

## On the Effective Size of Populations With Separate Sexes, With Particular Reference to Sex-Linked Genes

Armando Caballero

*Institute of Cell, Animal and Population Biology, University of Edinburgh, Edinburgh EH9 3JT, Scotland*

Manuscript received March 31, 1994

Accepted for publication October 8, 1994

### ABSTRACT

Inconsistencies between equations for the effective population size of populations with separate sexes obtained by two different approaches are explained. One approach, which is the most common in the literature, is based on the assumption that the sex of the progeny cannot be identified. The second approach incorporates identification of the sexes of both parents and offspring. The approaches lead to identical expressions for effective size under some situations, such as Poisson distributions of offspring numbers. In general, however, the first approach gives incorrect answers, which become particularly severe for sex-linked genes, because then only numbers of daughters of males are relevant. Predictions of the effective size for sex-linked genes are illustrated for different systems of mating.

**E**FFECTIVE population size (WRIGHT 1931) is a key parameter in population and quantitative genetics. Expressions for the effective population size for autosomal and X-linked genes in species with separate sexes have been derived using inbreeding or variance of drift approaches [see CABALLERO (1994) for a review]. Some of these expressions are general because they include variances and covariances of the number of offspring considering the four possible pathways from male or female parents to male or female offspring. The most common expressions in the literature, however, are derived on the assumption that the sex of the progeny cannot be identified and only specify variances of total number of progeny, irrespective of sex. For autosomal genes these equations agree when there are equal numbers of males and females and/or a Poisson distribution of offspring numbers but not in general. For X-linked genes the implications are more severe because males can pass genes only to female offspring, so there are inconsistencies in the literature (NAGYLAKI 1981; POLLAK 1990). In this paper I point out the reasons for these inconsistencies, first aiming to summarize and clarify the situations under which different equations apply for autosomal loci before addressing X-linked loci. In all the results presented I assume sufficiently large population sizes ( $N$ ) that terms of order  $1/N^2$  can be neglected relative to  $1/N$ . Discrete generations with constant numbers of males and females and random mating are also assumed, so that inbreeding and variance effective sizes are the same because both inbreeding and variance of gene frequencies can be used to predict the increase in homozygosity. Sources of variation in family size are due to noninherited causes.

### AUTOSOMAL LOCI

A general equation for the effective size ( $N_e$ ) of populations with separate sexes regarding autosomal genes is given by HILL (1979), following the methodology of LATTER (1959),

$$\frac{1}{N_e} = \frac{1}{16M} \left[ 2 + \sigma_{mm}^2 + 2 \left( \frac{M}{F} \right) \sigma_{mm,mf} + \left( \frac{M}{F} \right)^2 \sigma_{mf}^2 \right] + \frac{1}{16F} \left[ 2 + \sigma_{ff}^2 + 2 \left( \frac{F}{M} \right) \sigma_{fm,ff} + \left( \frac{F}{M} \right)^2 \sigma_{fm}^2 \right], \quad (1)$$

where  $M(F)$  is the number of male (female) parents in the population,  $\sigma_{sm}^2$  ( $\sigma_{sf}^2$ ) is the variance of the number of male (female) offspring (which will reach reproductive age) from parents of sex  $s$  ( $m$  or  $f$ ) and  $\sigma_{sm,sf}$  is the covariance of the number of male and female offspring from parents of sex  $s$ . Equation 1 was derived by using drift variance arguments. An alternative derivation following inbreeding arguments can be straightforwardly obtained from CABALLERO and HILL (1992), where the derivation was simplified to the case of  $M = F$ . With another derivation following drift variance arguments, CROW and DENNISTON (1988) obtained an expression for the effective size with separate sexes (Equations 26 and 28 of their paper) that, for large population size and random mating (when deviations from Hardy-Weinberg proportions are negligible), is identical to Equation 1.

When numbers of male and female parents are equal ( $M = F = N/2$ ), (1) reduces to

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

[WRIGHT (1938), assuming large  $N$ ], where  $\sigma_k^2$  is the variance of family size (the variance of the number of successful gametes, those that will produce zygotes, from each parent). For Poisson distribution of offspring numbers, where  $\sigma_{sm}^2 (\sigma_{sf}^2)$  equals the mean number of male (female) offspring from parents of sex  $s$ , *i.e.*,  $\mu_{sm} = M/S$  ( $\mu_{sf} = F/S$ ), where  $S$  is  $M$  or  $F$  and  $\sigma_{sm,sf} = 0$ , (1) reduces to  $N_e = 4MF/(M+F)$  (WRIGHT 1931), and (2) gives  $N_e = N$ , as expected.

