

# On the methods for predicting the effective size of populations under selection

TETSURO NOMURA

*Department of Biotechnology, Faculty of Engineering, Kyoto Sangyo University, Kyoto 603-8555, Japan*

Inconsistencies between equations for the effective population size under selection obtained by two different approaches are explained. In one approach, the effective population size is predicted from the drift in the frequency of a neutral allele, accounting for the accumulation of selective advantage over generations. The second approach is based on the rate of inbreeding, using the concept of long-term genetic contributions. It is shown that the long-term genetic contribution approach leads to an identical result to the drift approach, if the effect of mates on the long-term genetic contributions is correctly accounted for.

**Keywords:** effective population size, genetic contribution, genetic drift, inbreeding, selection.

## Introduction

The effective population size is one of the most important parameters in the definition of selection programmes, because the magnitude of this parameter determines the effects of both random genetic drift and inbreeding (Falconer, 1989; Caballero, 1994). In the absence of selection, all the individuals of a population have the same expected number of offspring, and differences in family size are caused only by random sampling of individuals among families. However, in populations under selection, families do not have equal probabilities of contributing to the next generation because of inherited or noninherited causes. This leads to a reduction in the effective population size. The effective size of selected populations, however, cannot be predicted solely from the variance of the family size at a given generation, because selection affects the variance of the family size not only in the parental generation but also in the ancestral generations (Wray *et al.*, 1990).

There are two approaches to predict the effective size of populations under selection. First, the effective population size (variance effective size) can be predicted from the variance of gene frequency, using the idea of accumulation of selective advantage over generations (Robertson, 1961). Following this approach, Santiago & Caballero (1995) obtained accurate predictions for selected populations with discrete generations. The method has been extended to monoecious populations with partial selfing (Caballero & Santiago, 1996),

dioecious populations with overlapping generations (Nomura, 1996), sex-linked loci (Nomura, 1997), linked autosomal loci (Santiago & Caballero, 1998) and non-random mating (Wang, 1998). Secondly, the effective population size (inbreeding effective size) can be predicted from the rate of inbreeding, using the concept of long-term genetic contributions (Wray & Thompson, 1990). Wray & Thompson (1990) showed that the rate of inbreeding is proportional to the sum of squared long-term genetic contributions of ancestors, and derived a method to predict rates of inbreeding in populations under mass selection. Their method, however, requires complex iterative computations. In order to overcome this problem, Woolliams *et al.* (1993) proposed a method for predicting without iterations. Wray *et al.* (1994) extended the method to index selection with family information. Later, Woolliams & Thompson (1994) corrected this method for some omissions and the equation they obtained has been used to optimize breeding programmes (Villanueva *et al.*, 1996; Villanueva & Woolliams, 1997).

Although the variance and inbreeding effective sizes should be the same under constant population size over generations (Crow & Kimura, 1970; Caballero, 1994), the two approaches to predict the effective size of selected populations give different expressions. It is unclear as yet how these two approaches differ in their assumptions and approximations (Woolliams & Thompson, 1994). In this paper, the author clarifies the meaning of the difference based on the long-term genetic contributions approach. It is shown that the equation of Santiago & Caballero (1995) can also be derived from this approach, if the effect of mates on the

\*Correspondence. E-mail: nomurat@cc.kyoto-su.ac.jp

long-term genetic contributions is correctly accounted for.

**Model and assumptions**

Although the theory can be extended to a population with different numbers of male and female parents, for the purpose of this paper we assume a population in which the same numbers of male and female parents are randomly pair-mated to create  $N/2$  families (couples) each generation. Each family contributes  $n$  progeny of each sex to candidates of selection. Among the progeny of each sex, the best  $N/2$  individuals are selected on the basis of their phenotypic values. The proportion selected is  $p = 1/n$ , and the selection intensity is  $i = z/p$ , where  $z$  is the normal ordinate corresponding to the truncation point of selection ( $x$ ). The trait under selection is assumed to be determined by an infinite number of additive loci, each with infinitesimal effect (the infinitesimal model; Bulmer, 1980). The variance reduction factor due to the Bulmer effect (Bulmer, 1980) is  $k = i(i - x)$  and the coefficient of competitiveness defined by Woolliams *et al.* (1993) is  $c = (1 - kh^2)/2$ , where  $h^2$  is the heritability. The population is assumed to be at the steady-state under the infinitesimal model. Then asymptotic values are used for all genetic parameters. Constant population size and discrete generations are assumed.

