Letter to the Editor

F_{ST} AND G_{ST} STATISTICS IN THE FINITE ISLAND MODEL

In WRIGHT's (1931, 1943) island model of population structure it is assumed that a population consists of an infinite number of subpopulations each of the same size N, and in each subpopulation the proportion m of the total gene pool is derived from immigrants that may be considered a random sample of the entire population. The extent of genetic differentiation of subpopulations is measured by the fixation index (F_{ST}), which is defined as $\sigma_x^2/[\bar{x}(1 - \bar{x})]$. Here, $\bar{x}$ and σ_x^2 are the mean and variance of gene frequency over all subpopulations. In the island model the equilibrium value of F_{ST} is given by

$$F_{ST} = 1 \left/ \left[1 + 2N \left\{ \frac{1}{(1 - m)^2} - 1 \right\} \right] \right. \tag{1}$$

This equation has been derived by considering a pair of alleles at a locus without mutation.

With the development of the infinite allele model of neutral mutations (KIMURA and CROW 1964), a number of authors have extended this model to the case of a finite number of subpopulations (*e.g.*, MARUYAMA 1970; LATTER 1973; NEI 1975; LI 1976; TAKAHATA 1983). In particular, NEI (1975) derived the expectation of his coefficient of gene differentiation (NEI 1973), which is denoted by G_{ST} and is defined as the ratio of the intersubpopulational gene diversity (D_{ST}) to the total gene diversity (H_T). G_{ST} is regarded as an extension of WRIGHT's F_{ST} to the case of multiple alleles, since it is a weighted mean of F_{ST} over all extant alleles. However, NEI's (1975) formula for G_{ST} is quite complicated, and its relationship with WRIGHT's formula in (1) is not clear. Recently TAKAHATA (1983) derived a simple equation for G_{ST} using diffusion approximations; yet, its relationship with WRIGHT's or NEI's formula is not immediately clear. The purpose of this note is to present a simplified form of NEI's exact formula and clarify its relationship with other related equations.

Before discussing this problem, however, we would like to indicate that in the case of the finite island model two different schemes of migration have been used in the literature, and this seems to have caused some confusion among readers. In the first scheme the genes entering into a subpopulation by migration are assumed to be a random sample of genes from the entire population. In this case a certain proportion of genes are expected to be returned to the subpopulation from which they originally came. Let x_i be the frequency of an allele in the ith subpopulation in a generation. If we disregard the effect

of genetic drift, the allele frequency, x'_i, in the next generation is then given by

$$x'_i = (1 - m)x_i + \frac{m}{s} \sum_{j=1}^{s} x_j, \tag{2}$$

where m is the migration rate $(0 \leq m \leq 1)$, and s is the number of subpopulations. This migration scheme has been used by MARUYAMA (1970), NEI (1975), LI (1976) and NEI, CHAKRAVARTI and TATENO (1977). In the second migration scheme the gene pool of the ith subpopulation is assumed to contain a proportion $1 - m_L$ of gametes derived from the same subpopulation in the previous generation and a proportion $m_L/(s - 1)$ of gametes derived from each of the remaining $s - 1$ subpopulations. Therefore, x'_i is given by

$$x'_i = (1 - m_L)x_i + \frac{m_L}{s - 1} \sum_{j \neq i}^{s} x_j, \tag{3}$$

where $0 \leq m_L \leq (s - 1)/s$. This scheme of migration has been used by LATTER (1973), TAKAHATA (1983) and Nagylaki (1983).

It should be noted that, although the mathematical formulations for the two migration schemes are different, they are actually interchangeable. Indeed, if we replace m_L by $m(s - 1)/s$ in (3) the second model becomes identical with the first model, as noted by LATTER (1973). NEI (1975) and NAGYLAKI (1983) both derived equations for the gene identities within and between subpopulations. Superficially, they look different, but if we note the relationship between m and m_L, they are identical. In a similar study LATTER (1973) considered a different set of quantities for the finite island model. His equations are naturally different from NEI's and NAGYLAKI's, but they can be converted into the latter by a proper transformation, as NAGYLAKI (1983) noted.

As mentioned earlier, NEI's formula for G_{ST} is quite complicated, but it is possible to reduce it in the following form.

$$G_{ST} = 1 \left/ \left[1 + 2N\left(\frac{s}{s - 1}\right) \left\{ \frac{1}{(1 - m)^2 (1 - v)^2} - 1 \right\} \right], \right. \tag{4}$$

which is approximately equal to

$$G_{ST} \approx 1 \left/ \left[1 + 4N\left(\frac{s}{s - 1}\right)(m + v) \right], \right. \tag{4a}$$

where $m, v \ll 1$. Here, v is the mutation rate per generation. When $s = \infty$ and $v = 0$, equation (4) becomes identical with WRIGHT's equation (1). This clearly indicates the equivalence between F_{ST} and G_{ST} (see also NEI 1977 for this equivalence). When s is finite and $v \neq 0$, (4) is not the same as (1), strictly speaking, but if s is sufficiently large and $m \gg v$, it is approximately the same as (1) or its simplified version $1/(1 + 4Nm)$ (WRIGHT 1951).

