of selected individuals after a single generation of selection. The method includes higher order terms

\* Present address: Edinburgh School of Agriculture, West Mains Road, Edinburgh EH9 3JG, UK

that were ignored by Robertson (1961), but it considers only the selective advantage from parent to offspring. Woolliams (1989) presented a method that incorporates the same principles as Burrows (1984a, b) into a transition matrix approach. Using a drift variance argument, Latter (1959) derived an expression for rate of inbreeding in terms of variances and covariances of family size. Although this approach was not derived for the selection case, it has immediate appeal here because of the impact of selection of variances of family size.

We reappraise the method of Burrows (1984a, b) and the method proposed by Woolliams (1989), with particular emphasis on the prediction of coselection of sibs. These methods only consider the proliferation of lineal relatives in a single generation, therefore the development of these to account for proliferation over two generations of selective advantage, i.e. from grandparents to grandoffspring, is considered. A close relationship exists between the transition matrix method and the method of Latter (1959), and a two-generation formula in terms of variances of family size is also derived. Finally, the methods are compared to results from simulations.

### Assumptions and notation

Rate of inbreeding is assumed to apply to genes that are neutral with respect to the selected trait and that are unlinked to genes controlling the selected trait. However, if the selected trait is assumed to be controlled by many unlinked loci, each of small additive effect (the infinitesimal model), then the rate of inbreeding at selected loci is expected to be the same as at neutral loci. Selected parents are assumed to be mated at random. Generations are assumed to be nonoverlapping.

$M$  and  $F$  denote the total number of males and females available for selection per generation, and  $m$  and  $f$  (not subscripts) denote the number selected. The letters  $m$  and  $f$  are used in subscripts in the text as labels for males and females. Let  $\alpha$  denote the proportion selected so that  $\alpha_m = m/M$  and  $\alpha_f = f/F$ . Throughout,  $x$  and  $y$  are used as subscripts for a single sex, where either  $m$  or  $f$  could be appropriate; similarly,  $X$ ,  $Y$  each represent either  $M$  or  $F$ ,  $x$ ,  $y$  represent  $m$  or  $f$  and  $\alpha_x$ ,  $\alpha_y$  represent either  $\alpha_m$  or  $\alpha_f$ . Subscripts  $i$  and  $j$  are used to describe families and, where both are used together,  $i$  and  $j$  are assumed to index male and female parents, respectively. Thus  $m_{ij}$  denotes the number of males selected from the full-sib family of sire  $i$  and dam  $j$ . In addition,  $n_x$  represents the number of individuals of sex  $x$  available from each family for selection. Asymptotic inbreeding rate is denoted by  $\Delta F$  and effective population size by  $N_e = 1/2 \Delta F$ . Finally,  $\Phi(a_1)$  denotes  $P(w \leq a_1)$ , where  $w$  is any normally distributed random variable and  $\Phi(a_1, a_2; \rho)$  denotes  $P(w_1 \leq a_1, w_2 \leq a_2)$ , where  $w_1$  and  $w_2$  are bivariate normal random variables with correlation  $\rho$ .

### Prediction of inbreeding by considering the probability of identity by descent

#### The method of Burrows

Before generalising methods based on identity by descent, it is useful to review the methods of Burrows (1984a, b). The effective

population size was defined by  $1/N_e = Q_I$ , where  $Q_I$  is the probability that a pair of genes randomly chosen from distinct, selected individuals (in generation 1 of selection) were contributed by the same individual of the previous generation.  $Q_I$  defined here is equivalent to the  $Q$  defined by Burrows (1984b). This definition is equivalent to the inbreeding effective size of Ewens (1982). The inbreeding coefficient of progeny obtained by random mating of the selected individuals is  $0.5 Q_I$  (assuming no existing inbreeding). Burrows (1984a) shows that with full-sib families of fixed size  $n_x$  and random mating his definition reduces to that of Wright (1931), which assumes Poisson distribution of family size,  $N_e = 4MF/(M+F)$  as  $n_x \rightarrow \infty$ . The condition of large family size is expected, since family size of chosen individuals follows a hypergeometric distribution when available family size ( $n_x$ ) is constant but, as  $n_x$  increases, the Poisson distribution becomes a good approximation to the hypergeometric distribution.

The method of Burrows recognises no sexual dimorphism in the selection process and each selected individual may act as either male or female. This does not adequately account for sex differences in animal selection programmes, and alternative methods to account for this are addressed in the following sections.  $Q_I$  depends on the probability of coselection of full and half sibs in the notional gene sampling from among the selected individuals. When indices of relatives are uncorrelated (i.e. when  $h^2 = 0$  and family indices are not involved), selection is at random and these probabilities are obtained from the hypergeometric distribution. Let  $x_{ij}$  be the number of individuals (of sex  $x$ ) selected from family  $(i, j)$  of size  $n_x$ . Then the probability of selecting two full sibs is given by

$$\sum_{i,j} E \left[ \frac{x_{ij}(x_{ij}-1)}{x(x-1)} \right] \quad (1)$$

and

$$E[x_{ij}(x_{ij}-1)] = \frac{x(x-1)n_x(n_x-1)}{X(X-1)} \quad (2)$$

The probability of selecting two sibs from a common sire is given by

$$\begin{aligned} & \frac{1}{x(x-1)} \sum_i E \left[ \left( \sum_j x_{ij} \right) \left( \sum_j x_{ij} - 1 \right) \right] \\ & = \frac{1}{x(x-1)} \left\{ \sum_{i,j} E[x_{ij}(x_{ij}-1)] + \sum_i \sum_{j \neq j'} E[x_{ij} x_{ij'}] \right\}. \end{aligned} \quad (3)$$

The difference between Eqs. 3 and 1 is the probability of selecting two paternal half sibs. The evaluation thus requires the further result,

$$E[x_{ij} x_{ij'}] = \frac{x(x-1)n_x^2}{X(X-1)} \quad (4)$$

However, when correlations between index values of half sibs ( $\rho_{HS}$ ) and of full sibs ( $\rho_{FS}$ ) exist then the estimation of  $Q_I$  is more complex. If the selection indices of sibs follow a bivariate normal distribution, then  $Q_I$  can be approximated by bivariate normal probabilities. The approximation requires the following result, which is more general than that of Burrows (1984a, b), and this generality will be used later when considering two sexes.

Consider a single family containing two groups of size  $n_1$  and  $n_2$ . Assign an index  $G$  to each individual,  $G_{kl}$  ( $k=1, 2; l=1, \dots, n_k$ ) normally distributed, so that

$$E[G_{kl}] = 0, \quad E[G_{kl}^2] = 1, \quad E[G_{kl} G_{kl'}] = \rho_k$$

$$\text{and } E[G_{1l} G_{2l'}] = \rho, \quad \text{where } 0 \leq \rho \leq \rho_k.$$

Define  $T_k$  as the number of individuals of group  $k$  whose index exceeds a threshold  $a_k$ . The joint probability generating

function,  $M(\theta, \phi)$  of  $T_1$  and  $T_2$  for  $p_{uv} = P(T_1 = u; T_2 = v)$  is,  $M(\theta, \phi) = \sum_1^{n_1} \sum_1^{n_2} p_{uv} \theta^u \phi^v = E_{z_1, z_2} [\{\Phi(b_1) + \theta[1 - \Phi(b_1)]\}^{n_1} \{\Phi(b_2) + \phi[1 - \Phi(b_2)]\}^{n_2}]$ , where  $b_k = \frac{a_k - z_k \varrho_k^{1/2}}{(1 - \varrho_k)^{1/2}}$  for  $k=1, 2$  and  $z_1$  and  $z_2$  are distributed as a standardised bivariate normal distribution with correlation coefficient  $\varrho_1^{-1/2} \varrho_2^{-1/2}$ .

