The Expected Number of Heterozygous Sites in a Subdivided Population

Thomas Nagylaki

Department of Ecology and Evolution, University of Chicago, Chicago, Illinois 60637

Manuscript received December 5, 1997
Accepted for publication March 16, 1998

ABSTRACT

A simple, exact formula is derived for the expected number of heterozygous sites per individual at equilibrium in a subdivided population. The model of infinitely many neutral sites is posited; the linkage map is arbitrary. The monoecious, diploid population is subdivided into a finite number of panmictic colonies that exchange gametes. The backward migration matrix is arbitrary, but time independent and ergodic (*i.e.*, irreducible and aperiodic). With suitable weighting, the expected number of heterozygous sites is $4N_e u$, where N_e denotes the migration effective population number and u designates the total mutation rate per gene (or DNA sequence). For diploid migration, this formula is a good approximation if $N_e \gg 1$.

ONE of the most important measures of genetic variability at the molecular level is the expected number of heterozygous nucleotide sites per individual, $\bar{d}_0$. For a panmictic population at equilibrium and without selection, Kimura (1969) showed that in his model of infinitely many sites,

$$\bar{d}_0 = 4N_e u, \tag{1}$$

where N_e represents the effective number of monoecious, diploid individuals and u signifies the total mutation rate per gene (or DNA sequence). Ewens (1974) and Watterson (1975) have presented alternative derivations of this basic result.

Since natural populations are frequently subdivided, considerable effort has been devoted to extending (1) to subdivided populations. Li (1976) proved that (1) holds for the island model without recombination if $N_e = N_T$, the total population number. Slatkin (1987) generalized Li's result by demonstrating that (1) holds if (i) there is no recombination; (ii) the backward migration matrix M is symmetric and ergodic; (iii) $\bar{d}_0$ is calculated by weighting each deme by the reciprocal of the number of individuals in it; and (iv) $N_e = n\tilde{N}$, where n signifies the number of demes and $\tilde{N}$ denotes the harmonic mean of the subpopulation numbers. Strobeck (1987) established (1) with $N_e = N_T$ for weak evolutionary forces and conservative migration. See Griffiths (1981), Takahata (1988), Tajima (1989), Notohara (1990, 1993, 1997), Hey (1991), Nath and Griffiths (1993), and Herbots (1994) for related studies.

In this note, we prove for gametic dispersion that, for suitably weighted calculation of $\bar{d}_0$, the formula (1) holds for every linkage map if M is arbitrary, but ergodic and time independent, and if N_e designates the migration effective population number (Nagylaki 1980, 1982, 1983, 1994). We show also that (1) is a good approximation for diploid migration if $N_e \gg 1$.

Address for correspondence: Department of Ecology and Evolution, University of Chicago, 1101 E. 57th St., Chicago, IL 60637.

GAMETIC DISPERSION

Generations are discrete and nonoverlapping; the monoecious, diploid population is subdivided into a finite number of panmictic colonies that exchange gametes in a fixed pattern. We apply the model of infinitely many neutral sites with an arbitrary linkage map to a gene or DNA sequence. Thus, we posit that the mutation rate per site is so low that mutation occurs at each site at most once and then only at monomorphic sites. This approximation requires that the proportion of polymorphic sites be much less than one. Let u denote the total mutation rate per gene.

At the beginning of the life cycle, every one of the N_i adults in deme i produces the same very large number of gametes, which then disperse independently. Complete random union of gametes follows. Therefore, a proportion $1/N_i$ of the zygotes whose gametes originate in deme i are produced by self-fertilization. Mutation is next, and finally population regulation returns the number of individuals in deme i to N_i. Thus, random genetic drift operates through population regulation.

Before deriving our results, we introduce some essential concepts and parameters.

