

The Variance Effective Number for Mitochondrial Genes

Published in Genetika. 2000 Aug;36(8):1157-60

Dmitri V. Zaykin (Zaikin)^{*,†} and Alexander I. Pudovkin[†]

*Programs in Statistical Genetics and Biomathematics,
Department of Statistics,
North Carolina State University, Raleigh NC 27695-8203

[†]Institute of Marine Biology,
Far East Branch of the Russian Academy of Sciences,
Vladivostok 690041, Russia

Running head: Variance Effective Number

Keywords: variance effective population size

Address for correspondence:

Dmitri Zaykin
Program in Statistical Genetics
Department of Statistics
North Carolina State University
Raleigh NC 27695-8203.
phone: (919) 515 1919
FAX: (919) 515 7315
e-mail: zaykin@statgen.ncsu.edu

ABSTRACT

A formula for variance effective number (N_e) of a population considered for mitochondrial haplotypes is derived and discussed. The notion and the formula refer to the female part of the population. $N_e = \mu_k^2(N - 1)/\sigma_k^2$, where N is the number of females, σ_k^2 is the variance of reproductive success of individual females (measured in the number of progeny contributed to the next generation), and μ_k is the mean number of offspring averaged over all the females.

The effective number (N_e) of a population, a notion introduced by Wright (1931), is a critically important factor in consideration of natural populations (Lande and Barrowclough, 1987) and in simulations of population genetic processes (Beckenbach, 1994). N_e is the size of an ideal population that exhibits the same temporal change of population parameters (allelic and genotypic frequencies) as the real population under consideration. The ideal population has constant size, equal numbers of males and females, random mating, and multinomial contribution of parents to the next generation.

There are useful formulae (Crow and Kimura, 1970, pp. 345-365) which relate the size of this ideal population with the size of a real population, taking into account temporal changes of size, unequal sex ratio, and non-multinomial variance of individual reproductive contributions. Depending on what characteristic of the real population, inbreeding coefficient or intergeneration variance of allelic frequencies, is being compared to that of the ideal

population, one can speak of inbreeding or variance effective number. Crow and Kimura (1970) gave formulae for the inbreeding and variance N_e for nuclear genes. Recently, much research has focussed on mitochondrial genes (see Avise, 1994).

Birky et al. (1983) considered dynamics of frequency changes of mitochondrial and chloroplast genes in cell lineages within organisms and among individuals in populations for both diparentally and uniparentally transmitted organelle genes. However, these authors did not address the problem of effective number of breeders; they just maintained that the latter is approximately equal to the number of breeding individuals. Caballero (1994) gives a formula for the variance N_e for a constant size population of haploid organisms (which is applicable to the case of mitochondrial genes) and a formula for the case of n -ploid population. The latter formula (eq. 28) can be solved for $n = 1$ (haploid case), but the solution is different from our result (see discussion below).

We present here the formula for the variance N_e and its derivation. In our consideration of the problem we follow the approach of Crow and Kimura (1970, pp. 345-365) and Crow and Denniston (1988).

Though there are cases of diparental transmission of mitochondrial genes and occurrence of heteroplasmy (e.g. in mussels: Zouros et al., 1992; Skibinski et al., 1994; Wenne and Skibinski, 1995; see also Birky et al., 1989; Birky, 1995) in the majority of organisms mitochondrial genes are transmitted strictly maternally and all the individuals (both males

and females) are homoplasmic (Birky et al, 1983; Avise, 1994). We consider the following situation. Mitochondrial genes are transmitted strictly maternally; thus, we consider only the female part of the population, haplotypes of males being irrelevant for our consideration. Mothers produce eggs of the same haplotype (mutations and heteroplasmy are ignored). There are N females in the population, each of A or \bar{A} haplotype, p being the frequency of the A haplotype. The i th female transmits k_i gametes (eggs), with μ_k being the mean over all the females. σ_k^2 is the variance of the k_i 's. We consider non-overlapping generations.