The following probabilities can then be derived:

$$\begin{aligned} E[T_k] &= n_k (1 - \Phi(a_k)) \\ E[T_k(T_k - 1)] &= n_k(n_k - 1) (1 - 2\Phi(a_k) + \Phi(a_k, a_k; \varrho_k)) \\ E[T_1 T_2] &= n_1 n_2 (1 - \Phi(a_1) - \Phi(a_2) + \Phi(a_1, a_2; \varrho)) \end{aligned} \quad (5)$$

If the  $a_k$  are considered as selection thresholds based on the  $G_{kl}$ , it can be seen that the above has some similarities to the selection process. The key difference is that there are no constraints on the sum of the 'T' random variables across the two groups or across a number of families, as there are in the selection process, where only a predetermined total number are selected. Exact expressions for these quantities can be derived using order statistics, but these are very complex.

To use the formulae 5 in the method of Burrows (1984a, b), assume the population has a factorial or nested mating structure involving full and half sibs (full-sib family size  $n_x$ ). Define a common threshold  $a_x$  and let the correlation of full-sib family members be  $\varrho_{FS}$  and amongst half sibs  $\varrho_{HS}$ .  $\varrho_{HS}$  can vary according to whether the common parent is the sire of dam, but it is only necessary to consider the former case. Define  $T_{ij}$  analogous to the  $T_k$  as the number of individuals in family  $(i, j)$  greater than the threshold  $a_x$ , and apply the following conditions.

(A) for  $n_x$  and  $\alpha_x$  constant and for all  $0 \leq \varrho_{HS} \leq \varrho_{FS}$ , then

$$\frac{1}{x} \sum_{i,j} T_{ij} \rightarrow \alpha_x \quad \text{as } X \rightarrow \infty,$$

(B) as  $\varrho_{FS} \rightarrow 0$  (and, hence,  $\varrho_{HS} \rightarrow 0$ )

$$E[x_{ij}(x_{ij} - 1)] = k_1 E[T_{ij}(T_{ij} - 1)]$$

$$E[x_{ij} x_{ij'}] = k_2 E[T_{ij} T_{ij'}],$$

which can be used to derive  $\alpha_x$ ,  $k_1$  and  $k_2$ , where  $k_1$  and  $k_2$  are constants.

Substituting Eq. 5 into Eqs. 2 and 4 gives the results of Burrows (1984b). Formulae 1 and 3 can be used to derive an approximation of coselection probabilities and, hence,  $Q_f$  for all  $\varrho_{HS}$  and  $\varrho_{FS}$ . For example, the probability of coselecting full sibs is

$$\begin{aligned} \frac{1}{x(x-1)} \sum_{i,j} E[x_{ij}(x_{ij} - 1)] &\approx \frac{(n_x - 1)}{\alpha_x^2 (X - 1)} (2\alpha_x + \Phi(a_x, a_x; \varrho_{FS}) - 1) \\ &\approx \frac{(2\alpha_x + \Phi(a_x, a_x; \varrho_{FS}) - 1)}{s \alpha_x^2}, \end{aligned}$$

for large  $n_x$ , where  $s$  is the total number of full-sib families.

#### Methods using transition matrices

Woolliams (1989) presented generalisations of the recurrence relations of Wright (1931) using transition matrices to describe  $P_{xy}(t)$ , defined as the probability that the genes chosen from distinct selected individuals of sex  $x$  and  $y$  in generation  $t$  are identical by descent. Under three assumptions, the strongest being that the selective advantage of a selected individual over others selected is not inherited in any part by its offspring, then  $P_{mf}(t)$ ,  $P_{mm}(t)$  and  $P_{ff}(t)$  could be described in terms of  $P_{mf}(t-1)$ ,  $P_{mf}(t-2)$ ,  $P_{mm}(t-1)$  and  $P_{ff}(t-1)$ . An equivalent derivation substitutes  $P_{mm}(t-1)$  for  $P_{mf}(t-2)$ , where  $P_{mm}(t-1)$  is the proba-

bility that two genes sampled from a single individual of either sex are identical by descent. The coefficients of these relationships were defined by the probabilities that the two individuals sampled had the same maternal or paternal parent (see Woolliams 1989). Then, if  $\mathbf{h}'_t = [1 - P_{mm}(t) \quad 1 - P_{mf}(t) \quad 1 - P_{ff}(t) \quad 1 - P_{mm}(t)]$ ,  $\mathbf{h}'_t = \mathbf{A} \mathbf{h}'_{t-1}$  where  $\mathbf{A}$  is a matrix of form,

$$\begin{bmatrix} r_1 & 1/2 & r_2 & (1 - 2r_1 - 2r_2)/2 \\ r_3 & 1/2 & r_4 & (1 - 2r_3 - 2r_4)/2 \\ r_5 & 1/2 & r_6 & (1 - 2r_5 - 2r_6)/2 \\ 0 & 1/2 & 0 & 0 \end{bmatrix} \quad (6)$$

where  $0 \leq r_i \leq 1/4$  for all  $i$ .  $\Delta F$  is described by the behaviour of  $P_{mf}(t)$ , which is associated with the largest real eigenvalue of  $\mathbf{A}$  ( $\lambda_{\max}$ ), with  $\Delta F \approx (1 - \lambda_{\max})$ . The equivalence of the matrix  $\mathbf{A}$  given in Eq. 6 and that of Woolliams (1989) can be seen by noting that the latter is of the form  $\mathbf{D} \mathbf{A} \mathbf{D}^{-1}$ , where  $\mathbf{D}$  is a diagonal matrix and thus the two matrices have identical eigenvalues. The existence and bounds of  $\lambda_{\max}$  are shown in Appendix I, together with proof that  $P_{mf}(t)$  is  $O(\lambda_{\max})$ .

As in the method of Burrows (1984a, b) the coefficients of transition matrices require probabilities of coselection of full sibs and half sibs when sampling two males, two females and one male and one female. Woolliams (1989) generated these by Monte Carlo simulation, but this gives little indication of their behaviour in general terms. Using Eq. 5 these probabilities can be approximated by bivariate normal probabilities. For (male, male) and (female, female) sampling, the formulae derived in the last section for coselection probabilities of full sibs and half sibs can be used after appropriate substitution of parameters. For (male, female) sampling it is necessary to note that condition (A) defines two thresholds  $a_m$  and  $a_f$  and to add a further condition, namely,

(C) as  $\varrho_{FS}$  (and, hence,  $\varrho_{HS} \rightarrow 0$  for male and female pairs,

$$E[m_{ij} f_{ij}] = k_3 E[T_{m_{ij}} T_{f_{ij}}]$$

$$E[m_{ij} f_{ij'}] = k_4 E[T_{m_{ij}} T_{f_{ij'}}].$$

Since the male and female involved in the notional gene sampling do not contribute to the same limit on numbers selected, the moments for  $\varrho_{FS} = 0$  are of binomial form rather than hypergeometric, with the result that the constants  $k_3 = k_4 = 1$ . In addition,  $a_m$  and  $a_f$  are defined by  $\Phi(a_m) = 1 - \alpha_m$  and  $\Phi(a_f) = 1 - \alpha_f$ . By noting

$$\sum_i E \left[ \left( \sum_j m_{ij} \right) \left( \sum_j f_{ij} \right) \right] = \sum_{i,j} E[m_{ij} f_{ij}] + \sum_i \sum_{j \neq j'} E[m_{ij} f_{ij'}],$$

all coselection probabilities required for constructing the transition matrix can be calculated. For example, the probability of coselection of full sibs when sampling a male and female is given by

$$\frac{1}{m f} \sum_{i,j} E[m_{ij} f_{ij}] \cong \frac{(\alpha_m + \alpha_f + \Phi(a_m, a_f; \varrho_{FS}) - 1)}{s \alpha_m \alpha_f},$$

where  $s$  is the number of families and  $\varrho_{FS}$  is the correlation of indices between male and female full sibs.

