

## SHORT REVIEW

# Effective size of populations with heritable variation in fitness

T Nomura

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

The effective size of monogamous populations with heritable variation in fitness is formulated, and the expression obtained is compared with a published equation. It is shown that the published equation for dioecious populations is inappropriate for most animal and human populations, because the derivation is implicitly based on the assumption that

zygotes are produced by random union of gametes, each from conceptual male and female gametic pools. A convenient equation for practical use is proposed, and the application is illustrated with the estimation of the effective size of a rural human community in Japan.

*Heredity* (2002) 89, 413–416. doi:10.1038/sj.hdy.6800169

**Keywords:** effective population size; random genetic drift; fitness; heritability; human population; animal population

## Introduction

The effective size of a population is a parameter central to understanding evolution in small populations, because the magnitude of this parameter determines the genetic effects of both inbreeding and genetic drift (Falconer, 1989; Caballero, 1994). This parameter is also important for solving some practical issues; it has been used as a key parameter in designing strategies for conservation of endangered animal and plant species (eg, Lande and Barrowclough, 1987; Santiago and Caballero, 2000; Yonezawa *et al*, 2000) and breeding of domestic animals (eg, Caballero *et al*, 1996; Bijma *et al*, 2001).

The effective size of a population depends on various demographic and ecological variables (Nunney, 1991, 1993). Among these, the variation in the number of progeny per parent (family size) is one with the most pronounced effect (Crow and Denniston, 1988; Caballero, 1994). When the variation is due to non-inherited causes, the effective population size is simply a function of the variance of family size, and predictions have been developed for a variety of cases (Caballero, 1994; Wang, 1996; Wang and Caballero, 1999). However, when the differences in family sizes are, at least partially, due to inherited causes, formulation of the effective size is complicated by the fact that the offspring and later descendants of a parent with a large (or small) family size inherit the property.

The problem of heritable variation of family size in the formulation of the effective size was first addressed by Robertson (1961) in the context of artificial selection. He introduced the idea of the accumulation of selective advantages of individuals over generations. Following this approach, formulae applicable to various practical

situations have been developed (Santiago and Caballero, 1995; Nomura, 1996, 1997; Wang, 1998; Bijma *et al*, 2000, 2001). Only one extension to directional selection on fitness was given by Nei and Murata (1966). Based on the approach of Robertson (1961), they worked out an equation for monoecious populations. They also extended their derivation to dioecious populations. The equations have been used to estimate the effective population size of humans (Nei and Murata, 1966) and in wild animals (Ryman *et al*, 1981; Kelly, 2001).

In this paper, I will first derive an expression of the effective size of monogamous populations with heritable variation in fitness, and will compare it with the equations of Nei and Murata (1966). We will see that their equation for dioecious populations is inappropriate for most animal and human populations. Finally, a more useful equation for practical use is presented. The application will be illustrated with published data on a human community in Japan.

## Effective size of monogamous populations

Suppose an unstructured population of monogamous species, consisting of  $N/2$  couples each generation. Random mating and discrete generations are assumed. An autosomal neutral gene unlinked to the genes affecting fitness is considered. Formulation of the effective size is based on the amount of genetic drift in the frequency of the neutral allele. But as emphasized by Hill (1972), the same expression could be derived from the loss of heterozygosity (or the increase in inbreeding) in the neutral locus. For the effective size under selection, the identity has been proved by Nomura (1999) and Bijma *et al* (2000). Environmental correlation between parental and offspring fitness is assumed to be absent.

Let  $k_i$  be the number of progeny (which reach reproductive age) of couple  $i$ . Following the terminology in the related works (eg, Caballero and Santiago, 1995; Santiago and Caballero, 1995), we will refer to  $k_i$  as the 'fitness' of

Correspondence: T Nomura, Department of Biotechnology, Faculty of Engineering, Kyoto Sangyo University, Kyoto 603-8555, Japan. E-mail: nomurat@cc.kyoto-su.ac.jp