Following an inbreeding approach, KIMURA and CROW (1963) and CROW and DENNISTON (1988) derived an alternative equation to (1) for the effective size for populations with separate sexes. For large census sizes this approximates (2), with  $N = M + F$  and  $\sigma_k^2 = \sigma_m^2/\mu_m + \sigma_f^2/\mu_f + (\mu_m - \mu_f)^2/\mu_m\mu_f$ , where  $\mu_s = N/S$  ( $S$  is  $M$  or  $F$ ) and  $\sigma_s^2 = \sigma_{sm}^2 + \sigma_{sf}^2 + 2\sigma_{sm,sf}$  are the mean and variance of the total number of offspring from parents of sex  $s$ . This is also identical, on rearrangement, to

$$N_e = \frac{4N^2}{M(\sigma_m^2 + \mu_m^2 - \mu_m) + F(\sigma_f^2 + \mu_f^2 - \mu_f)}, \quad (3)$$

obtained by MALECOT (1951), MORAN and WATTERSON (1959) and POLLAK (1977, except for  $4N[N-1]$  in the numerator).

Equations 2 and 3 assume that the probability that two individuals in generation  $t$  have common ancestry in generation  $t-1$  is independent of their sex, which might be a reasonable assumption when the sex of the offspring cannot be identified (CROW and DENNISTON 1988), perhaps for some practical reason. Thus, Equation 3 can be rearranged to read  $1/N_e = (1/\mu_f + \sigma_m^2/\mu_m^2)/4M + (1/\mu_m + \sigma_f^2/\mu_f^2)/4F$ , showing that total variances of offspring number for each sex ( $\sigma_s^2$ ) are weighted inversely by their mean squares ( $\mu_s^2$ ). However, variances in brackets in (1) can be expressed as  $\sigma_{sm}^2/\mu_{sm}^2 + 2\sigma_{sm,sf}/\mu_{sm}\mu_{sf} + \sigma_{sf}^2/\mu_{sf}^2$ , *i.e.*, variances for each of the pathways of genes are weighted inversely by the corresponding mean squares. This different weighting of the variances implies that (1) and (2 or 3) agree when  $M = F$  and/or when the distribution of offspring numbers is Poisson and there is no covariance between male and female offspring numbers, but not in general. If there are differences in fertility or viability, for example, even if these are not inherited, a covariance term arises, and (1) and (2 or 3) give different results. Therefore, if the sex of the offspring can be identified, Equation 1 should be used instead of (2 or 3), because the variances and covariances of offspring numbers are properly weighted. In all cases Equations 2 and 3 seem to apply to a peculiar and unrealistic situation in which the sex of the parents in a given generation can be determined but not that of the parents in the next. A possibility is that progeny numbers are available only at an earlier stage (*e.g.*, eggs) in which the sex cannot be identified. This is, however, a differ-

ent problem that requires conversion of the measurements at the earlier stage into those at the adult stage, involving assumptions about the model of survival (CROW and MORTON 1955).

#### X-LINKED LOCI

The effective size for X-linked loci or haplo-diploid species is obtained by equating the variance of change in gene frequency or the rate of inbreeding of an autosomal gene in an idealized population to the actual values observed for the X-linked gene in the population under consideration. Following a variance of drift derivation similar to that of HILL (1979), POLLAK (1990) obtained an expression for  $N_e$  for X-linked genes,

$$\frac{1}{N_e} = \frac{1}{9M} \left[ 1 + 2 \left( \frac{M}{F} \right)^2 \sigma_{mf}^2 \right] + \frac{1}{9F} \left[ 1 + \sigma_{ff}^2 + 2 \left( \frac{F}{M} \right) \sigma_{f,m,ff} + \left( \frac{F}{M} \right)^2 \sigma_{jm}^2 \right] \quad (4)$$

(cf. Equation 1), where the male sex is heterogametic. In the case of equal number of male and female parents ( $M = F = N/2$ ), (4) reduces to

$$N_e = \frac{9N}{4 + 4\sigma_{mf}^2 + 2\sigma_j^2}. \quad (5)$$

For Poisson-distributed offspring numbers, (4) leads to  $N_e = 9MF/(4M + 2F)$  (WRIGHT 1933) and (5) to  $N_e = 3N/4$ .