The expressions of coselection probabilities derived so far can be simplified. Consider an individual of sex  $x$ . The probability of coselecting a relative of given degree of sex  $y$  is a product of a function of selection proportions and their index correlation, i.e.  $(\alpha_x + \alpha_y + \Phi(a_x, a_y; \varrho) - 1)/(\alpha_x \alpha_y)$ , and a function of family sizes, where the latter is the ratio of [number of relatives of sex  $y$ ] and [number of individuals of sex  $y$  (excluding self if  $x=y$ )] where the numbers are those prior to selection. For  $\alpha_x = \alpha_y$ , the function of the proportions is equal to the reciprocal of the  $R(\alpha, \varrho)$  of Burrows (1984a).

*A comparison of Burrows' method and transition matrix methods*

The method of the last section explicitly caters for two sexes with differences in their treatment, whilst Burrows' method does not. Woolliams (1989) gave examples which showed that the methods are not, in general, equivalent. Furthermore, in random selection with son replacing sire and daughter replacing dam in a hierarchical mating structure (such as that considered by Gowe et al. 1959), application of Burrows' method gives  $\Delta F = 1/8 M$  when  $F \gg M \gg 1$ , whilst the method in the previous section gives  $\Delta F = 3/32 M$  as shown by Gowe et al. (1959).  $Q_I$  can be written in terms of the elements of matrix **A**, where  $A_{ij}$  is the  $(i, j)$  element of **A**,

$$\frac{1}{2} Q_I = \frac{1}{2} \left[ \frac{M(M-1)}{(M+F)(M+F-1)} A_{14} + \frac{2MF}{(M+F)(M+F-1)} A_{24} + \frac{F(F-1)}{(M+F)(M+F-1)} A_{34} \right].$$

However, when the two sexes are selected on the same index with  $n_m = n_f = n_x$ ,  $\alpha_m = \alpha_f = \alpha_x$ , and  $M = F = X$ , the two methods are asymptotically equivalent as  $X \rightarrow \infty$  for constant  $\alpha$ . In these circumstances  $P_{mm}(t) = P_{ff}(t)$  and the matrix **A** is of the form,

$$\begin{bmatrix} 1/4 - \phi \delta/2 & 1/2 & 1/4 - \phi \delta/2 & \phi \delta \\ 1/4 - \delta/2 & 1/2 & 1/4 - \delta/2 & \delta \\ 1/4 - \phi \delta/2 & 1/2 & 1/4 - \phi \delta/2 & \phi \delta \\ 0 & 1/2 & 0 & 0 \end{bmatrix}$$

where  $\delta$  is the probability of sampling genes that were from the same individual of generation  $t-1$ , when sampling one gene from an individual of each sex in generation  $t$ , and  $\phi \delta$  is the similar probability when sampling one gene from two individuals of the same sex. Since each sex is treated similarly,  $\phi \leq 1$ , and the characteristic equation of the transition matrix is

$$C(\lambda) = \lambda^4 - (1 - \phi \delta) \lambda^3 - \frac{1}{2} \phi \delta \lambda^2 + \frac{1}{4} \delta (1 - \phi) \lambda.$$

Using Newton-Raphson approximation with an initial estimate of 1, for small  $\delta$  (ignoring terms of  $\delta^2$  or higher) a maximum root of  $1 - (1 + \phi) \delta/4$  is obtained. This gives  $\Delta F = (1 + \phi) \delta/4$ .

Using the definition of the section describing Burrows' method

$$\Delta F = \frac{[(x-1)\phi + x] \delta}{2(2x-1)}. \text{ Thus, as } x \rightarrow \infty, \text{ for constant } \alpha, \text{ the two definitions become asymptotically equivalent.}$$

*Comparison of the Latter-Hill equation and transition matrix methods*

Expressions for inbreeding rates in terms of variances and covariances of family size were developed by considering the variance in changes of gene frequency by Latter (1959) for discrete generations and Hill (1972, 1979) for overlapping generations,

$$\Delta F = \frac{1}{32 ML} \left[ 2 + \sigma_{mm(1)}^2 + 2 \left( \frac{M}{F} \right) \sigma_{mf(1)} + \left( \frac{M}{F} \right)^2 \sigma_{mf(1)}^2 \right] + \frac{1}{32 FL} \left[ 2 + \sigma_{ff(1)}^2 + 2 \left( \frac{F}{M} \right) \sigma_{fm(1)} + \left( \frac{F}{M} \right)^2 \sigma_{fm(1)}^2 \right], \quad (7)$$

where  $L$  is the generation interval,  $\sigma_{xy(1)}^2$  is the variance in family size of offspring of sex  $y$  contributed to the next generation from parents of sex  $x$  and  $\sigma_{xm,xf(1)}$  is the covariance in the family size of male and female offspring contributed by parents of sex  $x$ . The subscript (1) emphasises that these are variances and covariances taken over a single generation. Equation 7 was derived for the situation where variation in family size is the result of nonheritable causes. In the selection case, the drift variance derivation

must be interpreted as the variance in change of gene frequency at a locus neutral with respect to the selected trait.

Variances of family size can be predicted involving arguments similar to those for probabilities of coselection presented earlier. In fact, the elements of the transition matrix defined in Eq. 6 can be written in terms of variances of family size. From drift variance by Eq. 7 in terms of the elements in Eq. 6

$$\Delta F = \frac{1}{8} \left[ \left( \frac{M-1}{M} \right) A_{14} + 2A_{24} + \left( \frac{F-1}{F} \right) A_{34} \right].$$

In randomly selected, randomly mated populations of constant size, each generation, the rate of inbreeding calculated from drift variance and identity by descent arguments differ only in second-order terms (Crow and Kimura 1971). However, the formulation given here demonstrates an intrinsic difference in the respective measures. The characteristic equation for the transition matrix treats  $r_1$  ( $r_6$ ) differently from  $r_2$  ( $r_5$ ), but in the drift equation above only the sums of  $r_1+r_2$  and  $r_5+r_6$  are important. In nonrandom selection these differences can be of significance. For example, consider a population with  $M$  males each mated at random to  $M$  females, with each female having one male and  $M$  female offspring. If selection rules are imposed such that all  $M$  male replacements are chosen from one male parent, and one female replacement is chosen from each female parent, then predicted rates of inbreeding from the transition matrix and Latter-Hill methods are  $0.0249 + 0.0520/M$  and  $(M+1)^2/32 M^2$ , respectively. If, on the other hand, selection rules are imposed such that one male replacement is chosen from each male, but all female replacements are chosen from a single male, then the two rates of inbreeding are  $0.0312 + 0.0665/M$  and  $(M+3)/32 M$ . As  $M \rightarrow \infty$ , the Latter-Hill predictions are identical for the two scenarios at 0.0313, but the transition matrix predictions differ by 25% at 0.0249 and 0.0312. This example is perhaps extreme, but it demonstrates that, for selected populations, identity of the transition matrix and Latter-Hill methods are not guaranteed.

**Two generation methods**

The methods presented so far account for only one consequence of selection on inbreeding, namely, the increased frequency of selecting sibs. The second consequence of selection on inbreeding, namely, the influence of a superior ancestor (older than parent) on the probability that his descendants are selected, is ignored, and these methods are therefore likely to underpredict asymptotic rates of inbreeding. More offspring are likely to be selected from a genetically superior parent than from a genetically average of genetically inferior parent. However, more grand-offspring are likely to be selected from a genetically superior grandparent for two reasons. Firstly, the grandparent was genetically superior as a parent and has already contributed more offspring, and therefore has more grandoffspring available for selection. Secondly, the grandoffspring have inherited superior genes from the grandparent, and so are more likely to be selected than their contemporaries with average or inferior grandparents.

In this section, we attempt to account for the influence of grandparents on the selection of their grandoffspring, as well as the influence of parents on the selection of their offspring. Firstly, the transition matrix method is extended to two generations and, secondly, the variance of family size method of Latter (1959) and Hill (1972, 1979) is extended to include variances of family size over two generations.