Let m_{ij} designate the probability that a gamete in deme i after dispersion was produced in deme j. In the absence of selection, it is reasonable to assume that the backward migration matrix $M = (m_{ij})$ is constant (Nagylaki 1992, p. 135). We posit also that M is ergodic, *i.e.*, irreducible and aperiodic (Gantmacher 1959, pp. 50, 80, 88). Irreducibility guarantees that the descendants of individuals in each deme are able eventually to reach every other deme. Aperiodicity precludes

pathological cyclic behavior. Given irreducibility, the biologically trivial condition that individuals have positive probability of remaining in some deme, *i.e.*, that $m_{ii} > 0$ for some i, suffices for aperiodicity (Feller 1968, p. 426). Of course, M must be stochastic:

$$\sum_j m_{ij} = 1. \tag{2}$$

Let N_T and κ_i represent the total population number and the proportion of adults in deme i, respectively:

$$N_T = \sum_i N_i, \quad \kappa_i = N_i / N_T, \tag{3a}$$

$$0 < \kappa_i < 1, \quad \sum_i \kappa_i = 1. \tag{3b}$$

By the ergodicity of the nonnegative stochastic matrix M, the eigenvalue 1 of M is simple and exceeds all other eigenvalues in absolute value; we may choose the left eigenvector ν corresponding to this unit eigenvalue to have only positive components (Gantmacher 1959, Chapter 13). Thus, the conditions

$$0 < \nu_i < 1, \quad \sum_i \nu_i = 1, \quad \nu^T M = \nu^T, \tag{4}$$

where the superscript T signifies matrix transposition, determine ν uniquely. Note that ν is the unique stationary distribution of the Markov chain with transition matrix M.

The components of the last equation in (4) are

$$\nu_i = \sum_j \nu_j m_{ji} = \nu_i m_{ii} + \sum_{j:\, j\neq i} \nu_j m_{ji}.$$

With the aid of (2), we can rewrite this as

$$\nu_i = \sum_{j:\, j\neq i} \left(m_{ji} \Big/ \sum_{k:\, k\neq i} m_{ik} \right) \nu_j,$$

which reveals that the stationary distribution ν depends only on the relative migration rates. More precisely, if we replace m_{ij} by cm_{ij} for every i and j such that $i \neq j$, where the constant c is independent of i and j and the restriction

$$c \sum_{j:\, j\neq i} m_{ij} \leq 1$$

holds, then ν is unaltered.

Conservative migration patterns are those that do not change the subpopulation numbers; in this case, and only in this case, we have $\nu = \kappa$ (Nagylaki 1980). Conservative migration has many simple intuitive properties that do not always hold for arbitrary migration (Nagylaki 1980, 1982, 1983, 1985, 1986, 1992, pp. 135–136, 151; Nordborg 1997). In our model, the subpopulation numbers N_i refer to adults. However, since the number of gametes in each deme before dispersion is proportional to N_i, it is also true that the gametic numbers are unchanged by conservative migration, and only by conservative migration.

In our results, the vectors κ and ν enter combined in the migration effective population number N_e, defined by (Nagylaki 1980, 1982, 1983, 1994)

$$N_e = \beta N_T, \quad \beta = \left(\sum_i \nu_i^2 / \kappa_i \right)^{-1}. \tag{5}$$

We have $\beta \leq 1$ and hence $N_e \leq N_T$, with equality if and only if migration is conservative (Nagylaki 1980). This effective population number replaces the actual total population number in the strong-migration limit (Nagylaki 1980, 1983) and in certain aspects of geographical invariance (Nagylaki 1982, 1994). Observe that N_e is independent of the genetic model.

There is a simple, intuitive interpretation of N_e (*cf.* Nordborg 1997). In many cases, $1/N_e$ can be defined as the probability that two randomly chosen gametes in distinct individuals are descended from the same parent (Crow and Denniston 1988; Caballero and Hill 1992; Nagylaki 1992, pp. 243–247, 1995). For gametes in demes i and j, this probability is

$$\sum_k m_{ik} m_{jk} / N_k.$$

Averaging this with respect to the stationary distribution ν and using (4), (3a), and (5) yield

$$\sum_{i,j} \nu_i \nu_j \sum_k m_{ik} m_{jk} / N_k = \sum_k \nu_k^2 / (N_T \kappa_k) = 1 / N_e.$$

We are now prepared to deduce our results.