The genetic contribution of an ancestor  $j$  in generation  $n_1$  to descendants in a later generation  $n_2$  is defined as the total additive relationship between the ancestor and the descendants (Wray & Thompson, 1990). The long-term contribution is the contribution when  $n_2$  is much bigger than  $n_1$ . In the absence of selection, an ancestor is expected to contribute two offspring to the next generation, each with the additive relationship  $1/2$ . The expected total contribution is therefore 1. This expectation remains constant in the later generations, and the expected long-term contribution of an ancestor is 1. Under selection, genetically superior ancestors are expected to contribute more descendants than genetically inferior ancestors. Thus, ancestors have different expected genetic contributions, depending on their selective advantage. We relate the effective population size to the long-term contribution of ancestors in a given generation [following Wray & Thompson (1990), this generation is referred to as generation 1]. The long-term contribution of ancestor  $j$  of sex  $x$  ( $=m$  or  $f$ ) to descendants in generation  $t$  is denoted by  $r_{j(x),t}$ . Note that an ancestor and its mate contribute to their descendants in any generations with exactly the same rate, because there are no half-sib families in our model. Then, the long-term contributions of male and female parents in couple  $j$  are  $r_{j,t} \equiv r_{j(m),t} = r_{j(f),t}$ .

In the works of Wray & Thompson (1990) and Woolliams *et al.* (1993), the genetic contribution of an ancestor is formulated only with its own breeding value. However, as suggested by Woolliams & Thompson (1994), the expected long-term contribution of an ancestor should be conditional not only on its breeding value but also on the breeding value of mates. Let  $A_{j(m)}$  and  $A_{j(f)}$  be the deviations of breeding values of male and female parents in couple  $j$  in generation 1 from their selected contemporaries, respectively, and  $\sigma_P$  be the phenotypic standard deviation. Defining  $A_j^* = (A_{j(m)} + A_{j(f)})/2\sigma_P$ , the expectation of long-term contributions ( $\mu_{j,t}$ ) of parents in this couple, conditional on  $A_j^*$ , is:

$$\mu_{j,t} = E[r_{j,t}|A_j^*] = 1 + iS_t \ 2A_j^*$$

and the limiting value ( $\mu_{j,\infty}$ ) is:

$$\mu_{j,\infty} = E[r_{j,\infty}|A_j^*] = 1 + iS_\infty A_j^*,$$

where  $S_t = \sum_{i=0}^t c^i$  and  $S_\infty = \sum_{i=0}^\infty c^i = (1 - c)^{-1}$  (Woolliams *et al.*, 1993). Further, a genealogical pathway descended from this couple is expected to express the relative selective advantage in the selection of generation  $t$ :

$$\mu_{j,t-1,t} = 1 + ic^t \ 2A_j^*$$

(Woolliams *et al.*, 1993).

**Comparison of equations and discussion**

Santiago & Caballero (1995) derived a prediction equation for effective size ( $N_e$ ) in terms of the variance of change in the frequency of a neutral allele. Under random mating, the equation reduces to:

$$N_{e(SC)} = \frac{N}{1 + S_\infty^2 i^2 \rho_{FS}}, \tag{1}$$

where  $\rho_{FS}$  is the intraclass correlation of full-sibs.

An alternative viewpoint for predicting the effective size is in terms of the rate of inbreeding. According to Wray & Thompson (1990), the rate of inbreeding is related to the long-term contribution as:

$$\begin{aligned} \Delta F &= \frac{1}{4N^2} \sum_{j=1}^{N/2} (r_{j(m),\infty}^2 + r_{j(f),\infty}^2) \\ &= \frac{1}{2N^2} \sum_{j=1}^{N/2} r_{j,\infty}^2 \\ &= \frac{1}{4N} E[r_{j,\infty}^2], \end{aligned} \tag{2}$$

where  $E[r_{j,\infty}^2] = \sum_{j=1}^{N/2} r_{j,\infty}^2 / (N/2)$ . Instead of the direct computation of  $E[r_{j,\infty}^2]$ , Wray *et al.* (1994) derived the expression for  $E[r_{j,2}^2]$ , and then evaluated  $E[r_{j,\infty}^2]$  with the expression. The expression for  $E[r_{j,2}^2]$  is:

$$E[r_{j,2}^2] = 1 + \frac{1}{2} + i^2 \rho_{FS} = 1 + \sigma_{r(e),2}^2 + \sigma_{r(g),2}^2, \tag{3}$$

where the constant 1 is the mean squared under random selection,  $\sigma_{r(e),2}^2 = 1/2$  is the sampling variance under random selection (see also Wray & Thompson, 1990), and  $\sigma_{r(g),2}^2 = i^2 \rho_{FS}$  is the variance attributable to selective advantage of couples. To evaluate  $E[r_{j,\infty}^2]$  with this expression, Wray *et al.* (1994) took the following aspects into account.

**1** Under random selection (i.e.  $\sigma_{r(g),2}^2 = 0$ ), the mean of the long-term contributions of the ancestors is the same for  $t=2$  as for  $t=\infty$ , and so the first term in eqn (3) remains unchanged. This holds also under selection (Woolliams *et al.*, 1993).

**2** Under random selection, the sampling variance  $\sigma_{r(e),2}^2$  is increased by a factor of 2 from  $t=2$  to  $t=\infty$  (Wray & Thompson, 1990). Thus,  $\sigma_{r(e),\infty}^2 = 2\sigma_{r(e),2}^2 = 1$ . This leads to  $\Delta F = 1/2N$  or  $N_e = N$ , as expected.

**3** The doubling of the variance terms ( $\sigma_{r(e),2}^2 + \sigma_{r(g),2}^2$ ) is insufficient for the expression of  $E[r_{j,\infty}^2]$  under selection, because of the cumulative effect of selective advantage. The accumulation of selective advantage can be incorporated by multiplying  $\sigma_{r(g),2}^2$  by  $S_\infty^2$ .

Accounting for these points results in:

$$E[r_{j,\infty}^2] = 1 + 2 \left( \sigma_{r(e),2}^2 + S_\infty^2 \sigma_{r(g),2}^2 \right) = 2 \left( 1 + S_\infty^2 \sigma_{r(g),2}^2 \right).$$

Then:

$$\Delta F = \frac{1 + S_\infty^2 i^2 \rho_{FS}}{2N}$$

or:

$$N_{e(WT)} = \frac{N}{1 + S_\infty^2 i^2 \rho_{FS}}, \tag{4}$$

which agrees with eqn (1).

In the original work of Wray *et al.* (1994),  $\sigma_{r(e),2}^2 = 1/2 + i^2 \rho_y$ , and  $\sigma_{r(g),2}^2 = i^2 \rho_x$  were used, where  $\rho_x$  and  $\rho_y$  are the correlations of full-sibs arising from the breeding values of the ancestors and their mates, respectively, i.e.  $\rho_x = \rho_y = \rho_{FS}/2$ . Applying the rule to evaluate  $E[r_{j,\infty}^2]$ , they obtained an expression:

$$\frac{1}{N_e} = \frac{1}{N} + \frac{i^2 \rho_{FS}}{2N} + \frac{S_\infty^2 i^2 \rho_{FS}}{2N}. \tag{5}$$

But this expression has an omission, because the accumulation of the selective advantage from mates is not accounted for (Woolliams & Thompson, 1994). To correct this omission, the effect of mates [the second term of eqn (5)] should be accumulated in the same process as the last term of eqn (5). This correction leads to eqn (4).

Recently, Woolliams (1998) showed that eqn (2) can be approximated by:

$$\Delta F = \frac{1}{2N} E[\mu_{j,\infty}^2].$$

Noting that  $E[A_j^*] = 0$  and  $E[A_j^{*2}] = \rho_{FS}$ ,

$$E[\mu_{j,\infty}^2] = 1 + S_\infty^2 i^2 \rho_{FS}.$$

Then, we obtain again eqn (1).