*Two-generation transition matrix method*

Considering the relationships over two generations, individuals can be considered to be full sibs, half sibs, half cousins, full

cousins, double half cousins, one-and-a-half cousins, double full cousins or unrelated at the grandparental level. The probabilities of selecting a gene identical by descent, when sampling two males, two females or a male and a female can be related to the same probabilities in generation  $t-2$  via a transition matrix,  $\mathbf{B}$ , analogous to  $\mathbf{A}$  the one-generation method.  $\Delta F$  is estimated by  $(1 - \lambda_{\max}^{1/2})$ , where  $\lambda_{\max}$  is the largest real eigenvalue of  $\mathbf{B}$ . Estimating the elements of  $\mathbf{B}$  is more complex than for  $\mathbf{A}$ , and perhaps is more easily accomplished using simulation. Under random selection  $\mathbf{B} = \mathbf{A}^2$ .

#### Two-generation variance of family size method

The drift variance method of Latter (1959) and Hill (1979) can be extended to relate rates of inbreeding to variances of family size from grandparents to grandoffspring, as well as from parents to offspring. The same assumptions as for the one-generation method apply but, in addition, it is assumed that the variance of change in gene frequency over two generations ( $V(\delta_{q_i})$ ) can be expressed as  $V(q_2 - q_0) = V(q_2 - q_1 + q_1 - q_0) = V(q_2 - q_1) + V(q_1 - q_0) \approx 2q(1-q)/2N_e$ , where  $q_t$  is the gene frequency in generation  $t$  and  $q_0 = q$ . Implicit in this assumption is that there is no covariance between changes in gene frequency over the two generations and that  $V(q_2 - q_1) = V(q_1 - q_0)$ , which are conditions that may be violated in selected populations.

In the one-generation derivation, variance of change in gene frequency can be attributed to two types of sampling processes, namely, sampling between parents and sampling within heterozygous parents (the 2's, which are the first elements in the square brackets of Eq. 7, are attributed to the latter cause). In the two-generation derivation, variance of change in gene frequency can be attributed to three types of sampling, sampling between grandparents, sampling within heterozygous grandparents and sampling within heterozygous parents. Details of the derivation are presented in Appendix II and the resulting expression for rate of inbreeding is

$$\begin{aligned} \Delta F = & \frac{1}{256 ML} \left[ \sigma_{mm(2)}^2 + 2 \left( \frac{M}{F} \right) \sigma_{mm, mf(2)} + \left( \frac{M}{F} \right)^2 \sigma_{mf(2)}^2 \right] \\ & + \frac{1}{256 FL} \left[ \sigma_{ff(2)}^2 + 2 \left( \frac{F}{M} \right) \sigma_{fm, ff(2)} + \left( \frac{F}{M} \right)^2 \sigma_{fm(2)}^2 \right] \\ & + \frac{1}{128 ML} \left[ \sigma_{mm(1)}^2 + 2 \left( \frac{M}{F} \right) \sigma_{mm, mf(1)} + \left( \frac{M}{F} \right)^2 \sigma_{mf(1)}^2 \right] \\ & + \frac{1}{128 FL} \left[ \sigma_{ff(1)}^2 + 2 \left( \frac{F}{M} \right) \sigma_{fm, ff(1)} + \left( \frac{F}{M} \right)^2 \sigma_{fm(1)}^2 \right] \\ & + \frac{1}{16 ML} + \frac{1}{16 FL}, \end{aligned} \quad (8)$$

where the variance and covariance terms have the same interpretation as in Eq. 7, except that the subscript (2) now represents the variance in family size from grandparents to grandoffspring. For example,  $\sigma_{mm(2)}^2$  is the variance in family size of male grandoffspring from male grandparents, which can be written as

$$\sigma_{mm(2)}^2 = \sigma_{mmm}^2 + \sigma_{mf m}^2 + 2 \sigma_{mmm, mfm},$$

where  $\sigma_{mmm}^2$  and  $\sigma_{mf m}^2$  are the variances in family size from male grandparents to male grandoffspring via male and female offspring, respectively, and  $\sigma_{mmm, mfm}$  is the covariance between them. Under random selection and Poisson distribution of family size, Eq. 8, like Eq. 7 reduces to Wright's rate of inbreeding of  $1/(8 ML) + 1/(8 FL)$ , and when equal numbers (as far as possible) are chosen from each family, both equations reduce to  $3/(32 ML) + 1/(32 FL)$ , as expected (Gowe et al. 1959). The elements of the two-generation transition matrix can also be expressed in terms of variances of family size.

## Applications and examples

### Parameters

Inbreeding rate will depend on the amount of genetic variation and the covariance of family members, which will affect family size. The methods described reflect this dependence. However, during the selection process such parameters are not constant. The additive-genetic variation will decrease with the initial generations principally through a loss of between-family variation. Provided that inbreeding is not rapid, an equilibrium will be attained in which the loss of between-family variation is regenerated through Mendelian sampling variance originating within families (Bulmer 1971). The genetic variances and covariances are not equal to those in the initial generation nor are they in the same proportions relative to each other. In particular, correlations between relatives decrease and can be as little as half their initial values when  $h^2$  is large. Therefore, there are two potential sets of parameters for use in one-generation formulae, initial and at equilibrium (denoted  $\rho_0$  and  $\rho_e$ ). In the examples, both sets have been used for the one-generation transition matrix method and the method of Burrows (1984 a, b), but only the equilibrium parameters have been used for the other methods.

### Probabilities of coselection

Simulations were carried out to test the accuracy of the bivariate normal approximations to the coselection probabilities. In the first instance, mass selection was simulated for  $M=20$ , and  $F=20, 40, 100, 200$  with  $n_m = n_f = 6$ . Since the approximation was designed to be exact for index correlations of zero, the simulations used large correlations for demonstration purposes with  $\rho_{HS} = 0.6$  and  $\rho_{FS} = 0.8$  for both sexes. Simulations are the result of 1,000 realisations of the selection process for each scheme, and the results are shown in Table 1.

The approximation can be seen to have good agreement with the probabilities derived from simulation, with the exception of the probability of selecting two half-sib males when  $F=100$  and 200. There are two possible causes for this. Firstly, it may be due to the size of the probability and, secondly, it may be the result of a bias in approximation that occurs when the number of individuals selected is less than the family size, e.g. for  $F=100$ , 20 males are selected and the half-sib family size for males is 30.

Therefore, a second series of simulations was carried out in which there was only one degree of relationship amongst family members (assumed to be of a single sex) and where the total number of individuals remained constant but family size varied. The selection proportion was kept constant and intraclass correlation was varied. Three thousand realisations were carried out for each scheme and the results are shown in Table 2. It is clear

**Table 1.** The probabilities of coselection of full and half sibs when sampling selected individuals in a variety of schemes. All assume  $M = 20$ ,  $n_m = n_f = 6$  with  $\rho_{HS} = 0.6$  and  $\rho_{FS} = 0.8$  for both sexes

		Probability of coselection					
		$(m, m)^a$	Full sibs			Half sibs	
			$(m, f)^a$	$(f, f)^a$	$(m, m)$	$(m, f)$	$(f, f)$
$F = 20$	Simulation	0.154	0.178	0.154	0	0	0
	Approximation	0.157	0.186	0.154	0	0	0
$F = 40$	Simulation	0.131	0.110	0.077	0.101	0.078	0.067
	Approximation	0.136	0.115	0.078	0.111	0.084	0.070
$F = 100$	Simulation	0.102	0.057	0.031	0.237	0.143	0.107
	Approximation	0.115	0.054	0.031	0.328	0.161	0.112
$F = 200$	Simulation	0.081	0.027	0.015	0.320	0.170	0.120
	Approximation	0.103	0.028	0.016	0.593	0.199	0.126

<sup>a</sup>  $(m, m)$ ,  $(m, f)$  and  $(f, f)$  denote notional sampling of two males, one male and one female and two females, respectively

**Table 2.** The probabilities of coselection of family members when sampling selected individuals in a variety of schemes. Assume a selection proportion of 0.1

Scheme			Correlation among family members			
			0.25	0.50	0.75	0.90
Number of families (s)	Family size ( $n_x$ )					
25	4	Simulation	0.056	0.095	0.152	0.208
		Approximation	0.059	0.098	0.155	0.209
10	10	Simulation	0.161	0.239	0.408	0.576
		Approximation	0.176	0.295	0.466	0.626
4	25	Simulation	0.382	0.525	0.675	0.797
		Approximation	0.469	0.785	1.242	1.669

from Table 2 that the major problem in the approximation is a bias that occurs when selection need only involve one family. The bias is sufficient to give probabilities greater than one. The cause of this bias can be found by considering the derivation of the approximation. Condition (B) considers the limit as  $X \rightarrow \infty$ , whilst  $n_x$  and  $\alpha_x$  remain constant, and this must imply that the selection process involves many families. This condition becomes untenable when  $\rho \rightarrow 1$  and  $x \leq n_x$ . When  $\rho = 1$ , the approximation to the probability of coselection for a single sex yields  $\frac{(n_x - 1)}{(X - 1) \alpha_x}$  and this exceeds one when  $\alpha_x \leq \frac{(n_x - 1)}{(X - 1)}$ .