Let T_{ij} denote the mean coalescence time (in generations) of two distinct, homologous nucleotides chosen at random from adults just before gametogenesis, one from deme i and one from deme j. At equilibrium, considering ancestry and coalescence in the preceding generation yields directly

$$T_{ij} = \sum_{k,l:\, k\neq l} m_{ik} m_{jl} (1 + T_{kl}) + \sum_k m_{ik} m_{jk} \left[\frac{1}{2N_k} + \left(1 - \frac{1}{2N_k} \right) (1 + T_{kk}) \right],$$

and (2) simplifies this at once to

$$T_{ij} = 1 + \sum_{k,l} m_{ik} m_{jl} T_{kl} - \sum_k (2N_k)^{-1} m_{ik} m_{jk} T_{kk}. \tag{6}$$

Clearly, (6) applies also to a model with $2N_i$ haploid individuals in deme i.

Define the global and local means (*cf.* Nagylaki 1982)

$$\overline{T} = \sum_{i,j} \nu_i \nu_j T_{ij}, \quad \overline{T}_0 = \beta \sum_i (\nu_i^2 / \kappa_i) T_{ii}. \tag{7}$$

Averaging (6) according to (7) and appealing to (4) and (3a) yield

$$\overline{T} = 1 + \overline{T} - \frac{\overline{T}_0}{2N_T \beta},$$

whence (5) gives

$$\overline{T}_0 = 2N_e. \tag{8}$$

Therefore, summing mutations over sites shows that the expected number of nucleotide differences is

$$\bar{d}_0 = 2u\bar{T}_0 = 4N_e u. \tag{9}$$

Thus, with the weighting (7), the exact effect of population subdivision in (8) and (9) is to replace the actual total population number N_T by the migration effective population number N_e.

In the strong-migration limit, $N_i \to \infty$ for every i with κ and M fixed. Then $T_{ij} \sim 2N_e$ for every i and j (Notohara 1993), where the notation means that $T_{ij}/(2N_e) \to 1$. This demonstrates independently the asymptotic validity of the exact formula (8) whenever migration dominates random drift.

We discuss special cases of (8) after presenting a different proof of (9).

An alternative proof of (9): Since (8) and (9) are geographical-invariance relations, the following instructive approach is natural. Suppose the model of infinitely many alleles (Malécot 1946, 1948, 1951; Wright 1948; Kimura and Crow 1964) applies to each site: every nucleotide at site s mutates to new nucleotides at rate u_s. (We soon let $u_s \to 0$, so the fact that there are really only four nucleotides will not matter.)

Let $f_{ij}^{(s)}$ denote the probability that two distinct nucleotides at site s chosen at random from adults just before gametogenesis, one from deme i and one from deme j, are the same. Define the weighted means (Nagylaki 1982)

$$\bar{f}^{(s)} = \sum_{i,j} \nu_i \nu_j f_{ij}^{(s)}, \quad \bar{f}_0^{(s)} = \beta \sum_i (\nu_i^2/\kappa_i) f_{ii}^{(s)}, \tag{10}$$

$$\bar{h}_0^{(s)} = 1 - \bar{f}_0^{(s)}; \tag{11}$$

clearly, $\bar{h}_0^{(s)}$ is the weighted average nucleotide heterozygosity. At equilibrium, these satisfy the geographical-invariance relation (Nagylaki 1982)

$$\bar{h}_0^{(s)} = \left[\frac{2N_e u_s(2 - u_s)}{(1 - u_s)^2}\right]\bar{f}^{(s)}. \tag{12}$$