Using the iterative regression method of Wray & Thompson (1990), Woolliams *et al.* (1993) derived an alternative expression for  $E[r_{j,\infty}^2]$ ,

$$E[r_{j,\infty}^2] = E[P_{j,3}^2 \mu_{j,2}^2] + E\left[ \sum_{t=2}^{\infty} P_{j,t+1}^2 V_t \right], \tag{6}$$

where  $P_{j,t} = \prod_{l=t}^{\infty} \mu_{j,t-1,l}$  and  $V_t$  is the sampling variance of long-term genetic contribution arising in generation  $t$ . If we neglect the contributions of  $O(A_j^{*2})$  or higher following the suggestion of Woolliams & Thompson (1994), then:

$$\begin{aligned} P_{j,3} &= (1 + ciA_j^*) (1 + c^2iA_j^*) (1 + c^3iA_j^*) \dots \\ &= 1 + \sum_{t=1}^{\infty} c^t iA_j^* = 1 + (S_\infty - 1)iA_j^* \end{aligned}$$

and:

$$P_{j,3} \mu_{j,2} = (1 + iA_j^*) [1 + (S_\infty - 1)iA_j^*] = 1 + S_\infty iA_j^*.$$

Thus the first term of eqn (6) equals  $E[\mu_{j,\infty}^2] = 1 + S_\infty^2 i^2 \rho_{FS}$ . The second term of eqn (6) accounts for the accumulation of sampling variance (we denote this term by  $E[\sigma_{j,\infty}^2]$ ). An approximated expression of  $E[\sigma_{j,\infty}^2]$  may be obtained by using the terms listed in appendix 3 of Woolliams *et al.* (1993). Using the terms A3.6, A3.10 and A3.16 in Woolliams *et al.* (1993),

$$E[\sigma_{j,\infty}^2] = 1 + [1 + B_\infty (S_\infty - 1)] i^2 \rho_{FS}, \tag{7}$$

where  $B_\infty = \sum_{j=0}^{\infty} (c^2/2)^j = 2/(2-c^2)$ . From this approximation, the prediction equation of Woolliams & Thompson (1994),

$$N_{e(\text{WT})} = \frac{N}{1 + \frac{1}{2}[1 + S_\infty^2 + B_\infty(S_\infty - 1)]i^2\rho_{\text{FS}}} \quad (8)$$

is obtained. If we further include term A3.7 in Woolliams *et al.* (1993), eqn (7) becomes:

$$E[\sigma_{j,\infty}^2] = 1 + [1 + B_\infty(S_\infty - 1) + (B_\infty - 1)S_\infty^2]i^2\rho_{\text{FS}}, \quad (9)$$

leading to:

$$N_{e(\text{WT})} = \frac{N}{1 + \frac{1}{2}[1 + S_\infty^2 + B_\infty(S_\infty - 1) + (B_\infty - 1)S_\infty^2]i^2\rho_{\text{FS}}}, \quad (10)$$

which is the simplified expression of the equation used by Villanueva & Woolliams (1997). [In these derivations,  $\rho_x$  (the correlation of full-sibs arising from the breeding value of the parent of sex  $x$ ) in A3.7 and A3.10 should be replaced by  $\rho_{\text{FS}}$ , because the cumulative effect of the selective advantage of the mates is omitted in the work of Woolliams *et al.* (1993). See also appendix of Woolliams & Thompson (1994).]

In the case of the iterative regression method [eqn (6)], eqn (4) is derived by assuming:

$$E[\mu_{j,\infty}^2] = E[\sigma_{j,\infty}^2] = 1 + S_\infty^2 i^2 \rho_{\text{FS}}.$$

The equality of  $E[\mu_{j,\infty}^2] = 1 + S_\infty^2 i^2 \rho_{\text{FS}}$  was given in this paper, and the equality of  $E[\mu_{j,\infty}^2] = E[\sigma_{j,\infty}^2]$  is obtained by assuming that the litter size of each family, i.e. the number of scored individuals in each family, follows a Poisson distribution (Woolliams, 1998; J. A. Woolliams & P. Bijma, pers. comm.). As seen from eqns (8) and (9) of Santiago & Caballero (1995), the assumption of a Poisson distribution of litter size has a very small effect on the prediction of  $N_e$  (the effect on the prediction is of the order of  $1/T$ , where  $T$  is the total number of scored individuals). In fact, eqn (1) is the approximation obtained from a Poisson distribution of litter size (Santiago & Caballero, 1995). On the other hand, in the derivation of eqns (7) and (9), the terms resulting from an approximated expansion of  $E[\sigma_{j,\infty}^2]$  were listed (Woolliams *et al.*, 1993), and only terms with a larger effect on  $N_e$  were included. All the terms of the order of  $1/T$  were also neglected in the derivation. Thus, eqns (8) and (10) may be regarded as approximations to eqn (1). Although simulation studies have shown that the approximated eqns (8) and (10) give satisfactory

predictions (Woolliams & Thompson, 1994; Villanueva & Woolliams, 1997), eqn (1) is preferable to these equations because of the neater expression involving less approximation.