Received 25 October 2001; accepted 27 July 2002

couple  $i$ . Assuming an additive model, the fitness of couple  $i$  can be partitioned as:

$$k_i = g_{m(m)i} + g_{f(f)i} + e_i,$$

where  $g_{s(s)i}$  is the additive genetic component from parent of sex  $s$  ( $= m$  or  $f$ ), and  $e_i$  is the environmental (including the nonadditive and sampling) component. The subscript  $s$  in parentheses in  $g_{s(s)i}$  is attached to emphasize the sex of the parent. In general,  $g_{s(u)i}$  is the genetic component of fitness of parent of sex  $u$  ( $= m$  or  $f$ ), when the genotype is expressed in sex  $s$ . Note that for  $s \neq u$ , the genotype is not expressed in the fitness of the couple, but will be expressed when it is transmitted into descendants of sex  $s$ . The two genetic components of fitness ( $g_{m(m)i}$  and  $g_{f(f)i}$ ) may be partially controlled by different sets of genes, because some components of fitness, such as prenatal and postnatal maternal abilities, are sex-limited. Thus, we treat the contributions of male and female parents as two different sex-limited traits with genetic correlation  $r_g$ . The variance of  $k_i$  is decomposed as

$$V_k = V_{gm} + V_{gf} + V_e,$$

where  $V_{gs}$  and  $V_e$  are the variances of  $g_{s(s)i}$  and  $e_i$ , respectively. Putting

$$h_s^2 = V_{gs} / V_k,$$

the heritability of fitness may be written as

$$h^2 = \frac{V_{gm} + V_{gf}}{V_k} = h_m^2 + h_f^2. \quad (1)$$

The effective population size (denoted by  $N_e$ ) after one generation is approximated by

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

(Crow and Denniston, 1988; Caballero, 1994). The cumulative effect of selection can be incorporated into the above expression by considering the partial persistence of the random association between the relative fitness and the frequency of a neutral allele for which the effective size is defined (Santiago and Caballero, 1995, 1998). Let  $p_i$  ( $= 0, 1/4, 1/2, 3/4$  or  $1$ ) be the frequency of the neutral allele in couple  $i$ , and  $f_{s(u)i}$  ( $= g_{s(u)i}/2$ ) be the expected contribution of parent of sex  $u$  in couple  $i$  to the relative fitness of couple when the genotype is expressed in sex  $s$ . The frequency of the neutral allele in the population is  $p = \sum_{i=1}^{N/2} p_i / (N/2)$ . Only the two relative fitnesses,  $f_{m(m)i}$  and  $f_{f(f)i}$ , contribute to the change ( $\Delta p$ ) in the frequency of the neutral allele after one generation as

$$\Delta p = \text{cov}(p_i, f_{m(m)i}) + \text{cov}(p_i, f_{f(f)i})$$

(Santiago and Caballero, 1995, 1998). But in the following generations, all the associations between  $p_i$  and  $f_{s(u)i}$  are relevant to the total cumulative change of the gene frequency. Noting that under weak selection the associations are halved each generation due to segregation and recombination (Santiago and Caballero, 1995, 1998), and half of the persisting association contributes to the change in the gene frequency because of the sex-limited expression of  $f_{s(u)i}$ , the total cumulative change ( $\Sigma \Delta p$ ) can be obtained as

$$\Sigma \Delta p = \{1 + 1/2(1/2 + 1/4 + 1/8 + \dots)\} \text{cov}(p_i, f_{m(m)i}) + 1/2(1/2 + 1/4 + 1/8 + \dots) \text{cov}(p_i, f_{f(f)i})$$

$$\begin{aligned} & + 1/2(1/2 + 1/4 + 1/8 + \dots) \text{cov}(p_i, f_{m(f)i}) \quad (3) \\ & + \{1 + 1/2(1/2 + 1/4 + 1/8 + \dots)\} \text{cov}(p_i, f_{f(f)i}) \\ & = Q_1 \Delta p_{m(m)} + Q_2 \Delta p_{f(m)} + Q_2 \Delta p_{m(f)} + Q_1 \Delta p_{f(f)}, \end{aligned}$$

where

$$\begin{aligned} Q_1 &= 1 + 1/2(1/2 + 1/4 + 1/8 + \dots) = 3/2 \\ Q_2 &= 1/2(1/2 + 1/4 + 1/8 + \dots) = 1/2 \end{aligned}$$

and

$$\Delta p_{s(u)} = \text{cov}(p_i, f_{s(u)i}).$$

Following the argument of Santiago and Caballero (1995), the variance of  $\Delta p_{s(u)}$  can be written as

$$V(\Delta p_{s(u)}) = \frac{p(1-p)}{8N} V_{gs}.$$

Analogously, the covariance between  $\Delta p_{m(u)}$  and  $\Delta p_{f(u)}$  is expressed as

$$\text{cov}(\Delta p_{m(u)}, \Delta p_{f(u)}) = \frac{p(1-p)}{8N} r_g \sqrt{V_{gm} V_{gf}}.$$