We put forward a suggestion to overcome this problem, namely, that the value  $\alpha_x$  used in the formulae where this problem is encountered is substituted by

$$\alpha'_x = (1 - \rho) \alpha_x + \rho \max(\alpha_x, \alpha^*),$$

where  $\alpha^*$  is the ratio of the family size, including all relatives of the same or higher degree to that being con-

sidered, and the total number of individuals available for selection. A heuristic argument for  $\alpha^*$  is that as  $\rho$  approaches one, members of families become less and less distinct, thus selection becomes a two-stage process of, in the first instance, selection of a single family with a selection proportion equal to  $\alpha^*$ , followed by a random selection with selection proportion  $\alpha_x/\alpha^*$ . As an example consider the scheme in Table 1, with  $M = 20$ ,  $F = 200$ , where the probability of coselection of half-sib males is to be estimated. Each individual has 54 male half sibs and 5 male full sibs, thus  $\alpha^* = 60/1,200 = 0.05$  with  $\rho_{HS} = 0.6$ ,  $\alpha_m = 0.0367$ . Use of  $\alpha'_m$  in formulae gives the probability of coselection as 0.345, close to the value of 0.320 estimated by simulation.

*Prediction of inbreeding by transition matrix and variance of family size methods*

Predictions of rates of inbreeding derived from one- and two-generation transition matrix and variance of family size methods were compared to rates of inbreeding calculated from simulation of populations undergoing mass selection. A range of mating ratios ( $M = 20$ ,  $F = 20, 40, 100, 200$ ) with three or six offspring of each sex available for selection from each dam was used. An individual's phenotype was simulated as a sum of a breeding value and an environmental effect. The breeding value was sampled from a normal distribution with the mean being the average breeding value of its parents and variance  $\frac{1}{2}(1 - F) \sigma_a^2$ , where  $\sigma_a^2$  is the additive-genetic variance and  $F$  is the mean inbreeding coefficient of the parents. The individual environmental component was sampled from a normal distribution with mean zero and variance  $\sigma_e^2$ . A range of heritabilities  $\sigma_a^2/(\sigma_a^2 + \sigma_e^2)$  was considered:  $10^{-6}$ , 0.1, 0.2, 0.4 or 0.6. Selected parents were mated at random. The rate of inbreeding calculated

**Table 3.** Rates of inbreeding ( $\times 100$ ) from simulation ( $\Delta F_{\text{sim}}$ ) and those predicted using Burrows (1984 a, b) ( $\Delta F_{\text{B}}$ ), one and two-generation transition matrix methods ( $\Delta F_{1,\text{T}}$  and  $\Delta F_{2,\text{T}}$ ) and one (Eq. 7) and two (Eq. 8) generation variance of family size equations ( $\Delta F_{1,\text{LH}}$  and  $\Delta F_{2,\text{LH}}$ )

$h^2$	$\Delta F_{\text{sim}}$	$\Delta F_{\text{B}}$		$\Delta F_{1,\text{T}}$		$\Delta F_{2,\text{T}}$		$\Delta F_{1,\text{LH}}$	$\Delta F_{2,\text{LH}}$
		$\varrho_0^*$	$\varrho_e^{**}$	$\varrho_0$	$\varrho$	$\varrho_e$	$\varrho_e$	$\varrho_e$	
<i>M=20, F=20, n<sub>f</sub>=3</i>									
0.0	1.07	1.05	1.05	1.04	1.04	1.05	1.03	1.03	
0.1	1.23	1.11	1.11	1.10	1.10	1.13	1.08	1.11	
0.2	1.33	1.18	1.15	1.17	1.14	1.17	1.12	1.16	
0.4	1.42	1.31	1.22	1.30	1.21	1.25	1.18	1.24	
0.6	1.50	1.45	1.26	1.43	1.25	1.32	1.23	1.31	
<i>M=20, F=20, n<sub>f</sub>=6</i>									
0.0	1.13	1.15	1.15	1.14	1.14	1.14	1.13	1.13	
0.1	1.44	1.29	1.27	1.27	1.25	1.28	1.22	1.27	
0.2	1.61	1.42	1.36	1.41	1.34	1.39	1.30	1.38	
0.4	1.92	1.72	1.69	1.49	1.45	1.58	1.40	1.57	
0.6	1.98	2.04	1.57	2.01	1.55	1.70	1.54	1.68	
<i>M=20, F=40, n<sub>f</sub>=3</i>									
0.0	0.83	0.83	0.83	0.83	0.83	0.83	0.82	0.82	
0.1	0.98	0.88	0.87	0.89	0.88	0.89	0.87	0.89	
0.2	1.10	0.93	0.91	0.95	0.92	0.95	0.90	0.95	
0.4	1.18	1.04	0.96	1.07	0.98	1.03	0.95	1.01	
0.6	1.23	1.15	0.99	1.19	1.01	1.08	1.00	1.07	
<i>M=20, F=40, n<sub>f</sub>=6</i>									
0.0	0.88	0.88	0.88	0.88	0.88	0.88	0.87	0.87	
0.1	1.17	0.98	0.97	0.98	0.97	1.01	0.96	1.00	
0.2	1.30	1.08	1.03	1.09	1.04	1.09	1.02	1.09	
0.4	1.50	1.30	1.13	1.33	1.15	1.24	1.12	1.23	
0.6	1.50	1.55	1.19	1.59	1.20	1.29	1.17	1.28	
<i>M=20, F=100, n<sub>f</sub>=3</i>									
0.0	0.71	0.69	0.69	0.71	0.71	0.70	0.70	0.70	
0.1	0.82	0.73	0.72	0.76	0.75	0.76	0.74	0.75	
0.2	0.94	0.76	0.74	0.81	0.79	0.81	0.77	0.81	
0.4	1.05	0.83	0.78	0.93	0.84	0.87	0.82	0.87	
0.6	1.02	0.91	0.80	1.05	0.87	0.91	0.85	0.90	
<i>M=20, F=100, n<sub>f</sub>=6</i>									
0.0	0.73	0.72	0.72	0.72	0.72	0.71	0.70	0.70	
0.1	0.95	0.78	0.78	0.81	0.80	0.82	0.78	0.81	
0.2	1.10	0.85	0.82	0.90	0.86	0.90	0.83	0.90	
0.4	1.24	0.99	0.88	1.10	0.94	1.02	0.92	1.01	
0.6	1.23	1.15	0.91	1.35	0.99	1.06	0.96	1.05	
<i>M=20, F=200, n<sub>f</sub>=3</i>									
0.0	0.66	0.65	0.65	0.66	0.66	0.66	0.66	0.66	
0.1	0.82	0.68	0.68	0.72	0.71	0.72	0.70	0.72	
0.2	0.86	0.71	0.69	0.77	0.74	0.78	0.73	0.74	
0.4	1.00	0.76	0.72	0.89	0.80	0.84	0.79	0.84	
0.6	0.99	0.82	0.73	1.02	0.82	0.87	0.81	0.86	
<i>M=20, F=200, n<sub>f</sub>=6</i>									
0.0	0.87	0.67	0.67	0.67	0.67	0.67	0.67	0.67	
0.1	0.87	0.72	0.71	0.75	0.74	0.77	0.73	0.76	
0.2	1.03	0.77	0.75	0.84	0.80	0.84	0.78	0.84	
0.4	1.12	0.88	0.79	1.03	0.87	0.94	0.85	0.94	
0.6	1.19	1.00	0.82	1.24	0.91	0.99	0.89	0.95	