TAKAHATA's (1983) equation for G_{ST} is given in the following form.

$$G_{ST} = 1 \left/ \left[1 + 4N\left(\frac{s}{s - 1}\right)^2 \left(m_L + \frac{s - 1}{s} v\right) \right]. \right. \tag{5}$$

TABLE 1

G_{ST} values obtained from equations (4) and (6) for various values of population size (N),
migration rate (m) and number of subpopulations (s)

s	Equation no.	m					
		0.01	0.1	0.5	0.001	0.005	0.01
		N = 5			N = 500		
2	(4)	0.711	0.176	0.016	0.200	0.047	0.024
	(6)	0.819	0.212	0.018	0.236	0.050	0.025
10	(4)	0.816	0.277	0.029	0.310	0.082	0.042
	(6)	0.829	0.285	0.030	0.318	0.083	0.043
10^2	(4)	0.830	0.297	0.032	0.331	0.089	0.046
	(6)	0.831	0.298	0.032	0.332	0.090	0.047
10^3	(4)	0.831	0.299	0.032	0.333	0.090	0.047
	(6)	0.831	0.300	0.032	0.333	0.090	0.047

The mutation rate used is 10^{-6}.

If we replace m_L by $m(s - 1)/s$, this becomes identical with (4a). Thus, NEI's and TAKAHATA's equations are essentially the same.

Equation (4) is based on the assumption that the population is in equilibrium with respect to the effects of mutation, migration and genetic drift. For this equilibrium to be attained, however, it requires a large number of generations—often of the order of the reciprocal of the mutation rate. Since the migration pattern of natural populations almost never remains the same for a long time, the applicability of (4) may be questioned. Fortunately, however, F_{ST} or G_{ST} reaches the equilibrium value relatively quickly for polymorphic loci (NEI, CHAKRAVARTI and TATENO 1977). A similar observation was also made by J. F. CROW (personal communication) independently. In the finite island model all polymorphic loci eventually become monomorphic in the absence of mutation. Yet, G_{ST} reaches an equilibrium value, and this equilibrium value is given by

$$G_{ST} = 1 - 2Ns(1 - \lambda), \tag{6}$$

where $1 - \lambda$ is the rate of change of total gene diversity (H_T) per generation. Equation (6) is valid for any number of alleles, since it is derived from the relationship between D_{ST} and H_T. It is also valid for any type of migration.

In the case of the finite island model the exact value of λ is known (LI 1976), so that we can compare (6) and (4). Table 1 shows the numerical values of G_{ST}'s given by (4) and (6). It is clear that (4) and (6) give virtually the same values for most parameter values examined. This indicates that for practical purposes (4) can be used even for transient populations if H_T reaches the stage of steady decay. This result suggests that (6) can be expressed in a form similar to (4). In practice, however, λ includes a square root term, and it is not simple to get such an equation.

Finally, it should be mentioned that G_{ST} in this note refers to the ratio of

the *expected* intersubpopulational gene diversity to the *expected* total gene diversity. Therefore, it should be estimated by using the average gene diversities for many loci (see NEI 1977). Theoretically, G_{ST} can be computed by using single-locus values of D_{ST} and H_T. However, such a single-locus G_{ST} is subject to a large stochastic error, and the mean of single-locus G_{ST}'s for many loci is not necessarily a good estimate of the G_{ST} discussed in this note (NEI, CHAKRAVARTI and TATENO 1977; TAKAHATA 1983).

LITERATURE CITED

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

LATTER, B. D. H., 1973   The island model of population differentiation: a general solution. Genetics **73**: 147–157.

LI, W.-H., 1976   Effect of migration on genetic distance. Am. Nat. **110**: 841–847.

MARUYAMA, T., 1970   Effective number of alleles in a subdivided population. Theor. Pop. Biol. **23**: 216–240.

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

NEI, M., 1973   Analysis of gene diversity in subdivided populations. Proc. Natl. Acad. Sci. USA **70**: 3321–3323.

NEI, M., 1975   *Molecular Population Genetics and Evolution.* North-Holland, New York.

NEI, M., 1977   *F*-statistics and analysis of gene diversity in subdivided populations. Ann. Hum. Genet. **41**: 225–233.

NEI, M., A. CHAKRAVARTI and Y. TATENO, 1977   Mean and variance of F_{ST} in a finite number of incompletely isolated populations. Theor. Pop. Biol. **11**: 291–306.

TAKAHATA, N., 1983   Gene identity and genetic differentiation of populations in the finite island model. Genetics **104**: 497–512.

WRIGHT, S., 1931   Evolution in Mendelian populations. Genetics **16**: 97–159.

WRIGHT, S., 1943   Isolation by distance. Genetics **28**: 114–138.

WRIGHT, S., 1951   The genetical structure of populations. Ann. Eugen. **15**: 323–354.

NAOYUKI TAKAHATA and MASATOSHI NEI
Center for Demographic and Population Genetics
The University of Texas Health Science Center at Houston
P.O. Box 20334
Houston, Texas 77225