Equilibrium Values of Measures of Population Subdivision for Stepwise Mutation Processes

François Rousset

Laboratoire Génétique et Environnement, Institut des Sciences de l'Évolution, Université de Montpellier II, 34095 Montpellier, France

Manuscript received September 26, 1995
Accepted for publication November 29, 1995

ABSTRACT

Expected values of WRIGHT's *F*-statistics are functions of probabilities of identity in state. These values may be quite different under an infinite allele model and under stepwise mutation processes such as those occurring at microsatellite loci. However, a relationship between the probability of identity in state in stepwise mutation models and the distribution of coalescence times can be deduced from the relationship between probabilities of identity by descent and the distribution of coalescence times. The values of F_{IS} and F_{ST} can be computed using this property. Examination of the conditional probability of identity in state given some coalescence time and of the distribution of coalescence times are also useful for explaining the properties of F_{IS} and F_{ST} at high mutation rate loci, as shown here in an island model of population structure.

The estimation of WRIGHT's *F*-statistics is standard practice (*e.g.*, WEIR 1990) and has multiple uses (*e.g.*, HARTL and CLARK 1989; SLATKIN 1993). Most discussions of their properties have been based on the values of probabilities of identity by descent in the infinite allele model. In a recent paper SLATKIN (1995) has emphasized that such theory may not be appropriate for the analysis of loci evolving under a stepwise mutation process with a high mutation rate, taking the example of the estimation of the number of migrants in an island model.

In its simplest form the stepwise mutation model assumes that mutation acts by increasing or decreasing the electrophoretic mobility of an allele by one unit. This model was initially considered in charge state models of protein polymorphism (OHTA and KIMURA 1973; WEHRHAHN 1975) and has been reconsidered as a model of evolution of microsatellite allele sizes (SHRIVER *et al.* 1993; VALDES *et al.* 1993). Microsatellite loci have high mutation rates and there is some evidence from pedigree data for a stepwise pattern of mutation (WEBER and WONG 1993). However the exact nature of the mutation process is still not clear (ESTOUP *et al.* 1995a,b), and variants of the stepwise model including changes by more than one unit must be considered (DI RIENZO *et al.* 1994).

In such models alleles can be identical in state without being identical by descent. It is then appropriate to define *F*-statistics as intraclass correlations for the probability of identity in state ("IIS correlations"), *i.e.*,

$$F_{IS} \equiv \frac{Q_1 - Q_2}{1 - Q_2},$$

$$F_{ST} \equiv \frac{Q_2 - Q_3}{1 - Q_3}, \qquad (1)$$

where the Q_j are probabilities of identity in state, Q_1 for pairs of genes within individuals, Q_2 between individuals within subpopulations, and Q_3 between subpopulations (COCKERHAM and WEIR 1987). Throughout this paper the standard notation $\equiv$ is used to distinguish the definition of parameters from their values under particular models, and the j indices (j = 1 to 3) will always refer to the same pairs of genes.

Infinite allele models provide the value of the probabilities of identity by descent θ_j for the different pairs of genes, and correlations for the probability of identity by descent (IBD correlations) are defined in the same way: $\beta_{IS} \equiv (\theta_1 - \theta_2)/(1 - \theta_2)$ and $\beta_{ST} \equiv (\theta_2 - \theta_3)/(1 - \theta_3)$. In this paper, probabilities of identity in state are given for stepwise mutation models (SMMs). The values of F_{ST} and F_{IS} in this case are compared to their values in the infinite allele model (IAM) and in a simple *k*-allele model (KAM), and some explanations are provided for the properties of *F*-statistics under the different mutation models.

EQUILIBRIUM VALUES IN AN ISLAND MODEL

As an example, I consider an island model with gametic migration. n subpopulations of N diploid individuals exchange migrant gametes at rate m. The following quantities are useful (NAGYLAKI 1983): after migration, the frequency of pairs of genes within a subpopulation that originate from one subpopulation in the previous generation is $a = (1 - m)^2 + m^2/(n - 1)$, and the frequency of pairs of genes from different subpopula-

Address for correspondence: Laboratoire Génétique et Environnement, Institut des Sciences de l'Evolution, CC065, USTL, Place E. Bataillon, 34095 Montpellier Cedex 05, France.
E-mail: rousset@isem.univ-montp2.fr

tions that also originate from one subpopulation in the previous generation is $b = (1 - a)/(n - 1)$. The mutation rate is u for all alleles. Let $\gamma \equiv (1 - u)^2$. Some offspring are produced by selfing in each subpopulation: if s is the conditional probability of selfing given both genes are copies of genes from one subpopulation in the previous generation, the overall amount of selfing is as. In a random mating hermaphroditic subpopulation, $s = 1/N$. The notation $\sigma \equiv s - 1/N$ will be used.