\*  $\varrho_0$  using initial genetic variances, covariances and correlations prior to selection

\*\*  $\varrho_e$  using genetic variances, covariances and correlations at equilibrium

from the simulations was  $\Delta F_{\text{sim}}$ ,

$$\Delta F_{\text{sim}} = \frac{1}{10} \sum_{t=5}^{14} \frac{F_t - \Delta F_{t-1}}{1 - \Delta F_{t-1}},$$

where  $F_t$  is the mean level of inbreeding in generation  $t$  expected from all possible matings and averaged over 100 replicates. The standard error of  $F_t$  increased, with  $t$  but did not exceed 0.0030 by generation 14 for any of the examples that are discussed (Table 3). The standard error of  $\Delta F_{\text{sim}} \times 100$  never exceeded 0.03 for any of the examples.  $\Delta F_{\text{sim}}$  was compared to that expected from theory when  $h^2 = 10^{-6}$  (random selection) and to results from other published simulations (Hill 1985; Verrier 1989) and they were found to agree well.

In Table 3 rates of inbreeding calculated in the simulations  $\Delta F_{\text{sim}}$ , are presented along with the predictions calculated from Burrows (1984 a, b)  $\Delta F_{\text{B}}$ ; from a one-generation transition matrix approach,  $\Delta F_{1,\text{T}}$ ; from the two-generation transition matrix,  $\Delta F_{2,\text{T}}$ ; from the Latter-Hill one-generation Eq. 7,  $\Delta F_{1,\text{LH}}$ ; and from the two-generation Latter-Hill Eq. 8,  $\Delta F_{2,\text{LH}}$ . The variances of family size required for the Latter-Hill equations were calculated within the simulation described above. The probabilities of coselection necessary for the one- and two-generation transition matrices were calculated as the mean of 1,000 independent simulations.

Under random section ( $h^2 = 10^{-6}$ ), all methods of predicting rates of inbreeding agree well with each other and with  $\Delta F_{\text{sim}}$ . Consider first the predictions using the equilibrium parameters ( $\varrho_e$ ). There is generally a very good agreement between the one-generation predictions  $\Delta F_{1,\text{T}}$  and  $\Delta F_{1,\text{LH}}$  and between the two generation predictions  $\Delta F_{2,\text{LH}}$  and  $\Delta F_{2,\text{T}}$  for these populations. The two-generation predictions are always better than the one-generation predictions, but both underestimate the asymptotic rates of inbreeding calculated in the simulation;  $\Delta F_{1,\text{T}}$  underpredicts  $\Delta F_{\text{sim}}$  by 10–25%, and  $\Delta F_{2,\text{T}}$  underpredicts  $\Delta F_{\text{sim}}$  by 9–18% in these examples. However, perhaps a more appropriate comparison is the proportion of the observed increase in rate of inbreeding in a selected over a non selected population accounted for the prediction;  $\Delta F_{1,\text{T}}$  predicts 31–53%  $\Delta F_{2,\text{T}}$  predicts 44–70% of the total increase in inbreeding.  $\Delta F_{\text{B}}$  is slightly superior to  $\Delta F_{1,\text{T}}$  when  $M = F$ , but becomes considerably inferior as  $F/M$  increases, since the different treatment of the sexes becomes important.

Rates of inbreeding calculated using  $\varrho_0$  in some cases appear superior in predicting the true rate of inbreeding to equivalent predictors using  $\varrho_e$ . This apparent superiority is, in fact, coincidental; the overestimation of family correlations by  $\varrho_0$  compensates for the incomplete account of the inheritance of selective advantage. Indeed, in some cases, rates of inbreeding calculated using  $\varrho_0$  are greater than the true rate of inbreeding.

## Discussion

The significance of the methods presented in this paper may best be discussed by restating the assumptions given by Woolliams (1989) required for  $\Delta F_{1,T}$  to equal inbreeding rate (since the methods are closely related, the assumptions will also apply to  $\Delta F_{1,LH}$  and  $\Delta F_{1,B}$ ): (i) mating is carried out at random; (ii) genetic variances and covariances remain constant; and (iii) selective advantage relative to others selected is not inherited. Whilst assumption (i) can be satisfied by the breeder, the other two are not, and the results show the importance of both of these.

The importance of assumption (ii) is demonstrated by the discrepancy between the examples using initial and equilibrium parameters. This discrepancy increases as  $h^2$  increases (mass selection), as would be expected from the discrepancy between initial and equilibrium correlations among relatives. However, although initial values are close to true values for high heritabilities, it is more appropriate and natural to use the equilibrium variances since, in doing so, assumption (ii) is then satisfied and there appears to be a more stable relationship with the true inbreeding rate.

Since assumptions (i) and (ii) are satisfied by parameters at equilibrium, the difference between the one-generation models using these and the simulation values is due to breaking assumption (ii), the inheritance of selective advantage since by definition  $\Delta F$  is a rate independent of previous generations in which genetic variances are equilibrating. The difference is still large: for  $h^2=0.1$ ,  $\Delta F_{1,T}$  predicts only 30% of the extra inbreeding over random mating, and for  $h^2=0.6$ , only 50% of the extra inbreeding. This trend may be expected from the results of Wray and Thompson (1990), who showed that terms describing inheritance of selective advantage over many generations are functions of  $(1-kh^2)$ , i.e. of decreasing magnitude as  $h^2$  increases.

The development of the two-stage methods was made in an attempt to introduce the concept of inherited selective advantage into the one-generation methods. Nevertheless, they account for 40–60% of the additional inbreeding through selection, giving only a marginal improvement. It is possible to conceive of  $n$  generation methods ( $\Delta F_n$ ) that will converge to  $\Delta F$ , however, the evidence is that convergence is slow and not obviously predictable from  $\Delta F_1$  and  $\Delta F_2$ . Furthermore, the parameters required for the estimation of  $\Delta F_n$  will require either extensive simulation or the use of the theory developed by Wray and Thompson (1990). It can be shown (data not presented) that if their formulae for predicting the additional selective advantage from ancestors to descendants are forced to zero for descendants born after one or two generations, then there is good agreement of the resulting estimates of rate of inbreeding with  $\Delta F_1$  or  $\Delta F_2$ , respectively. This has two consequences: firstly, it confirms the

cause of the discrepancy between estimates and true rates and, secondly, by examining the results from forcing selective advantage terms to zero after three or more generations, it confirms the slow convergence. As a result, it is unlikely that further development of one- and two-generation methods will be worthwhile.

Verrier et al. (1990) have recently presented a method to predict levels of inbreeding in each generation for populations undergoing selection. The method is an extension of the method of Burrows (1984b), but correctly accounts for sexual dimorphism. The method is recursive, using the level of inbreeding in generation  $t-1$  to predict the rate of inbreeding in generation  $t$  and incorporating genetic parameters each generation that account for the effects of selection. However, the method is still 'one-generation' in the sense that it accounts only for the selective advantage of parents to offspring. The asymptotic rate of inbreeding calculated from this method is expected to be equal to the one-generation transition matrix method using  $q_e$ .