Following an inbreeding approach NAGYLAKI (1981) arrived at an expression that, for large  $N$ , is approximately

$$N_e = \frac{9N^2}{2M(\sigma_m^2 + \mu_m^2 - \mu_m) + 4F(\sigma_f^2 + \mu_f^2 - \mu_f)} \quad (6)$$

(cf. Equation 3) (more exactly, Nagylaki's expression has  $9N[N-1]$  in the numerator). Expression 6 was also obtained by MORAN and WATTERSON (1959) and ETHIER and NAGYLAKI (1980).

With equal number of male and female parents ( $M = F = N/2$ ),  $\mu_m = \mu_f = 2$  and  $\sigma_m^2 = \sigma_f^2 = \sigma_k^2$ , (6) leads to

$$N_e = \frac{3N}{2 + \sigma_k^2}. \quad (7)$$

For Poisson-distributed offspring numbers, (6 and 7) are in agreement with (4 and 5). However, they do not agree under some circumstances even for equal numbers of male and female parents. The difference can be illustrated under this latter situation when  $M = F$ . Results given by Equations 5 and 7 differ most in the extreme case of  $\sigma_{mf}^2 = \sigma_j^2 = \sigma_k^2 = 0$  (equal numbers of successful gametes produced per parent), for which

(5) gives  $N_e = 9N/4$ , whereas (7) gives  $N_e = 3N/2$ . The reason for this discrepancy is, again, that (6) is based on the assumption that the sex of the progeny cannot be identified. Thus, Equation 6 specifies variances of total offspring from males, whereas only numbers of daughters are relevant for X-linked loci. In what follows, a derivation of  $N_e$  following an inbreeding approach similar to that of CABALLERO and HILL (1992) for autosomal loci is made, which leads to the same result obtained by the variance of drift approach (Equation 4).

Let  $x_t$  ( $z_t$ ) be the probability that two genes from two different males (females) in generation  $t$  are identical by descent and let  $y_t$  be the same probability for two genes, one each from a male and a female in generation  $t$ . With random mating the inbreeding coefficient in females is  $F_t = y_{t-1}$ . In generation  $t + 1$  the coefficients are obtained by adding up the probabilities that the pair of genes are sampled from the same male, from different males, from the same female, from different females and from a male and a female where appropriate. For instance,

$$x_{t+1} = \left[ \frac{\sum_i^F k_{fm_i} (k_{fm_i} - 1)}{M(M - 1)} \right] \left( \frac{1}{2} + \frac{F_t}{2} \right) + \left[ 1 - \frac{\sum_i^F k_{fm_i} (k_{fm_i} - 1)}{M(M - 1)} \right] z_t,$$

where  $k_{fm_i}$  is the number of male offspring from the  $i$ th female parent. Noting that  $\sum_i^F k_{fm_i} (k_{fm_i} - 1) / [M(M - 1)] = [(F/M) \sigma_{fm}^2 + (M/F) - 1] / (M - 1)$ , it follows that

$$x_{t+1} = \frac{1}{(M - 1)} \left[ \left( \frac{F}{M} \sigma_{fm}^2 + \frac{M}{F} - 1 \right) \times \left( \frac{1}{2} + \frac{F_t}{2} - z_t \right) \right] + z_t. \tag{8}$$

Analogously,

$$z_{t+1} = \frac{1}{4(F - 1)} \left[ \left( \frac{M}{F} \sigma_{mf}^2 + \frac{F}{M} - 1 \right) (1 - x_t) + \sigma_{mf}^2 \left( \frac{1}{2} + \frac{F_t}{2} - z_t \right) \right] + \frac{x_t + 2y_t + z_t}{4} \tag{9}$$

and

$$y_{t+1} = \frac{1}{2} \left( \frac{\sigma_{f_m f_f}}{M} + \frac{1}{F} \right) \left( \frac{1}{2} + \frac{F_t}{2} - z_t \right) + \frac{1}{2} (z_t + y_t). \tag{10}$$

Assume, for simplicity, that  $M$  and  $F$  are large enough that  $M \approx M - 1$  and  $F \approx F - 1$ . Denote by  $f_t = (x_t + 4z_t + 4y_t) / 9$  the average probability of identity in generation  $t$ , because  $1/3$  of the X-linked genes are expected to come from males and  $2/3$  from females. Note also that the vector  $(1/9, 4/9, 4/9)$  is the eigenvector corresponding to the unit eigenvalue (other eigenvalues being smaller than one) of the matrix that gives the values of  $x_{t+1}$ ,  $y_{t+1}$  and  $z_{t+1}$  as a function of  $x_t$ ,  $y_t$  and  $z_t$  in Equations 8–10, *i.e.*, when the terms in  $1/M$  and  $1/F$  in the right hand sides of the equations are neglected.