Identity by descent (infinite allele model): Recursions for the IBD probabilities follow from the definitions:

$$\theta_{1,t+1} = \gamma\left[a\left(s\frac{1+\theta_{1,t}}{2} + (1-s)\theta_{2,t}\right) + (1-a)\theta_{3,t}\right],$$

$$\theta_{2,t+1} = \gamma\left[a\left(\frac{1}{N}\frac{1+\theta_{1,t}}{2} + \left(1-\frac{1}{N}\right)\theta_{2,t}\right) + (1-a)\theta_{3,t}\right],$$

$$\theta_{3,t+1} = \gamma \times \left[b\left(\frac{1}{N}\frac{1+\theta_{1,t}}{2} + \left(1-\frac{1}{N}\right)\theta_{2,t}\right) + (1-b)\theta_{3,t}\right]. \quad (2)$$

This is a special case of the model considered by MARUYAMA and TACHIDA (1992). The equilibrium value of the IBD correlations are

$$F_{IS} = \beta_{IS} = \frac{\sigma a\gamma}{2 - \sigma a\gamma},$$

$$F_{ST} = \beta_{ST} = \frac{\gamma d}{\gamma d + N(2 - \sigma a\gamma)(1 - \gamma d)}, \quad (3)$$

where $d = a - b = (1 - m[n/(n-1)])^2$.

Symmetric k allele model: If there is a finite number k of possible alleles and each allele can mutate to another one at rate $u/(k-1)$, copies at generation $t+1$ of pairs of genes that were identical in state at generation t are identical at generation $t+1$ with probability $v = (1-u)^2 + u^2/(k-1)$, and pairs of genes that were different can become identical at rate $(1-v)/(k-1)$ (CROW and AOKI 1984). Hence

$$Q_{1,t+1} = a\left(s\frac{v+Q'_{1,t}}{2} + (1-s)Q'_{2,t}\right) + (1-a)Q'_{3,t},$$

$$Q_{2,t+1} = a\left(\frac{1}{N}\frac{v+Q'_{1,t}}{2} + \left(1-\frac{1}{N}\right)Q'_{2,t}\right) + (1-a)Q'_{3,t},$$

$$Q_{3,t+1} = b \times \left(\frac{1}{N}\frac{v+Q'_{1,t}}{2} + \left(1-\frac{1}{N}\right)Q'_{2,t}\right) + (1-b)Q'_{3,t}, \quad (4)$$

in which $Q'_{j,t}$ is the conditional IIS probability of a pair of genes given the type j of pair of genes sampled

$$Q'_{j,t} = vQ_{j,t} + (1 - Q_{j,t})(1-v)/(k-1) = \gamma' Q_{j,t} + (1-\gamma')/k, \quad (5)$$

where $\gamma' \equiv v - (1-v)/(k-1) = (1 - uk/(k-1))^2$. At equilibrium the IIS correlations have the same expression as the IBD correlations (3), with γ' instead of γ.

$$F_{IS} = \frac{\sigma a\gamma'}{2 - \sigma a\gamma'},$$

$$F_{ST} = \frac{\gamma' d}{\gamma' d + N(2 - \sigma a\gamma')(1 - \gamma' d)}. \quad (6)$$

This model is therefore equivalent to an infinite allele model with a higher mutation rate, $u' = ku/(k-1)$.