With these expressions, the variance of  $\Sigma \Delta p$  is obtained as

$$\begin{aligned} V(\Sigma \Delta p) &= \frac{p(1-p)}{8N} \left[ (Q_1^2 + Q_2^2) V_{gm} \right. \\ & \quad \left. + 4Q_1 Q_2 r_g \sqrt{V_{gm} V_{gf}} + (Q_1^2 + Q_2^2) V_{gf} \right]. \end{aligned}$$

Since the contribution of  $V(\Sigma \Delta p)$  to  $1/N_e$  is

$$\begin{aligned} & \frac{1}{4N} \left[ (Q_1^2 + Q_2^2) V_{gm} + 4Q_1 Q_2 r_g \sqrt{V_{gm} V_{gf}} + (Q_1^2 + Q_2^2) V_{gf} \right] \\ &= \frac{1}{4N} \left[ \frac{5}{2} V_{gm} + 3r_g \sqrt{V_{gm} V_{gf}} + \frac{5}{2} V_{gf} \right]. \end{aligned}$$

an expression of  $N_e$  is obtained from equation 2 as

$$N_e = \frac{4N}{2 + [1 + 3/2 (h_m^2 + h_f^2 + 2r_g h_m h_f)] V_k}. \quad (4)$$

If male and female components of fitness are controlled by the same set of genes, ie,  $h_m^2 = h_f^2$  and  $r_g = 1$ , equation 4 reduces to

$$N_e = \frac{4N}{2 + (1 + 3h^2) V_k}, \quad (5)$$

which is the same expression as the equation for a monocious population given by Nei and Murata (1966).

## Comparison with published equations

Nei and Murata (1966) extended equation 5 to a dioecious population with equal numbers of male and female parents. Their equation comparable to equation 4 can be written as

$$N_e = \frac{4N}{2 + [1 + 3/4 (h_m^2 + h_f^2 + 2r_g h_m h_f)] V_k}, \quad (6)$$

which is different from equation 4. In the following, we consider the reason for the difference.

The implicit model assumed in the derivation of Nei and Murata (1966) is a random union of gametes (RUG) model. The RUG model assumed by Nei and Murata (1966) is such that male parent  $i$  contributes  $k_{mi}$  gametes to a male gametic pool, and female parent  $j$  contributes

$k_{fj}$  gametes to a female gametic pool. Zygotes are produced by random union of gametes each from the male and female gametic pool. In this model, the expected fitness of a parent is determined only by the genotype of the parent. Although the variation in fitness under this model is assumed to be due to the difference of fertility among parents (fertility selection), the following argument can be applied to the case where the variation in fitness is caused by the difference of viability among progeny (viability selection). The RUG model under viability selection is such that all male and female parents contribute equally their gametes to the respective gametic pool, and after random union of gametes, the survival of zygotes is determined according to their viability. Note that the expected fitness of a parent under this model is not affected by the genotypes of mates, because they are randomly sampled from the population (Caballero and Santiago, 1995). Thus, irrespective of the types of selection, fitness in the RUG model should be defined for individual parents rather than for couples.

This difference of the definition of fitness results in the difference of the accumulation of the change in gene frequency. This can be understood by expressing the frequency of the neutral allele in couple  $i$  in our derivation as  $p_i = (p_{mi} + p_{fi})/2$ , where  $p_{si}$  ( $= 0, 1/2$  or  $1$ ) is the gene frequency in parent of sex  $s$  in the couple. As seen from equation 3, all possible covariances (eight covariances) among  $p_{si}$  and the relative fitness  $f_{u(v)i}$  ( $u, v = m$  or  $f$ ) are accounted for in our derivation, while under the RUG model, the relevant covariances are limited to only four covariances [ $\text{cov}(p_{si}, f_{u(s)i})$  ( $s, u = m$  or  $f$ )], because couples are not created. By this omission of the covariances, the amount of the total accumulated effect is halved in the RUG model, compared to monogamous populations. This is reflected in the coefficient  $3/4$  in the denominator of equation 6, which is half the corresponding coefficient in equation 4. Similar reasoning has been made by Santiago and Caballero (1998) and Nomura (1999) to explain the inconsistency between equations for  $N_e$  under selection.