For the ideal population, in which mothers contribute to the next generation binomially, so that each gamete is equally likely to come from each mother, the conditional variance of difference between haplotypic frequencies in adjacent generations is

$$\text{Var}(\Delta p \mid p) = \frac{p(1-p)}{N} \quad (1)$$

The variance N_e of a real population is the size of the ideal population, which has the drift variance of the same value as the real population. Hence, the variance N_e is

$$N_e = \frac{p(1-p)}{\text{Var}(\Delta p \mid p)} \quad (2)$$

It is shown in the Appendix that

$$\text{Var}(\Delta p \mid p) = \sigma_k^2 \frac{p(1-p)}{\mu_k^2(N-1)} \quad (3)$$

Substituting $\text{Var}(\Delta p \mid p)$ into (2) we obtain

$$\begin{aligned} N_e &= \frac{\mu_k(N-1)}{\frac{\sigma_k^2}{\mu_k}} \\ &= \frac{\mu_k^2(N-1)}{\sigma_k^2} \end{aligned} \quad (4)$$

For a population of constant size ($\mu_k = 1$) this reduces to

$$N_e = \frac{N-1}{\sigma_k^2} \quad (5)$$

When each female contributes binomially to the next generation,

$$k \sim \text{Bin}(N_t, \frac{1}{N_{t-1}}), \quad (6)$$

so the variance is

$$\sigma_k^2 = \mu_k \left(1 - \frac{1}{N_{t-1}}\right) \quad (7)$$

and N_e becomes $\mu_k N_{t-1} = N_t$. For the population of constant size $N_e = N$. If the whole next generation originates from a single female, $\sigma_k^2 = \mu_k^2(N_{t-1} - 1)$ and N_e becomes 1 as it should.

As mentioned earlier, Caballero (1994) gives a formula for the variance N_e for n -ploid population. When solved for $n = 1$ it is reduced to the following:

$$\begin{aligned}
N_e &= \frac{N-1}{\frac{\sigma_k^2}{\mu_k}} \\
&= \frac{\mu_k(N-1)}{\sigma_k^2}
\end{aligned}
\tag{8}$$

One can see that this formula differs from (4) in the first term μ_k (the average number of progeny), which is squared in our formula and is in the power 1 in (8). Our result seems quite reasonable as the variance N_e should be related to the number of offspring (see Crow and Kimura, 1970, p. 361) rather than parents as is in (8). We believe that this difference is caused by inconsistency in notation in Caballero (1994). He explicitly writes that N (in his eqs. 26, 27, 28) is the number of individuals in parental generation (though for the case of a constant population size he considers the distinction between the number of parents and offspring is irrelevant), while in the formula for the variance of change in gene frequency (Crow and Morton, 1955; eq. 11) N means the number of progeny. If one substitutes N in eq.28 (Caballero, 1994) for $N_{t-1}\mu_k$ and solves this equation for the haploid case ($n = 1$), one arrives at our solution (4) of the haploid N_e .

Chesser and Baker (1996) consider effective sizes in spatially subdivided populations for both uniparentally and diparentally inherited genes. Equation (34) in this paper gives the loss-of-variation effective size ($N_{e(L)}$), which is conceptually equivalent to the inbreeding effective size for a diploid population, for maternally transmitted genes in an undivided population:

$$N_{e(L)} = \frac{\mu_k(\mu_k N - 1)}{\sigma_k^2 + \mu_k(\mu_k - 1)} \quad (9)$$

For constant size populations ($\mu_k = 1$) $N_{e(L)}$ and $N_{e(V)}$, derived by us, coincide, but in other cases they differ.

Let us compare effective sizes of a population ($N_{e[nuc]}$ and $N_{e[mt]}$) for the nuclear and mitochondrial genes. Let N_{f+m} be the total number of females and males. According to Crow and Kimura (1970, p. 362) the variance N_e for nuclear genes for the population of constant size of a bisexual species is $N_{e[nuc]} = (4N_{f+m} - 4)/(\sigma_k^2 + 2)$. As shown before, $N_{e[mt]} = (N_f - 1)/\sigma_k^2$. With random mating and binomial contributions of individual parents $N_{e[nuc]} = 2N_{f+m}$ and $N_{e[mt]} = N_f$. When sex ratio is 1:1, $N_f = N_{f+m}/2$ and the ratio $N_{e[mt]}/N_{e[nuc]} = 0.25$ as is usually stated (Takahata and Maruyama, 1981 and many later publications).