In conclusion, an identical equation is obtained from the drift approach (Santiago & Caballero, 1995) and the long-term contribution approach (Wray & Thompson, 1990), as seen from eqns (1) and (4). The two approaches have, however, their respective appealing points. The long-term genetic contribution approach provides a prediction of genetic gain (Woolliams & Thompson, 1994; Woolliams, 1998), and gives an insight into the design of breeding schemes to maximize gain with constraints of inbreeding (Woolliams & Thompson, 1994; Villanueva & Woolliams, 1997). On the other hand, the prediction of effective size from the drift approach can be extended to complex genetic systems, such as sex-linked loci (Nomura, 1997), linked autosomal loci (Santiago & Caballero, 1998) and nonrandom mating populations (Santiago & Caballero, 1995; Wang, 1998). All the extensions seem to be difficult for the long-term genetic contribution approach.

## Acknowledgements

I thank two anonymous referees for their comments on the manuscript. I also thank Dr J. A. Woolliams for sending me his submitted manuscript.

## References

- BULMER, M. G. 1980. *The Mathematical Theory of Quantitative Genetics*. Clarendon Press, Oxford.
- CABALLERO, A. 1994. Developments in the prediction of effective population size. *Heredity*, **73**, 657–679.
- CABALLERO, A. AND SANTIAGO, E. 1996. Response to selection from new mutation and effective size of partially inbred populations. 1. Theoretical results. *Genet. Res.*, **66**, 213–225.
- CROW, J. F. AND KIMURA, M. 1970. *An Introduction to Population Genetics Theory*. Harper & Row, New York.
- FALCONER, D. S. 1989. *Introduction to Quantitative Genetics*, 3rd edn. Longman, Harlow, Essex.
- NOMURA, T. 1996. Effective size of selected populations with overlapping generations. *J. Anim. Breed. Genet.*, **113**, 1–16.
- NOMURA, T. 1997. Effective population size for a sex-linked locus in populations under selection. *Math. Biosci.*, **142**, 79–89.
- ROBERTSON, A. 1961. Inbreeding in artificial selection programmes. *Genet. Res.*, **2**, 189–194.
- SANTIAGO, E. AND CABALLERO, A. 1995. Effective size of populations under selection. *Genetics*, **139**, 1013–1030.
- SANTIAGO, E. AND CABALLERO, A. 1998. Effective size and polymorphism of linked neutral loci in populations under directional selection. *Genetics*, **149**, 2105–2117.
- VILLANUEVA, B. AND WOOLLIAMS, J. A. 1997. Optimization of breeding programmes under index selection and constrained inbreeding. *Genet. Res.*, **69**, 145–158.

- VILLANUEVA, B., WOOLLIAMS, J. A. AND GJERDE, B. 1996. Optimum designs for breeding programmes under mass selection with an application in fish breeding. *Anim. Sci.*, **63**, 563–576.
- WANG, J. 1998. Effective size of populations under selection: some extensions. *Proc. 6th World Congr. Genet. Applied to Livestock Prod.*, **26**, 13–16.
- WOOLLIAMS, J. A. 1998. A recipe for the design of breeding schemes. *Proc. 6th World Congr. Genet. Applied to Livestock Prod.*, **25**, 427–431.
- WOOLLIAMS, J. A. AND THOMPSON, R. 1994. A theory of genetic contributions. *Proc. 5th World Congr. Genet. Applied to Livestock Prod.*, **19**, 127–134.
- 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 breeding in selected populations. *Genet. Res.*, **55**, 41–54.
- WRAY, N. R., WOOLLIAMS, J. A. AND THOMPSON, R. 1990. Methods for predicting rates of breeding in selected populations. *Theor. Appl. Genet.*, **80**, 503–512.
- WRAY, N. R., WOOLLIAMS, J. A. AND THOMPSON, R. 1994. Prediction of rates of breeding in populations undergoing index selection. *Theor. Appl. Genet.*, **87**, 878–892.