As in monogamous populations, the male and female heritabilities in equation 6 are defined as  $h_s^2 = V_{gs}/V_k$ . But equation 1 does not hold in the RUG model. When male and female components of fitness are controlled by the same set of genes, substituting  $h^2 \equiv h_m^2 = h_f^2$  and  $r_g = 1$  into equation 6 leads again to equation 5. But we should note that  $h^2$  in the equation for monogamous populations is the heritability of fitness of couples, while in the RUG model,  $h^2$  is the heritability of fitness of individuals.

## Application to estimation of $N_e/N$ in a human population

Before the application, we modify equation 4 to a more convenient form. In quantitative genetic studies of human populations, father-offspring and mother-offspring correlations of sibship size (or equivalently parent-son and parent-daughter correlations of the progeny number) have been estimated (Imaizumi *et al*, 1970, and references therein). Let  $COV_{k,km}$  and  $COV_{k,kf}$  be father-offspring and mother-offspring covariances of sibship size, respectively. Assuming monogamous marriages which will be reasonable for most human populations, at least approximately, the two covariances can be written as

$$COV_{k,ks} = \frac{1}{2} (h_s^2 + r_g h_m h_f) V_k$$

where  $s = m$  or  $f$ . Substituting this expression into equation 4 leads to

$$N_e = \frac{4N}{2 + V_k + 3(COV_{k,km} + COV_{k,kf})} \quad (7)$$

Note that by the use of this equation, we can estimate  $N_e$  without estimating the heritabilities and genetic correlation.

Imaizumi *et al* (1970) studied the heritability of human fertility in a rural community in Japan. From their results between 1921 and 1930, I obtained the following values:

$$V_k = 2.154; COV_{k,km} = 0.086; COV_{k,kf} = 0.252$$

Substituting these values into equation 7, we get  $N_e/N = 0.774$ . If heritable variation in fitness is neglected ( $h_m^2 = h_f^2 = 0$ ), the estimate of  $N_e/N$  from equation 2 is 0.963. On the assumption of the RUG model, the coefficient of covariance terms in equation 7 is halved, giving  $N_e/N = 0.858$ . However, these estimates should be seen cautiously because the covariances of sibship size over generations may be inflated by non-genetic causes, as discussed in the next section.

## Discussion

In theoretical population genetics, the RUG model has been frequently assumed. Some organisms do reproduce more or less in this fashion. For example, some aquatic animals release large numbers of gametes into the ocean, and some wind-pollinated plants release a vast number of male gametes (Nunney, 1993). However, in most animal mating systems, individuals (rather than gametes) come together to mate. In the present study, we have limited our concern to the simplest individual-based mating system, but the derived conclusion holds for more complex mating systems. To incorporate the heritable effect of fitness into the formulation of effective size of populations with individual-based mating systems, fitness should be defined for couples (or families in polygamy or polyandry), and all the random associations between the gene frequencies and the parental genetic components of fitness in a couple or family should be accounted for.

In our derivation, we assumed that the random association between the neutral gene and fitness is halved each generation. This assumption will lead to an overestimation of the cumulative change in the gene frequency, since selection reduces the random association through the reduction in the genetic variance of fitness (Santiago and Caballero, 1998). Caballero and Santiago (1995) formulated the effect of selection on the accumulation of change in the gene frequency in monoecious populations.

In human populations, however, traditional continuity or 'social inheritance' could have a more important effect on the persistence of the random association between the neutral gene and fitness. As suggested by Cavalli-Sforza and Bodmer (1971), the inheritance of family size might be psychological as well as be physiological. In fact, the family sizes in most human populations are apparently below the potential maximum (Cavalli-Sforza and Bodmer, 1971), indicating that some form of birth control must be in operation. In wild animal populations,

environmental correlation is likely an important cause for the resemblance of fitness between parents and their offspring (Crow and Denniston, 1988). The correlation could be positive if parent and progeny both occupy an especially favorable niche, leading to a persisting change in the frequency of the neutral allele in a similar process to inherited causes. For example, from mother-daughter regression, Kelly (2001) estimated the heritability of female fitness of Serengeti cheetahs as 0.89. This high heritability strongly suggests the presence of environmental correlation between fitness of mothers and daughters. The persisting effect of the environmental correlation on the effective population size is kept completely as long as the progeny remain in the same niche and is partially reduced if a part of them migrate to other niches. Statistical analysis of fitness over three generations, such as a comparison of parent-offspring and parent-grandoffspring correlations, will provide an estimate of the persisting proportion over generations. With an analogous argument to the present study, the persisting proportion could be used for expressing the cumulative effect on the effective population size.