Stepwise mutation model: It is useful to consider the generating function of the probabilities p_k that a randomly chosen gene differs by k steps from another randomly chosen gene. In a subdivided population, these probabilities $p_{k,j}$ differ for the different pairs of genes, so we consider the generating functions $\psi_j(z) \equiv \Sigma_{-\infty}^{+\infty} p_{k,j} z^k$. Pairs of genes can be sampled symmetrically in two ways, therefore $p_{-k,j} = p_{k,j}$ and $\psi_j(e^{ix}) = p_{0,j} + 2\Sigma_1^{+\infty} p_{k,j}\cos(kx)$ (MORAN 1975) from which individual probabilities $p_{k,j}$ can be extracted by integration of $\psi_j(e^{ix})\cos(kx)$. Thus, IIS probabilities are

$$Q_j = p_{0,j} = \frac{1}{\pi}\int_0^\pi \psi_j(e^{ix})\,dx. \quad (7)$$

Mutation occurs at rate u as above. According to the one-step mutation model, it increases the size difference of a pair of genes with regard to the size difference of their ancestors in the previous generation by +1 with probability $\approx u$, by −1 with probability $\approx u$, and will not change it with probability $\approx 1 - 2u$. Hence the effect of mutation is to change the generating function $\psi(z)$ by a factor $r(z) \approx 1 - 2u + uz + u/z$ (MORAN 1975; WEHRHAHN 1975). The exact expression is $r(z) = [1 - (2 - z - 1/z)u/2]^2$, the square of the effect of mutation on each allele.

The results given below are valid whenever changes in allele size due to mutation are independent of the nature of the allele. A two-phase mutation model was proposed by DI RIENZO *et al.* (1994) to take into account the fact that some mutations may increase or decrease microsatellite size by more than one repeat. In this model with probability p, mutation increases or decreases allele size difference by one repeat, and with probability $1 - p$ it increases or decreases allele size difference by k repeats, where k follows some probability distribution. DI RIENZO *et al.* (1994) considered a truncated geometric distribution where $\Pr(k) = (1-q)q^{k-1}$ for $k \geq 1$. In this case

$$r(z) = \left[1 - u + \frac{1}{2}\left(up(z + z^{-1}) + u(1 - p)\frac{1 - q}{q}\sum_{k=1}^{\infty} q^k(z^k + z^{-k})\right)\right]^2$$

$$= \left[1 - u + \frac{1}{2}\left(up(z + z^{-1}) + u(1 - p)(1 - q)\left(\frac{z}{1 - qz} + \frac{z^{-1}}{1 - qz^{-1}}\right)\right)\right]^2, \quad (8)$$

so that

$$r(e^{ix}) = \left[1 - u + up\cos(x) + u(1 - p) \times \frac{(1 - q)(\cos(x) - q)}{1 - 2q\cos(x) + q^2}\right]^2.$$

In all cases the recursions are

$$\psi_{1,t+1} = r(z) \times \left[a\left(s\frac{1 + \psi_{1,t}}{2} + (1 - s)\psi_{2,t}\right) + (1 - a)\psi_{3,t}\right],$$

$$\psi_{2,t+1} = r(z) \times \left[a\left(\frac{1}{N}\frac{1 + \psi_{1,t}}{2} + \left(1 - \frac{1}{N}\right)\psi_{2,t}\right) + (1 - a)\psi_{3,t}\right],$$

$$\psi_{3,t+1} = r(z) \times \left[b\left(\frac{1}{N}\frac{1 + \psi_{1,t}}{2} + \left(1 - \frac{1}{N}\right)\psi_{2,t}\right) + (1 - b)\psi_{3,t}\right]. \quad (9)$$

These recursions are mathematically identical to those for IBD. At equilibrium the generating functions are

$$\psi_1(z) = \frac{r(z)(a - dr(z)) + aN\sigma(1 - r(z))(1 - dr(z))}{V(z)},$$

$$\psi_2(z) = \frac{r(z)(a - dr(z))}{V(z)},$$

$$\psi_3(z) = \frac{br(z)}{V(z)}, \quad (10)$$

where $V(z) = r(z)(a - dr(z)) + N(1 - r(z))(1 - dr(z))(2 - a\sigma r(z))$.