An interesting case is when equal numbers of progeny are contributed by each female. σ_k^2 is zero in this case and $N_{e[mt]}$ will be infinity. This result is quite reasonable. When σ_k^2 is zero, the haplotype frequency in the progeny will be the same as in mothers, as would have been in the ideal infinite population (with zero drift variance). Interestingly, variance $N_{e[nuc]}$ for this case of $\sigma_k^2 = 0$ will be twice the number of breeding individuals, that is very different from the mitochondrial gene case. Incidentally, the inbreeding N_e for the case of equal parental contribution is also infinity (Crow and Kimura, 1970; p. 361). This case of

zero variance of reproductive contribution is instructive in the respect that it clearly shows that it is not always possible to simply replace $2N_{e[nuc]}$ with N_e in adapting for mitochondrial case the equations derived for nuclear genes as is recommended by Birky et al., 1983 (pp. 516, 518).

Although we derived the formula (4) having in mind the case of mitochondrial genes, the formula is equally applicable for any case of uniparentally transmitted genes (e.g. maternally inherited in many plants chloroplast genes, paternally inherited in mammals Y-chromosome genes) or for haploid populations (e.g. bacterial).

ACKNOWLEDGMENTS

This research was partly supported by National Institutes of Health grant GM-43544 to North Carolina State University and by a grant of the National Science Foundation OCE-9301416 to Bodega Marine Laboratory, UC Davis. The authors thank D Hedgecock, BS Weir, AT Beckenbach and two anonymous reviewers for comments on the manuscript.

APPENDIX

Let k_i be a number of gametes transmitted by k th female. The frequency of allele A in the following generation is $\sum_A k_i / \sum k_i$, where \sum_A denotes summation extending through carriers of A . N_{t-1} and $N_{A; t-1}$ are the census population size (the number of breeding females) and the number of alleles A (equals the number of carriers of A) at generation $(t-1)$, respectively.

The frequency change is

$$\Delta p = \frac{\sum_A k_i}{\sum k_i} - p \quad (10)$$

$$= \frac{\sum_A k_i}{\sum k_i} - \frac{N_{A; t-1}}{N_{t-1}} \quad (11)$$

Omitting $t-1$ subscript in $N_{A; t-1}$ and N_{t-1} for now, and writing $\text{Var}(\Delta p) \equiv \text{Var}(\Delta p | p)$

$$\text{Var}(\Delta p | N_A) = E\{(\Delta p | N_A)^2\} - (E\{\Delta p | N_A\})^2 \quad (12)$$

The first term, $E\{(\Delta p | N_A)^2\}$ is

$$E\{(\Delta p | N_A)^2\} = E\left\{\left(\frac{\sum_A k_i}{\sum k_i} - \frac{N_{A; t-1}}{N_{t-1}}\right)^2\right\} \quad (13)$$

$$= E\left\{\left(\frac{N \sum_A k_i - N_A \sum k_i}{N^2 \mu_k}\right)^2\right\} \quad (14)$$

$$= \frac{1}{N^4 \mu_k^2} E\left\{\left(N \sum_A k_i - N_A \sum k_i\right)^2\right\} \quad (15)$$

$$= \frac{1}{N^4 \mu_k^2} E\left\{\left(N \sum_A k_i - (\sum_A 1) \mu_k N\right)^2\right\} \quad (16)$$

$$= \frac{1}{N^4 \mu_k^2} E\left\{\left(N \sum_A k_i - N \sum_A \mu_k\right)^2\right\} \quad (17)$$

$$= \frac{1}{N^4 \mu_k^2} E\left\{N^2 \left(\sum_A (k_i - \mu_k)\right)^2\right\} \quad (18)$$

$$= \frac{1}{N^2 \mu_k^2} E\left\{\left(\sum_A (k_i - \mu_k)\right)^2\right\} \quad (19)$$

$$= \frac{1}{N^2 \mu_k^2} E \left\{ \sum_A (k_i - \mu_k)^2 + \sum_{A; i \neq j} (k_i - \mu_k)(k_j - \mu_k) \right\} \quad (20)$$