Subdivided population structure can be a more important cause for the resemblance of fitness between parents and their progeny, if there are permanent environmental differences among subpopulations and a part of progeny remain in the same subpopulation as their parents (Wang and Caballero, 1999). Since the resemblance is partially reduced by the migration among subpopulations, the persisting effect on the effective size depends on the migration rate. Thus, the effect of population subdivision could be theoretically modeled with an analogy to the model of linkage between the neutral and selected loci, in which the persisting effect on the effective size is formulated with the recombination rate (Santiago and Caballero, 1998). Incorporation of the persisting effects due to such ecological causes could extend our understanding of evolution in small populations.

## References

- Bijma P, Van Arendonk JAM, Woolliams JA (2000). A general procedure to predict rates of inbreeding in populations undergoing mass selection. *Genetics* **154**: 1865–1877.
- Bijma P, Van Arendonk JAM, Woolliams JA (2001). Predicting rate of inbreeding for livestock improvement schemes. *J Anim Sci* **79**: 840–853.
- Caballero A (1994). Developments in the prediction of effective population size. *Heredity* **73**: 637–679.
- Caballero A, Santiago E (1995). Response to selection from new mutation and effective size of partially inbred populations. I. Theoretical results. *Genet Res* **66**: 213–225.
- Caballero A, Santiago E, Toro MA (1996). Systems of mating to reduce inbreeding in selected populations. *Anim Sci* **62**: 431–442.
- Cavalli-Sforza LL, Bodmer WF (1971). *The Genetics of Human Populations*. W. H. Freeman and Company: San Francisco.
- Crow JF, Denniston C (1988). Inbreeding and variance effective population numbers. *Evolution* **42**: 482–495.
- Falconer DS (1989). *Introduction to Quantitative Genetics*, 3rd edn. Longman: Essex.
- Hill WG (1972). Effective size of populations with overlapping generations. *Theor Pop Biol* **3**: 278–289.
- Imaizumi Y, Nei M, Furusho T (1970). Variability and heritability of human fertility. *Ann Hum Genet* **33**: 251–259.
- Kelly M (2001). Lineage loss in Serengeti cheetahs: consequences of high reproductive variance and heritability of fitness on effective population size. *Conserv Biol* **15**: 137–147.
- Lande R, Barrowclough GF (1987). Effective population size, genetic variation, and their use in population management. In: Soule ME (ed) *Variable Populations for Conservation*, Cambridge University Press: Cambridge. pp 87–123.
- Nei M, Murata M (1966). Effective population size when fertility is inherited. *Genet Res* **8**: 257–260.
- 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.
- Nomura T (1999). On the methods for predicting the effective size of populations under selection. *Heredity* **83**: 485–489.
- Nunney L (1991). The influence of age structure and fecundity on effective population size. *Proc R Soc B* **246**: 71–76.
- Nunney L (1993). The influence of mating system and overlapping generations on effective population size. *Evolution* **47**: 1329–1341.
- Robertson A (1961). Inbreeding in artificial selection programmes. *Genet Res* **2**: 189–194.
- Ryman N, Baccus R, Reuterwall C, Smith MH (1981). Effective population size, generation interval, and potential loss of genetic variability in game species under different hunting regimes. *OIKOS* **36**: 257–266.
- Santiago E, Caballero A (1995). Effective size of populations under selection. *Genetics* **139**: 1013–1030.
- Santiago E, Caballero A (1998). Effective size and polymorphism of linked neutral loci in populations under directional selection. *Genetics* **149**: 2105–2117.
- Santiago E, Caballero A (2000). Application of reproduction technologies to the conservation of genetic resources. *Conserv Biol* **14**: 1831–1836.
- Wang J (1996). Inbreeding and variance effective sizes for non-random mating populations. *Evolution* **50**: 1786–1794.
- Wang J (1998). Effective size of populations under selection: some extensions. *Proc 6th World Congr Genet Applied to Livestock Prod* **26**: 13–16.
- Wang J, Caballero A (1999). Developments in predicting the effective size of subdivided populations. *Heredity* **82**: 212–226.
- Yonezawa K, Kinoshita E, Watano Y, Zentho H (2000). Formulation and estimation of the effective size of stage-structured populations in *Frillaria camtschatscensis*, a perennial herb with a complex life history. *Evolution* **54**: 2007–2013.