To obtain the expected number of heterozygous sites in the model of infinitely many sites, we must let $u_s \to 0$ and sum over s in (12). Evidently, $\bar{f}^{(s)} \to 1$ as $u_s \to 0$, so (12) reduces to

$$\bar{h}_0^{(s)} \sim 4N_e u_s, \tag{13}$$

whence we get

$$\bar{d}_0 = \sum_s \bar{h}_0^{(s)} = 4N_e u, \tag{14}$$

where

$$u = \sum_s u_s. \tag{15}$$

Conservative migration: If migration is conservative, then $N_e = N_T$, so (9) reduces to a result established by Strobeck (1987) for weak evolutionary forces. In this case, $\nu = \kappa$, and hence the averages in (8) and (9) simplify to weighting by the demic proportions:

$$\bar{T}_0 = \sum_i \kappa_i T_{ii} = 2N_T. \tag{16}$$

Examples of conservative migration are random outbreeding and site homing (Nagylaki 1992, pp. 136, 149, and refs. therein), the island model (Nagylaki 1983, 1986, and refs. therein), and the circular stepping-stone model (Nagylaki 1983, 1986, and refs. therein). The choice $m_{ij} = \kappa_j$ corresponds to panmixia in the entire population.

Note that (16) is independent of the migration pattern, provided the latter is conservative. This raises the following apparent paradox. If there is no migration, then $T_{ii} = 2N_i$, whence

$$\bar{T}_0 = 2\sum_i \kappa_i N_i \equiv 2\bar{N}, \tag{17}$$

but (17) is obviously not the limit of (16) as the migration rates converge to zero. This phenomenon can be understood on several levels.

From the formal point of view, note that if there is no migration, then M is the identity matrix. Therefore, contrary to our assumption of ergodicity, M is reducible and ν is undefined. Thus, (16) does not apply.

A more illuminating explanation is that, as the migration rates tend to zero, so does the probability of descent from a different deme, but the mean interdeme coalescence times (T_{ij} for $i \neq j$) diverge and make a finite, positive contribution to the mean intrademe coalescence times T_{ii}. This behavior is exemplified by the island model with migration rate m: Li's (1976) solution shows that $T_{ij} = O(m^{-1})$ for $i \neq j$ and that the interdeme contribution is $O(1)$ as $m \to 0$. For two islands, Nath and Griffiths (1993) demonstrate that the distribution of the intrademe coalescence time converges to the single-deme distribution, but the mean intrademe coalescence time does not converge to the single-deme mean.

Doubly stochastic backward migration matrix: Here we assume, in addition to (2), that

$$\sum_i m_{ij} = 1 \tag{18}$$

for every j. Then

$$\nu^T = \frac{1}{n}(1, 1, \ldots, 1) \tag{19}$$

is the unique solution of (4) for n demes. Therefore, (5) yields

$$N_e = n\tilde{N}, \tag{20a}$$

where

$$\tilde{N} = \left(\frac{1}{n}\sum_i \frac{1}{N_i}\right)^{-1} \tag{20b}$$

designates the harmonic mean of the subpopulation numbers. Thus, N_e can be much smaller than N_T.

A natural subclass of doubly stochastic M is homogeneous M: in this case, $m_{ij} = m_{i-j}$, which depends only on displacement, rather than on both the initial and final positions. Examples are the island and circular stepping-stone models, but, as observed above, these migration patterns are also conservative (Nagylaki 1992, p. 136).

Symmetric M is another subclass of doubly stochastic M. In this case, the formula $\bar{d}_0 = 4n\tilde{N}u$ was derived by Slatkin (1987) under the assumption of no recombination. (His mutation rate, however, should be per gene, not per site.)

Two demes: Parametrizing M as

$$M = \begin{pmatrix} 1 - m_1 & m_1 \\ m_2 & 1 - m_2 \end{pmatrix}, \tag{21}$$

from (4) and (5) we find

$$\nu = \frac{1}{m_1 + m_2}\begin{pmatrix} m_2 \\ m_1 \end{pmatrix}, \tag{22}$$

$$N_e = \frac{N_1 N_2 (m_1 + m_2)^2}{N_1 m_1^2 + N_2 m_2^2}. \tag{23}$$

We can confirm (23) by calculating $\overline{T}_0$ from Notohara's (1990, p. 69) formulas for T_{ij}.