Denoting

$$C_m = \frac{1}{9M} \left[ 1 + 2 \left( \frac{M}{F} \right)^2 \sigma_{mf}^2 \right] \tag{11}$$

and

$$C_f = \frac{1}{9F} \left[ 1 + \sigma_{ff}^2 + 2 \left( \frac{F}{M} \right) \sigma_{f_m f_f} + \left( \frac{F}{M} \right)^2 \sigma_{fm}^2 \right] \tag{12}$$

and substituting (8–10) into  $f_{t+1}$ , we obtain

$$f_{t+1} = \frac{1}{2} (C_m + C_f) (1 - f_t) + \frac{1}{2} C_m (f_t - x_t) + \frac{1}{2} C_f (F_t - 2z_t + f_t) + f_t. \tag{13}$$

By manipulating Equations 8–10 it can be seen that absolute differences among  $x_t$ ,  $y_t = F_{t+1}$ ,  $z_t$  and  $f_t$  are of order  $1/M$  or  $1/F$  (say  $1/N$ ) in the long run. As  $C_m$  and  $C_f$  are also of order  $1/N$ , terms  $C_m(f_t - x_t)$  and  $C_f(F_t - 2z_t + f_t)$  in (13) are of order  $1/N^2$  and can be neglected for large population sizes. Hence, noting that  $1/2N_e \approx (f_{t+1} - f_t) / (1 - f_t)$ , from (13) we get  $1/N_e \approx C_m + C_f$ , which using (11–12) gives Equation 4.

### DISCUSSION

Two different approaches have been used in the literature to predict effective population sizes for populations with separate sexes. The most common is based on the assumption that the sex of the progeny cannot be identified. A more general approach incorporates identification of the sexes of both parents and offspring. These approaches lead to the same results under some restrictive situations, but, in general, only the second gives correct answers. For sex-linked genes this effect is more severe because only numbers of daughters of males are relevant. NAGYLAKI (1981) derived an equation for the effective population size for X-linked genes under the first approach. He used the probabilities that two distinct, homologous, randomly chosen paternal ( $P_m$ ) or maternal ( $P_f$ ) genes in generation  $t$  came from the same individual in generation  $t - 1$  and obtained

$$N_e = \frac{9}{2P_m + 4P_f} \quad (14)$$

(his Equation 5). He treated these probabilities similarly, such that  $P_m$  and  $P_f$  are given approximately by  $P_s \approx S(\sigma_s^2 + \mu_s^2 - \mu_s) / N^2$  (from his Equation 8), where  $S$  is  $M$  or  $F$  and  $s$  is  $m$  or  $f$ . However, although female parents can pass gametes carrying  $X$ -linked genes to males, females or one of each, male parents can pass them only to females. If these differences in contributions from males and females are considered, we can write

$$P_m = \frac{\sum_i^M k_{mfi}(k_{mfi} - 1)}{F(F - 1)} \quad (15)$$

and

$$P_f = \frac{1}{4} \left( \frac{\sum_i^F k_{fmi}(k_{fmi} - 1)}{M(M - 1)} \right) + \frac{1}{2} \left( \frac{\sum_i^F k_{fmi}k_{ffi}}{MF} \right) + \frac{1}{4} \left( \frac{\sum_i^F k_{ffi}(k_{ffi} - 1)}{F(F - 1)} \right). \quad (16)$$

Proceeding as before,

$$P_m \approx \frac{1}{M} \left[ 1 + \left( \frac{M}{F} \right)^2 \sigma_{mf}^2 - \frac{M}{F} \right] \quad (17)$$

and

$$P_f \approx \frac{1}{4F} \left[ 3 + \sigma_{ff}^2 + 2 \left( \frac{F}{M} \right) \sigma_{f mff} \right] + \left( \frac{F}{M} \right)^2 \sigma_{fm}^2 - \frac{F}{M}. \quad (18)$$

Substituting (17–18) into (14), we obtain again (4).

Note also that Equations 8–10 are a corrected version of Equations 1b–1d of NAGYLAKI (1981), where  $k$ ,  $g$ ,  $l$  and  $f$ , in his notation, correspond to  $x$ ,  $y$ ,  $z$  and  $F_t$ , respectively. The difference between these two systems of equations is that there should not be a single  $P_2$  that appears in (1b–1d), but three different probabilities:  $P_{2,11}$  in (1c),  $P_{2,12}$  in (1b) and  $P_{2,22}$  in (1d), which correspond to the three terms in parenthesis in (16), respectively. Finally,  $P_1$  in (1d) corresponds to  $P_m$  from (15).