There is no simple relationship between N, m, σ, u and the Qs, so that the Fs need to be evaluated by numerical methods. However, SLATKIN (1995) has shown that the variances of the distributions of size difference, S_j, are proportional to the average coalescence times T_j of the respective pairs of genes, which themselves have simple relationships to the different parameters describing population structure. This is confirmed by computation of these variances, $S_j \equiv E[k_j^2] = (d^2\psi_j(z)/dz^2)|_{z=1}$ (since $E[k] = 0$). Thus

$$S_1 = 2uT_1 = \frac{4Nu(1 - d)(1 - a\sigma)}{b} = 4Nun(1 - a\sigma),$$

$$S_2 = 2uT_2 = S_1 + \frac{2Nua(1 - d)\sigma}{b} = S_1 + 2Nuna\sigma,$$

$$S_3 = 2uT_3 = S_2 + \frac{2ud}{b}. \quad (11)$$

Quantities analogous to F-statistics are

$$\rho_{IS} \equiv \frac{S_2 - S_1}{S_2} \quad \text{and} \quad \rho_{ST} \equiv \frac{S_3 - S_2}{S_3}. \quad (12)$$

In the stepwise mutation model the values of the ρs are the values of the ratios of average coalescence times, $C_{IS} \equiv (T_2 - T_1)/T_2$ and $C_{ST} \equiv (T_3 - T_2)/T_3$. They have the same expression as the values of the βs (3), with 1 instead of γ. Hence $2\rho_{IS}/(1 + \rho_{IS}) = a\sigma$ and $\rho_{ST} \approx [1 + 4\alpha Nm(1 - \sigma/2)]^{-1}$ for m small, where $\alpha \equiv n/(n - 1)$.

Table 1 shows values of the functions $\dot{\sigma}$ and $\dot{M}$ defined as

$$\dot{\sigma}(F_{IS}) \equiv \frac{2F_{IS}}{a(1 + F_{IS})} \quad \text{and}$$

$$\dot{M}(F_{ST}) \equiv \frac{1}{4\alpha(1 - \sigma/2)}\left(\frac{1}{F_{ST}} - 1\right), \quad (13)$$

evaluated for the different mutation processes and different values of m and σ. These definitions are chosen only for ease of comparison to common reference values that are $\dot{\sigma}(C_{IS}) = \sigma$ and $\dot{M}(C_{ST}) = Nm$. For the infinite allele model $\dot{M}(F_{ST})$ is computed using the value of F_{ST} according to (3), for IIS in the k-allele model according to (6), and for IIS in the stepwise mutation model according to (7) and (10).

$\dot{M}$ is larger than Nm, the relative difference between them increasing with decreasing values of Nm, and decreasing with decreasing mutation rate. These results are in agreement with SLATKIN's (1995) simulation results for the SMM. However, $\dot{M}$ is larger and differs more from Nm in the IAM and is still larger in the KAM. The discrepancies between σ and $\dot{\sigma}$ are very small.

RELATIONSHIP TO COALESCENCE TIMES

The island model with gametic migration illustrates some relationships between IBD, IIS in the k-allele model, and generating functions in stepwise mutation models. These results are a simple consequence of the fact that, given some coalescence time t, the conditional IBD probability is γ^t, the IIS probability in the KAM is $[1 + (k - 1)\gamma'^t]/k$ [from (5)], and the generating function of allele size differences is $r(z)^t$. If c_t is the probability of coalescence of a pair of genes at time t

TABLE 1

Examples of values of $\dot{M}$ and $\dot{\sigma}$

Mutation process	$2Nu$	σ	Parameter	Nm 1000	100	10	1	0.1	0.01
KAM, $k = 5$	10	0	$\dot{M}$	1182	108	16.2	7.19	6.29	6.20
IAM	10	0	$\dot{M}$	1181	107	15	5.95	5.05	4.96
One-step	10	0	$\dot{M}$	1181	107	14.2	3.59	1.71	1.40
	10	0	$\dot{\sigma}$	0	0	0	0	0	0
	10	0.5	$\dot{M}$	1256	107	14.3	3.65	1.85	1.56
	10	0.5	$\dot{\sigma}$	0.499	0.499	0.499	0.499	0.499	0.499
	10	0.99	$\dot{M}$	1401	109	14.3	3.76	2.07	1.82
	10	0.99	$\dot{\sigma}$	0.989	0.989	0.989	0.989	0.989	0.989
KAM, $k = 5$	0.755	0	$\dot{M}$	1175	102	10.5	1.47	0.57	0.48
IAM	0.755	0	$\dot{M}$	1175	102	10.4	1.37	0.47	0.38
One-step	0.755	0	$\dot{M}$	1175	102	10.4	1.35	0.38	0.27
Two-phase, $q = 0.8682$	0.755	0	$\dot{M}$	1175	102	10.4	1.35	0.41	0.31
	0.755	0	$\dot{\sigma}$	0	0	0	0	0	0
	0.755	0.5	$\dot{M}$	1250	103	10.4	1.35	0.42	0.32
	0.755	0.5	$\dot{\sigma}$	0.500	0.500	0.500	0.500	0.500	0.500
	0.755	0.99	$\dot{M}$	1394	104	10.4	1.36	0.43	0.33
	0.755	0.99	$\dot{\sigma}$	0.990	0.990	0.990	0.990	0.990	0.990