Migration is conservative if $\nu = \kappa$, which is equivalent to $N_1 m_1 = N_2 m_2$. This condition means that the same number of individuals migrate from deme 1 to deme 2 as vice versa.

DIPLOID MIGRATION

In this section, we provide support for the robustness of (8) and (9) by proving that (8) is a good approximation for diploid migration if $N_e \gg 1$. We derive exact results for conservative migration and weak- and strong-migration approximations for the general case.

We modify the model in the preceding section so that selfing is excluded and zygotes (rather than gametes) disperse, still before population regulation.

Let S_i designate the mean coalescence time of two distinct, homologous nucleotides chosen at random just before gametogenesis from an adult in deme i. Let T_{ij} signify the mean coalescence time of two homologous nucleotides chosen from distinct adults just before gametogenesis, one from deme i and one from deme j. A moment's reflection shows that at equilibrium,

$$S_i = \sum_k m_{ik}(1 + T_{kk}) = 1 + \sum_k m_{ik} T_{kk}, \tag{24a}$$

$$\begin{aligned} T_{ij} &= \sum_{k,l:\, k\neq l} m_{ik} m_{jl}(1 + T_{kl}) + \sum_k m_{ik} m_{jk} \\ &\quad \times \left\{ \frac{1}{N_k}[\tfrac{1}{2} + \tfrac{1}{2}(1 + S_k)] + \left(1 - \frac{1}{N_k}\right)(1 + T_{kk}) \right\} \\ &= 1 + \sum_{k,l} m_{ik} m_{jl} T_{kl} - \sum_k (2N_k)^{-1} m_{ik} m_{jk}(2T_{kk} - S_k). \end{aligned} \tag{24b}$$

We retain (7) and define

$$\overline{S}_0 = \beta \sum_i (\nu_i^2 / \kappa_i) S_i. \tag{25}$$

Averaging (24b) and using (4), (3a), (5), and (25), we obtain the invariance formula

$$2\overline{T}_0 - \overline{S}_0 = 2N_e. \tag{26}$$

Conservative migration: Since the number of zygotes in each deme before migration is proportional to the number of adults, therefore it is conservative migration, and only conservative migration, that leaves the zygotic numbers invariant. If migration is conservative, then $\overline{S}_0$ and $\overline{T}_0$ are averaged with respect to κ, as in (16). Therefore, averaging (24a) yields

$$\overline{S}_0 = 1 + \overline{T}_0. \tag{27}$$

Recalling that $N_e = N_T$ and solving (26) and (27) simultaneously, we find

$$\overline{S}_0 = 2(N_T + 1), \quad \overline{T}_0 = 2N_T + 1. \tag{28}$$

Let $\bar{d}_0$ and $\bar{e}_0$ denote the expected number of nucleotide differences between two homologous genes in the same individual and in different individuals in the same deme, respectively. Summing mutations over sites, we have

$$\bar{d}_0 = 2u\overline{S}_0 = 4u(N_T + 1), \tag{29a}$$

$$\bar{e}_0 = 2u\overline{T}_0 = 2u(2N_T + 1). \tag{29b}$$

If $N_T \gg 1$, then (28) and (29) are very close to (8) and (9), respectively.

One can also deduce (29) by the alternative approach presented in the preceding section: approximate Equations (19b) and (20) of Nagylaki (1985) for weak mutation and sum over sites.