There are two interesting points about control populations with minimal inbreeding. In the case where each male parent has one male and  $F/M$  female offspring, and each female parent has one female and zero or one male offspring ( $M/F$ , on average),  $\sigma_{fm}^2 = (M/F)(1 - M/F)$ ,  $\sigma_{mf}^2 = \sigma_{ff}^2 = \sigma_{f mff} = 0$  and (4) yields  $N_e = 9M/2$ , *i.e.*, independent of the number of females. For species where females are the heterogametic sex,

as in poultry, the same mating procedure gives  $N_e = 9MF/(3F - M)$ . This indicates that, for minimal inbreeding, as few females as possible for each male should be used. In fact, the largest effective size would be obtained by mating one female to each male ( $N_e = 9N/4$ ). An analogous conclusion, perhaps more important from a practical point of view, can be made for haplo-diploid species, such as bees, in which females can mate each to many males [see CROW and ROBERTS (1950) for an analysis of the inbreeding produced by various systems of mating in honey bees]. If the minimal inbreeding procedure is followed, as few males as possible should be mated to each female.

A few examples illustrate these results, which were also checked by stochastic simulation. Using a population of constant size  $N = 80$  (half of each sex) with equal family sizes (monogamous matings with one male and one female offspring per couple) run for 15 generations with 10,000 replicates, the effective size, evaluated from the average rate of decrease in heterozygosity between generations 5 and 15 of a neutral  $X$ -linked gene, was  $177.0 \pm 2.4$ , while expectations from Equations 5 and 7 are 180 and 120, respectively. With  $M$  males mated to  $F/M$  females each and minimal inbreeding as explained above, effective sizes were  $45.3 \pm 0.4$  and  $44.7 \pm 1.7$  for  $M = 10$  and  $F = 20$  and  $F = 160$ , respectively. The expectation with Equation 4 is 45 for both cases. Similar simulations when females are heterogametic gave values of  $34.9 \pm 0.7$  and  $29.9 \pm 0.1$ , whereas expectations with Equation 4 are 36 and 30.6, respectively.

I thank the Agricultural and Food Research Council for financial support. I am indebted to W. G. HILL for pointing out the results with minimal inbreeding and different number of parents of each sex, to T. NAGYLAKI for pointing out that there are also discrepancies for autosomal loci equations and to them, M. A. TORO and an anonymous referee for useful comments on the manuscript.

#### LITERATURE CITED

- CABALLERO, A., 1994 Developments in the prediction of effective population size. *Heredity* (in press).
- CABALLERO, A., and HILL, W. G., 1992 A note on the inbreeding effective population size. *Evolution* **46**: 1969–1972.
- CROW, J. F., and DENNISTON, C., 1988 Inbreeding and variance effective population numbers. *Evolution* **42**: 482–495.
- CROW, J. F., and MORTON, N. E., 1955 Measurement of gene frequency drift in small populations. *Evolution* **9**: 202–214.
- CROW, J. F., and ROBERTS, W. C., 1950 Inbreeding and homozygosity in bees. *Genetics* **35**: 612–621.
- ETHIER, S. N., and NAGYLAKI, T., 1980 Diffusion approximations of Markov chains with two time scales and applications to population genetics. *Adv. Appl. Probab.* **12**: 14–49.
- HILL, W. G., 1979 A note on effective population size with overlapping generations. *Genetics* **92**: 317–322.
- KIMURA, M., and CROW, J. F., 1963 The measurement of effective population number. *Evolution* **17**: 279–288.
- 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.
- MALECOT, G., 1951 Un traitement stochastique des problèmes linéaires (mutation, linkage, migration) en génétique de population. *Ann. Univ. Lyon Sci. Sec. A* **14**: 79–117.

- 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.
- NAGYLAKI, T., 1981 The inbreeding effective population number and the expected homozygosity for an X-linked locus. *Genetics* **97**: 731-737.
- POLLAK, E., 1977 Effective population numbers and their interrelations. Proceedings of the Washington State University Conference on Biomathematics and Biostatistics, Washington State Univ., Spokane, WA, pp. 115-144.
- POLLAK, E., 1990 The effective population size of an age-structure population with a sex-linked locus. *Math. Biosci.* **101**: 121-130.
- WRIGHT, S., 1931 Evolution in Mendelian populations. *Genetics* **16**: 97-159.
- WRIGHT, S., 1933 Inbreeding and homozygosis. *Proc. Natl. Acad. Sci. USA* **19**: 411-419.
- WRIGHT, S., 1938 Size of population and breeding structure in relation to evolution. *Science* **87**: 430-431.

Communicating editor: B. S. WEIR