The accuracy of these one- or two-generation methods may be expected to be better when family indices are used rather than mass selection. This is because, in the latter case, the correlation between relatives has been assumed to be entirely genetic, whereas with family indices it is in part environmental. Indeed, for low  $h^2$  the index correlations of sibs can be principally (but never entirely) of environmental origin. If it were entirely environmental, then the methods  $\Delta F_{1,LH}$  and  $\Delta F_{1,T}$  are appropriate and unbiased. Thus, there is an a priori case for the bias in estimating additional inbreeding due to selection to be less when using one- and two-generation methods with family indices. Preliminary simulations (N. R. Wray, unpublished results) support this.

An important consideration in the use of the methods of this paper is their flexibility in enabling the modelling of population structure, including overlapping generations. Thus, providing the methods give a reliable 'measure' of  $\Delta F$  so that the ranking of alternative breeding schemes can be evaluated (as they do in the simulations presented), the methods will have a role in the planning of breeding schemes.

In conclusion, this paper has reviewed and improved various methods for estimating the rate of inbreeding under selection. Whilst these methods are biased and thus cannot have the reliability of the recursion of Wray and Thompson (1990), they are relatively easy to apply and are likely to give the correct ranking of possible breeding programmes. Therefore, it is valuable to demonstrate the bias with which they predict the asymptotic rate of inbreeding.

*Acknowledgements.* We would like to thank Prof. W. G. Hill for his critical reading of the manuscript of this paper and for a long series of discussions on this subject.



## Appendix I

*Eigenvalues and eigenvectors of gene transition matrices*

Let  $C(\lambda)$  be the characteristic equation of a matrix  $A$  of the form given in Eq. 6,

$$C(\lambda) = \lambda^4 + c_3 \lambda^3 + c_2 \lambda^2 + c_1 \lambda + c_0,$$

where

$$c_0 = \frac{1}{4} (r_1 r_4 + r_2 r_5 + r_3 r_6 - r_1 r_6 - r_2 r_3 - r_4 r_5)$$

$$c_1 = \frac{1}{2} (r_2 r_5 - r_1 r_6) + \frac{1}{4} (r_1 + r_6 - r_3 - r_4)$$

$$c_2 = \left( r_1 r_6 - r_2 r_5 + \frac{1}{2} r_1 + \frac{1}{2} r_6 - \frac{1}{4} \right)$$

$$c_3 = -\left( \frac{1}{2} + r_1 + r_6 \right).$$

Various properties of  $\lambda_{\max}$  can be shown. Firstly,  $C(1) \geq 0$  with equality if and only if  $r_i = 1/4$ , for  $i = 1 \dots 6$ ; furthermore, since  $\frac{\partial^j c}{\partial \lambda^j} > 0$  for  $\lambda = 1$ , for  $j = 1 \dots 4$ , then  $C(\lambda) > 0$  for  $\lambda > 1$ ; therefore,  $\lambda_{\max} \leq 1$  should it exist. Secondly, let  $k = \frac{1}{4}(1 + \sqrt{5})$ .  $C(k) \leq 0$  with equality if and only if  $r_3 = r_4 = 0$  ( $k$  is the  $\lambda_{\max}$  for full-sib mating). Thus, since  $C(1) \geq 0$  and  $C(k) \leq 0$ ,  $C(\lambda)$  has at least one root in the interval  $k < \lambda \leq 1$ ; therefore,  $\lambda_{\max}$  exists such that  $k \leq \lambda_{\max} \leq 1$ .

Finally, it needs to be shown that  $P_{mf}(t)$  is itself of  $O(\lambda_{\max})$ . Three cases can occur:

(i) if  $r_3, r_4 > 0$ , then  $P_{mf}(t)$  is a linear combination with positive coefficients of the other terms and, since one of these must be  $O(\lambda_{\max})$ , so is  $P_{mf}(t)$ ;

(ii) if  $r_3 = 0, r_4 \geq 0$ , then all individuals are from a common sire and  $r_1 = r_5 = 0$ , and the result immediately follows as in (i), unless  $P_{mm}(t)$  alone is  $O(\lambda_{\max})$ ; but this leads to a contradiction, since  $P_{mm}(t)$  is a positive linear combination of the others;

(iii) if  $r_4 = 0, r_3 \geq 0$ , then  $r_2 = r_6 = 0$  and the result follows from (ii).

This then establishes that  $P_{mf}(t)$  is  $O(\lambda_{\max})$  and  $2/(3 - \sqrt{5}) \leq N_e$ . Equality occurs if  $r_3 = r_4 = 0$ , which is a standard result for full-sib mating. The population size is infinite (it cannot be otherwise) when all selected individuals are unrelated ( $r_i = 1/4$  for  $i = 1 \dots 6$ ) in all generations.

## Appendix II

*Derivation of the two-generation variance of family size method*

The variance of change in gene frequency over two generations,  $V(\delta_{q(2)})$  can be expressed as the variance of the mean change in gene frequency from grandparents of either sex to grandoffspring of either sex ( $\delta q_{xy(2)}$ ),

$$V(\delta_{q(2)}) = E \left[ \frac{1}{4} (\delta q_{mm(2)} + \delta q_{mf(2)} + \delta q_{fm(2)} + \delta q_{ff(2)}) \right]^2.$$

Expanding this and omitting terms like  $E[\delta q_{mx(2)}][\delta q_{fx(2)}]$  which are zero, it follows that

$$V(\delta_{q(2)}) = \frac{1}{16} \{ E[\delta_{qm(2)}^2] + E[\delta_{mf(2)}^2] + 2 E[\delta q_{mm(2)}][\delta q_{mf(2)}] \} + \frac{1}{16} \{ E[\delta_{ff(2)}^2] + E[\delta_{fm(2)}^2] + 2 E[\delta q_{ff(2)}][\delta q_{fm(2)}] \}. \quad (A1)$$

The variance of change in gene frequency is attributable to three types of sampling: that due to sampling of genes between

grandparents (VGP), that due to sampling of genes within heterozygous grandparents (VHGP) and that due to sampling of genes within heterozygous parents (VHP). Heterozygous parents are generated by the random union of genes from either homozygous or heterozygous grandparents, and so VHP is independent of the other terms. Each term in Eq. A1 has components due to VGP, VHGP and VHP.

The VGP component of each term in Eq. A1 is analogous to the second equation on page 501 of Latter (1959) and with  $\sigma_q^2 = q(1-q)/2$ , it follows that

$$E[\delta_{qxy(2)}^2] = \frac{q(1-q)}{2} \frac{\sigma_{xy(2)}^2}{X \mu_{xy(2)}^2},$$

and, similarly,

$$E[\delta q_{xm(2)}][\delta q_{xf(2)}] = \frac{q(1-q)}{2} \frac{\sigma_{xm,xf(2)}}{X \mu_{xm(2)} \mu_{xf(2)}}.$$

It follows that

$$\text{VGP} = \frac{q(1-q)}{2} \frac{1}{16M} \left[ \frac{\sigma_{mm(2)}^2}{\mu_{mm(2)}^2} + 2 \frac{\sigma_{mm,mf(2)}}{\mu_{mm(2)} \mu_{mf(2)}} + \frac{\sigma_{mf(2)}^2}{\mu_{mf(2)}^2} \right] + \frac{q(1-q)}{2} \frac{1}{16F} \left[ \frac{\sigma_{ff(2)}^2}{\mu_{ff(2)}^2} + 2 \frac{\sigma_{fm,ff(2)}}{\mu_{fm(2)} \mu_{ff(2)}} + \frac{\sigma_{fm(2)}^2}{\mu_{fm(2)}^2} \right].$$

$\mu_{xy(2)}$  can be written as the sum of the mean number of grandoffspring of sex  $y$  from grandparents of sex  $x$  via male offspring ( $\mu_{xmy}$ ) and the equivalent via female offspring ( $\mu_{xfy}$ ), and is the same whether or not selection is taking place.

$$\mu_{xy(2)} = \mu_{xmy} + \mu_{xfy} = \mu_{xm(1)} \mu_{my(1)} + \mu_{xf(1)} \mu_{fy(1)} = \frac{M}{X} \frac{Y}{M} + \frac{F}{X} \frac{Y}{M} = 2 \frac{Y}{X}.$$