Weak migration: Let m represent the largest total migration rate:

$$m = 1 - \min_i m_{ii}. \tag{30}$$

We derive an approximation for $m \ll 1$ and arbitrary subpopulation numbers N_i. From (2) and (30) we see that

$$m_{ij} = \delta_{ij} + O(m) \tag{31}$$

as $m \to 0$, where δ_{ij} denotes the Kronecker delta ($\delta_{ii} = 1$, and $\delta_{ij} = 0$ if $i \neq j$). Substituting (31) into (24a) gives

$$S_i = 1 + T_{ii} + O(m), \tag{32}$$

and averaging (32) according to (25) and (7), we obtain

$$\overline{S}_0 = 1 + \overline{T}_0 + O(m). \tag{33}$$

The joint solution of (26) and (33) is our weak-migration approximation:

$$\overline{S}_0 = 2(N_e + 1) + O(m), \tag{34a}$$

$$\overline{T}_0 = 2N_e + 1 + O(m); \tag{34b}$$

$$\bar{d}_0 = 2u[2(N_e + 1) + O(m)], \quad (35a)$$

$$\bar{e}_0 = 2u[2N_e + 1 + O(m)] \quad (35b)$$

as $m \to 0$.

Observe that (34) and (35) agree with (28) and (29), respectively, for weak conservative migration.

Strong migration: If $N_i \gg 1$ but $m \not\ll 1$, then we can assume that $N_i \to \infty$ for every i with the backward migration matrix M fixed. In this limit, the result (8) indicates that we must have the asymptotic formulas

$$S_i \sim N_e S_i^*, \quad T_{ij} \sim N_e T_{ij}^* \quad (36)$$

as $N_e \to \infty$, where S_i^* and T_{ij}^* are independent of N_e. The leading terms in (24b) yield

$$T_{ij}^* = \sum_{k,l} m_{ik} m_{jl} T_{kl}^*. \quad (37)$$

By the ergodicity of M, the nonnegative, stochastic Kronecker-product matrix $M \otimes M$ in (37) has a simple maximal eigenvalue 1, and the corresponding right eigenvector has equal components. Therefore, $T_{ij}^* = T^*$, independent of i and j, and hence (24a) implies that $S_i^* = T^*$. From (26) we get

$$\bar{S}_0 \sim \bar{T}_0 \sim 2N_e, \quad \bar{d}_0 \sim \bar{e}_0 \sim 4N_e u \quad (38)$$

as $N_i \to \infty$ with M fixed.

This result agrees with (8) and (9), as expected from the strong-migration limit for diploid migration in the model of infinitely many alleles (Nagylaki 1983).

DISCUSSION

We have demonstrated that, for gametic dispersion and suitable averaging, the mean intrademe coalescence time is $\bar{T}_0 = 2N_e$, and the expected number of heterozygous nucleotide sites is $\bar{d}_0 = 4N_e u$, where N_e and u denote the migration effective population number (5) and the mutation rate per gene, respectively. If $N_e \gg 1$, these formulas are good approximations for diploid migration. Thus, for the simple functionals $\bar{T}_0$ and $\bar{d}_0$, population subdivision can be taken into account by replacing the actual total population number N_T by N_e. This reduction generally fails for higher moments of T_0 and d_0 and for measures of genetic variability other than $\bar{d}_0$, such as the expected number of segregating sites.

In contrast, the rate of gene substitution, K, is completely independent of population structure; unlike $\bar{T}_0$ and $\bar{d}_0$, it does not depend even on the stationary distribution ν and the deme proportions κ. To see this, note first that the fixation probability of a nucleotide with initial frequency p_i in deme i is its weighted average initial frequency,

$$P = \sum_i \nu_i p_i \quad (39)$$

(Nagylaki 1980). For a new mutant that appears in deme j, we have $p_j = 1/(2N_j)$ and $p_i = 0$ if $i \neq j$; so $P = \nu_j/(2N_j)$. Since a mutant appears in deme j with probability κ_j, the unconditional fixation probability is

$$\sum_j \kappa_j \nu_j/(2N_T \kappa_j) = 1/2N_T, \quad (40)$$

which yields Kimura's (1968) panmictic result,

$$K = 2N_T u/(2N_T) = u. \quad (41)$$

We close this note with some remarks on effective population number. Consult Caballero (1994) and Nagylaki (1995) for additional discussion and references.