$$= \frac{1}{N^2 \mu_k^2} \frac{N}{N-1} \left(N_A \sigma_k^2 + N_A(N_A - 1)(-1) \frac{1}{N-1} \sigma_k^2 \right) \quad (21)$$

$$= \frac{1}{N^2 \mu_k^2} \frac{N}{N-1} \sigma_k^2 \left(N_A - \frac{N_A(N_A - 1)}{N-1} \right) \quad (22)$$

$$= \frac{\sigma_k^2}{\mu_k^2 N(N-1)^2} N_A(N - N_A) \quad (23)$$

By a similar argument,

$$E\{\Delta p | N_A\} = \frac{1}{N \mu_k} E \left\{ \sum_A (k_i - \mu_k) \right\} = 0 \quad (24)$$

Then

$$\text{Var}(\Delta p) = \text{Var} \{ E(\Delta p | N_A) \} + E \{ \text{Var}(\Delta p | N_A) \} \quad (25)$$

$$= E \{ \text{Var}(\Delta p | N_A) \} \quad (26)$$

$$= \frac{\sigma_k^2}{\mu_k^2 N(N-1)^2} E \{ N_A(N - N_A) \} \quad (27)$$

$$= \frac{\sigma_k^2}{\mu_k^2 N(N-1)^2} N(N-1)p(1-p) \quad (28)$$

$$= \frac{\sigma_k^2 p(1-p)}{\mu_k^2 (N-1)} \quad (29)$$

LITERATURE CITED

Avise J. C., 1994 Molecular Markers, Natural History and Evolution. New York, Chapman and Hall.

Beckenbach A. T., 1994 Mitochondrial haplotype frequencies in oysters: neutral alternative to selection models. In: Non-Neutral Evolution (Golding B, ed.). New York: Chapman and Hall, 188-198.

Birky C.W., Jr. 1995 Uniparental inheritance of mitochondrial and chloroplast genes – mechanisms and evolution. Proc. Natl. Acad. Sci. USA **92**: 11331-11338.

Birky C.W., Jr., Maruyama a T., Fuerst P. 1983 An approach to population and evolutionary genetic theory for genes in mitochondria and chloroplasts, and some results. Genetics **103**: 513-527.

Birky C. W., Jr, Fuerst P., and Maruyama T., 1989 Organelle gene diversity under migration, mutation, and drift: equilibrium expectations, approach to equilibrium, effects of heteroplasmic cells, and comparison to nuclear genes. Genetics **121**: 613-627.

Caballero A. 1994 Developments in the prediction of effective population size. Heredity **73**:

657-679.

Chesser, R.K. and Baker, R.J. 1996 Effective sizes and dynamics of uniparentally and diparentally inherited genes. *Genetics* **144**: 1225-1235.

Crow J. F. and Kimura M., 1970 *An Introduction to Population Genetics Theory*, Harper and Row, New York.

Crow, J.F. and Morton N.E., 1955 Measurement of gene frequency drift in small populations. *Evolution* **9**: 202-214.

Crow J. F. and Denniston C 1988 Inbreeding and variance effective population numbers. *Evolution* **42**: 482-495.

Lande R. and Barrowclough G. F., 1987 Effective population size, genetic variation, and their use in population management. In: *Viable Populations for Conservation* (Soulé M, ed.). Cambridge, England: Cambridge University Press; 87-124.

Skibinski, D.O.F., Gallagher C., Beynon C.M. 1994 Sex-limited mitochondrial DNA transmission in the marine mussel *Mytilus edulis*. *Genetics* **138**: 801-809.

Takahata N., Maruyama T. 1981 A mathematical model of extranuclear genes and the genetic variability maintained in a finite population. *Genetical Research* **37**: 291-302.

Wenne R., Skibinski D.O.F. 1995 Mitochondrial DNA heteroplasmy in European populations of the mussel *Mytilus trossulus*. *Marine Biology* **122**: 619-624.

Wright S, 1931 Evolution in Mendelian population. *Genetics* **16**: 97- 159.

Zouros E., Freeman K.R., Ball A.O., Pogson G.H. 1992 Direct evidence for extensive paternal mitochondrial DNA inheritance in the marine mussel *Mytilus*. *Nature* **359**: 412-414.