It follows that  $\mu_{xy(2)}^2 = 4 \left( \frac{Y}{X} \right)^2$  and

$$\text{VGP} = \frac{q(1-q)}{128M} \left[ \sigma_{mm(2)}^2 + 2 \left( \frac{M}{F} \right) \sigma_{mm,mf(2)} + \left( \frac{M}{F} \right)^2 \sigma_{mf(2)}^2 \right] + \frac{q(1-q)}{128F} \left[ \sigma_{ff(2)}^2 + 2 \left( \frac{F}{M} \right) \sigma_{fm,ff(2)} + \left( \frac{F}{M} \right)^2 \sigma_{fm(2)}^2 \right].$$

Next, consider the sampling of genes within heterozygous grandparents (VHGP). The proportion of heterozygous grandparents is expected to be  $2q(1-q)$  and their sampling variance is  $1/8 : 1/4$  from the binomial variance of the sampling of two genes multiplied by  $1/2$ , since only half the grandoffspring of a given grandparent are expected to receive a gene from that grandparent. The genes from heterozygous grandparents are sampled to form offspring that are homozygous or heterozygous. Each offspring has equal probability of receiving either of the two genes from the grandparent and, since they are random events, there is no sampling covariance between offspring. The sex of the grandparent and the sex of the grandoffspring are therefore irrelevant to the sampling variance. The offspring genes are then sampled as usual to generate the grandoffspring. Each grandoffspring from an offspring family either receives a gene from the grandparent in question or it does not; all those grandoffspring from a given offspring that receive a gene from the grandparent receive the same gene. Therefore, there are  $n_{i(xy)}$  grandoffspring of sex  $y$  from offspring  $i$  of sex  $x$ . The sampling covariance summing over all offspring of sex  $x$  is  $\sum_{i=1}^X n_{i(xy)}^2$ , which must be averaged over all possible covariances  $\left( \sum_{i=1}^X n_{i(xy)} \right)^2$ . The

VHGP component of Eq. A 1 can then be written as:

$$\text{VHGP} = \frac{q(1-q)}{4} \frac{1}{16} \left[ \frac{\sum_{i=1}^M n_{i(mm)}^2}{\left(\sum_{i=1}^M n_{i(mm)}\right)^2} + 2 \frac{\sum_{i=1}^M n_{i(mm)} n_{i(mf)}}{\left(\sum_{i=1}^M n_{i(mm)}\right) \left(\sum_{i=1}^M n_{i(mf)}\right)} + \frac{\sum_{i=1}^M n_{i(mf)}^2}{\left(\sum_{i=1}^M n_{i(mf)}\right)^2} \right] + \frac{q(1-q)}{4} \frac{1}{16} \left[ \frac{\sum_{i=1}^F n_{i(ff)}^2}{\left(\sum_{i=1}^F n_{i(ff)}\right)^2} + 2 \frac{\sum_{i=1}^F n_{i(fm)} n_{i(ff)}}{\left(\sum_{i=1}^F n_{i(fm)}\right) \left(\sum_{i=1}^F n_{i(ff)}\right)} + \frac{\sum_{i=1}^F n_{i(fm)}^2}{\left(\sum_{i=1}^F n_{i(fm)}\right)^2} \right]$$

Notice that the terms in Eq. A 1 relate sex of grandparent to sex of grandoffspring, but terms in Eq. A 2 relate sex of offspring to sex of grandoffspring. Therefore, the terms in the two equations are not directly analogous. Each term in Eq. A 1 has components in several terms in Eq. A 2. The terms in Eq. A 2 are found to be directly dependent on one generation means and variances of family size,

$$\frac{\sum_{i=1}^X n_{i(xy)}^2}{\left(\sum_{i=1}^X n_{i(xy)}\right)^2} = \frac{X[\sigma_{xy(1)}^2 + \mu_{xy(1)}^2]}{X^2 \mu_{xy(1)}^2} = \left[ \frac{1}{X} \left( \frac{X}{Y} \right)^2 \sigma_{xy(1)}^2 + \frac{1}{X} \right]$$

and, similarly,

$$\frac{\sum_{i=1}^X n_{i(xy)} n_{i(xf)}}{\left(\sum_{i=1}^X n_{i(xm)}\right) \left(\sum_{i=1}^X n_{i(xf)}\right)} = \left[ \frac{1}{X} \frac{X}{F} \frac{X}{M} + \frac{1}{X} \right].$$

The  $1/X$  terms are due to the sampling of heterozygotic offspring. It follows that

$$\text{VHGP} = \frac{q(1-q)}{4} \frac{1}{16M} \left[ 4 + \sigma_{mm(1)}^2 + 2 \left( \frac{M}{F} \right) \sigma_{mm, mf(1)} + \left( \frac{M}{F} \right)^2 \sigma_{mf(1)}^2 \right] + \frac{q(1-q)}{4} \frac{1}{16F} \left[ 4 + \sigma_{ff(1)}^2 + 2 \left( \frac{F}{M} \right) \sigma_{fm, ff(1)} + \left( \frac{F}{M} \right)^2 \sigma_{fm(1)}^2 \right].$$

Finally, sampling of heterozygotic parents formed from the random union of genes from grandparents is simply the case of

one generation sampling of heterozygotic parents as, before,

$$\text{VHP} = \frac{2q(1-q)}{4} \frac{1}{16M} 2 + \frac{2q(1-q)}{4} \frac{1}{16F} 2 = \frac{q(1-q)}{16M} + \frac{q(1-q)}{16F}.$$

It follows that since  $V(\delta_{q(2)}) = \text{VGP} + \text{VHGP} + \text{VHP}$  and  $V(\delta_{q(2)}) = q(1-q)/N_e$ , and  $\Delta F = 1/(2N_e)$  (discrete generations  $L=1$ ), Eq. 8 results. By analogous arguments to Hill (1979), this derivation can be shown to hold for the asymptotic rate of inbreeding in populations with overlapping generations and, hence, each term in Eq. 8 is divided by  $L$ .

## References

- Bulmer MG (1971) The effect of selection on genetic variability. *Amer Nat* 105:201–211
- Burrows PM (1984a) Inbreeding under selection from unrelated families. *Biometrics* 40:357–366
- Burrows PM (1984b) Inbreeding under selection from related families. *Biometrics* 40:895–906
- Crow JF, Kimura M (1971) An introduction to population genetic theory. Harper and Row, New York, Evanston and London
- Ewens WJ (1982) On the concept of effective population size. *Theor Popul Biol* 21:373–378
- Gowe RSA, Robertson A, Latter BDH (1959) Environment and poultry breeding problems. 5. The design of poultry control strains. *Poult Sci* 38:462–471
- Hill WG (1972) Effective population size of populations with overlapping generations. *Theor Popul Biol* 3:278–289
- Hill WG (1979) A note on effective population size of populations with overlapping generations. *Genetics* 92:317–322
- Hill WG (1985) Fixation probabilities of mutant genes with artificial selection. *Genét Sél Evol* 17:351–358
- Latter BDH (1959) Genetic sampling in a random mating population of constant size and sex ratio. *Aust J Biol Sci* 12:500–505
- Robertson A (1961) Inbreeding in artificial selection programmes. *Genet Res Camb* 2:189–194
- Verrier E (1989) Prédiction de l'évolution de la variance génétique dans les populations animales d'effectif limité soumises à sélection. PhD thesis, INA-Paris-Grignon, France
- Verrier E, Colleau JJ, Foulley JL (1990) Predicting cumulated response to directional selection in finite panmictic populations. *Theor Appl Genet* 79:833–840
- Woolliams JA (1989) Modifications to MOET nucleus breeding schemes to improve rates of genetic progress and decrease rates of inbreeding in dairy cattle. *Anim Prod* 49:1–14
- Wray NR, Thompson R (1990) Prediction of rates of inbreeding in selected populations. *Genet Res Camb* 55:41–54
- Wright S (1931) Evolution in Mendelian populations. *Genetics* 16:97–159