First, it must be kept in mind that the introduction of an effective population number generally does not reduce exactly a complicated model to a simpler or ideal one. At most, such reduction occurs approximately or only for certain functionals of the evolutionary process. For example, the variance effective population number $N_e^{(v)}$ for a panmictic dioecious population is a parameter (rather than a random variable) only in the diffusion approximation, and it is only in this approximation that it reduces a dioecious model to a monoecious one (Ewens 1979, pp. 104–112, 1982; Nagylaki 1995). The inbreeding effective population number $N_e^{(i)}$ relates the asymptotic (*i.e.*, long-time) rate of decay of the probabilities of genetic identity (including the heterozygosity) in complicated models to their rate of decay in an ideal population.

Second, although effective population numbers have usually been defined in terms of some property of the evolutionary process, they are theoretically instructive and useful only if they can be evaluated as parameters, rather than random variables that depend on that process. This has been accomplished under a wide range of assumptions for both $N_e^{(v)}$ and $N_e^{(i)}$.

Third, a particular effective population number is useful only if it can be evaluated without analysis of the evolutionary process or if it predicts more than one property of that process. Again, $N_e^{(v)}$ and $N_e^{(i)}$ satisfy this criterion.

The migration effective population number N_e defined by (5) has all the desirable properties discussed above. Its evaluation from (5) is simple, explicit, and independent of the genetic model: N_e satisfies $N_e \leq N_T$ and depends only on the vector κ of demic proportions and on the unique stationary distribution ν of the Markov chain generated by the constant, ergodic backward migration matrix M. Of course, no effective population number can reduce a model of a subdivided population to that of a panmictic one. However, the above and earlier analyses (Nagylaki 1980, 1982, 1983, 1994) show that N_e replaces N_T in the strong-migration limit and in certain aspects of geographical invariance.

Finally, it should be noted that our definition of N_e differs from that of the various recently introduced effective population numbers for subdivided populations

(Nei and Takahata 1993; Wang 1997a,b; Whitlock and Barton 1997), which are defined in terms of the behavior of the evolutionary process.

I am very grateful to Magnus Nordborg for stimulating me to carry out this study by asking whether previous results were special cases of a general formula and for helpful comments on the manuscript. I thank Robert Griffiths for a discussion that led to the intuitive interpretation of the migration effective population number. This work was supported by National Science Foundation grant DEB-9706912.

LITERATURE CITED

Caballero, A., 1994 Developments in the prediction of effective population size. Heredity **73:** 657–679.

Caballero, A., and W. C. Hill, 1992 A note on the inbreeding effective population size. Evolution **46:** 1969–1972.

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

Ewens, W. J., 1974 A note on the sampling theory for infinite alleles and infinite sites models. Theoret. Pop. Biol. **6:** 143–148.

Ewens, W. J., 1979 *Mathematical Population Genetics*. Springer-Verlag, Berlin.

Ewens, W. J., 1982 On the concept of effective population size. Theoret. Pop. Biol. **21:** 373–378.

Feller, W., 1968 *An Introduction to Probability Theory and Its Applications*, Vol. I, Ed. 3. Wiley, New York.

Gantmacher, F. R., 1959 *The Theory of Matrices*, Vol. II. Chelsea, New York.

Griffiths, R. C., 1981 The number of heterozygous loci between two randomly chosen completely linked sequences of loci in two subdivided population models. J. Math. Biol. **12:** 251–261.

Herbots, H. M., 1994 Stochastic models in population genetics: genealogy and genetic differentiation in structured populations. Dissertation, University of London.

Hey, J., 1991 A multi-dimensional coalescent process applied to multi-allelic selection models and migration models. Theoret. Pop. Biol. **39:** 30–48.

Kimura, M., 1968 Evolutionary rate at the molecular level. Nature **217:** 624–626.

Kimura, M., 1969 The number of heterozygous nucleotide sites maintained in a finite population due to steady flux of mutations. Genetics **61:** 893–903.