They were computed as described in text, using *Mathematica* (WOLFRAM 1991). $n = 100$ and $2N = 20{,}000$ in all cases. In the two-phase model, with probability $1 - p = 0.25$ the increment in allele size is assumed to follow a geometric distribution. The values of u and q are chosen so that the variance of increment in allele size per generation due to the second phase mutations is 50 (SLATKIN 1995).

and $f(x) \equiv \Sigma c_t x^t$ is the generating function of coalescence times, then the IBD probability is $\Sigma c_t \gamma^t = f(\gamma)$ (SLATKIN 1991), the IIS probability in the KAM is $\Sigma c_t[1 + (k-1)\gamma'^t]/k = [1 + (k-1)f(\gamma')]/k$, and the generating function of allele size difference is $\Sigma c_t (r(z))^t = f(r(z))$. Then, if the IBD correlation is expressed as some function $g(\gamma)$, the IIS correlation in the KAM is $g(\gamma')$, and the ratio of coalescence times is $g(1)$.

TACHIDA (1985) has shown that the study of genealogical process and of the mutation process can be "separated" when the mutation process is a Markov chain. The reason is that any conditional IIS probability given the coalescence time t can be written as $\Sigma e_i \lambda_i^t$ where the λ_i are the eigenvalues of the transition matrix of the Markov chain, so that the unconditional IIS probability is $\Sigma e_i f(\lambda_i)$. The symmetric k-allele model is a simple illustration of this fact. It is also possible to analyze stepwise mutation processes by such an approach (MARUYAMA 1977).

Existing theory for IBD in various population structures (*e.g.*, MARUYAMA 1970; NAGYLAKI 1983; MARUYAMA and TACHIDA 1992) actually provides the generating functions of coalescence times (SLATKIN 1991) and can immediately be used for obtaining the equilibrium value of F_{ST} in stepwise mutation models and to compare them to ratios of average coalescence times that reflect the consequences of various factors of interest such as inbreeding, population size and migration patterns but not mutation. As another example consider a one-dimensional circular stepping-stone of d subpopulations. Each subpopulation receives $m/2$ immigrants from each of their two neighbors. The generating functions for pair of genes sampled in subpopulations distant by l steps are the solution of $\Psi = r(z)\mathbf{M}(\Psi + (1 - \psi_0)\mathbf{x}/2N)$ where $\mathbf{M}$ is a matrix describing the effects of migration and $\mathbf{x} = (1, 0, 0, \ldots, 0)^t$ (MARUYAMA 1970). The solution is

$$\psi_l(z) = \sum_{i=0}^{[(d+1)/2]} a_i \cos(2\pi i l/d), \qquad (14)$$

where $a_i = [\lambda_i(1 - r(z))a_0]/[(1 - r(z)\lambda_i)\Delta_i]$, $a_0 = r(z)/((1 - r(z))(2Nd + \Sigma_k r(z)\lambda_k/[(1 - r(z)\lambda_k)\Delta_k]))$, $\lambda_i = (1 - m(1 - \cos(2\pi i/d)))^2$, and $\Delta_i = 1$ if $i = 0$ or $i = k/2$, and $1/2$ otherwise. If the variances V_l of the distributions of differences between genes l locations apart are proportional to coalescence times, the following approximation holds for $\rho_{ST}(l) \equiv (V_l - V_0)/V_l$, assuming a small migration rate (SLATKIN 1993):