Kimura, M., and J. F. Crow, 1964 The number of alleles that can be maintained in a finite population. Genetics **49:** 725–738.

Li, W.-H., 1976 Distribution of nucleotide differences between two randomly chosen cistrons in a subdivided population: the finite island model. Theoret. Pop. Biol. **10:** 303–308.

Malécot, G., 1946 La consanguinité dans une population limitée. C.R. Acad. Sci. Paris **22:** 841–843.

Malécot, G., 1948 *Les mathématiques de l'hérédité*. Masson, Paris.

Malécot, G., 1951 Un traitment stochastique des problèmes linéaires (mutation, linkage, migration) en Génétique de Population. Ann. Univ. Lyon Sci. Sec. A **14:** 79–117.

Nagylaki, T., 1980 The strong-migration limit in geographically structured populations. J. Math. Biol. **9:** 101–114.

Nagylaki, T., 1982 Geographical invariance in population genetics. J. Theoret. Biol. **99:** 159–172.

Nagylaki, T., 1983 The robustness of neutral models of geographical variation. Theoret. Pop. Biol. **24:** 268–294.

Nagylaki, T., 1985 Homozygosity, effective number of alleles, and interdeme differentiation in subdivided populations. Proc. Natl. Acad. Sci. USA **82:** 8611–8613.

Nagylaki, T., 1986 Neutral models of geographical variation, pp. 216–237 in *Stochastic Spatial Processes*, edited by P. Tautu. Springer, Berlin.

Nagylaki, T., 1992 *Introduction to Theoretical Population Genetics*. Springer-Verlag, Berlin.

Nagylaki, T., 1994 Geographical variation in a quantitative character. Genetics **136:** 361–381.

Nagylaki, T., 1995 The inbreeding effective population number in dioecious populations. Genetics **139:** 473–485.

Nath, H. B., and R. C. Griffiths, 1993 The coalescent in two colonies with symmetric migration. J. Math. Biol. **31:** 841–852.

Nei, M., and N. Takahata, 1993 Effective population size, genetic diversity, and coalescence time in subdivided populations. J. Mol. Evol. **37:** 240–244.

Nordborg, M., 1997 Structured coalescent processes on different time scales. Genetics **146:** 1501–1514.

Notohara, M., 1990 The coalescent and the genealogical process in geographically structured population. J. Math. Biol. **29:** 59–75.

Notohara, M., 1993 The strong-migration limit for the genealogical process in geographically structured populations. J. Math. Biol. **31:** 115–122.

Notohara, M., 1997 The number of segregating sites in a sample of DNA sequences from a geographically structured population. J. Math. Biol. **36:** 188–200.

Slatkin, M., 1987 The average number of sites separating DNA sequences drawn from a subdivided population. Theoret. Pop. Biol. **32:** 42–49.

Strobeck, C., 1987 Average number of nucleotide differences in a sample from a single subpopulation: a test for population subdivision. Genetics **117:** 149–153.

Tajima, F., 1989 DNA polymorphism in a subdivided population: the expected number of segregating sites in the two-subpopulation model. Genetics **123:** 229–240.

Takahata, N., 1988 The coalescent in two partially isolated diffusion populations. Genet. Res. **52:** 213–222.

Wang, J., 1997a Effective size and *F*-statistics of subdivided populations. I. Monoecious species with partial selfing. Genetics **146:** 1453–1463.

Wang, J., 1997b Effective size and *F*-statistics of subdivided populations. II. Dioecious species. Genetics **146:** 1465–1474.

Watterson, G. A., 1975 On the number of segregating sites in genetical models without recombination. Theoret. Pop. Biol. **7:** 256–276.

Whitlock, M. C., and N. H. Barton, 1997 The effective size of a subdivided population. Genetics **146:** 427–441.

Wright, S., 1948 Genetics of populations, pp. 111–112 in *Encyclopaedia Britannica*, Vol. 10, Ed. 14. Encyclopaedia Britannica, Chicago.

Communicating editor: M. Slatkin