$$M(\rho_{ST}(l)) \equiv \frac{1}{4}\left(\frac{1}{\rho_{ST}(l)} - 1\right) \approx Nmd/(l(d-l)), \qquad (15)$$

but this approximation is not valid in general for $F_{ST}(l) \equiv (Q_0 - Q_l)/(1 - Q_l)$. $M(F_{ST}) \equiv 1/4(1/F_{ST} - 1)$ decreases much less than $M(\rho_{ST})$ in the more distant subpopulations (Figure 1). A similar lack of variation is observed for coalescence measures in some nonequilibrium models (SLATKIN 1993).

DISCUSSION

Effects of the mutation process on *F*-statistics: For the mutation process to matter, it is necessary that with

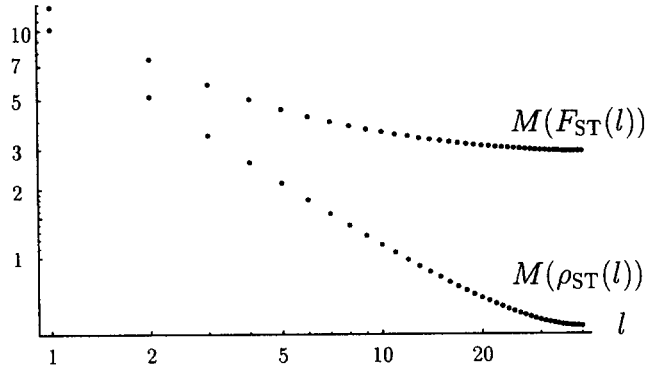

FIGURE 1.—$M(F_{ST})$ *vs.* $M(\rho_{ST})$ in a circular stepping stone. Note the logarithmic scale. l is the number of steps two demes are apart. The stepping stone has 80 subpopulations, subpopulation size is $N = 10^4$ and mutation follow the one-step mutation model with $u = 10^{-4}$.

high probability, coalescence times of genes between subpopulations are sufficiently long for two or more mutations events to occur. This is well illustrated by the stepping stone model in which the more distant the subpopulations, the longer the coalescence times and the more F_{ST} differs from the ratio of coalescence times C_{ST}. ρ_{ST} will also differ from C_{ST} when coalescence times get longer and the mutation process is not independent of the nature of the allele (BOWCOCK *et al.*, 1994; DEKA *et al.*, 1994; GARZA *et al.*, 1995; GOLDSTEIN *et al.*, 1995).

However, long coalescence times are not sufficient for the mutation process to matter. The fact that F_{IS} is virtually insensitive to this process is not due to the fact that most coalescence events for pairs of genes within subpopulations occur before the occurrence of two or more mutations, because they do not. For example in the case $Nm = 1$, $\sigma = 0.5$, $2N = 20000$ and $n = 100$, the probability of coalescence within 10^4 generations is only 0.49 for pairs within individuals and 0.23 for pairs between individuals within subpopulations (Figure 2). Yet σ is very close to $\dot{\sigma}$ in this case.

An examination of the distribution of coalescence times shows that a fraction $\approx x/(1-x)$ (where $x \equiv a\sigma/2$) of pairs of genes within individuals, and almost none between individuals, coalesce in the first generations in which case they are almost always identical. Thereafter the respective probabilities of coalescence at time t, $c_{1,t}$ and $c_{2,t}$ are proportional: $c_{1,t} \approx c_{2,t}(1-2x)/(1-x)$ (Figure 2). This is sufficient to ensure that $Q_1 \approx [x + (1-2x)Q_2]/(1-x)$ and that $F_{IS} \approx x/(1-x)$ is virtually independent of the mutation process. By the same argument, $S_1 \approx (1-2x)S_2/(1-x)$ so that $\rho_{IS} \approx x/(1-x)$ even if the mutation process is not stepwise. In contrast, it takes ~30,000 generations for $c_{3,t}$ to become proportional to $c_{2,t}$. The mutational processes occurring within this time affect the values of F_{ST} and ρ_{ST}. This time interval is longer when the migration rate decreases, so that F_{ST} is always performing worse when m decreases.

There are two aspects of the mutation process that

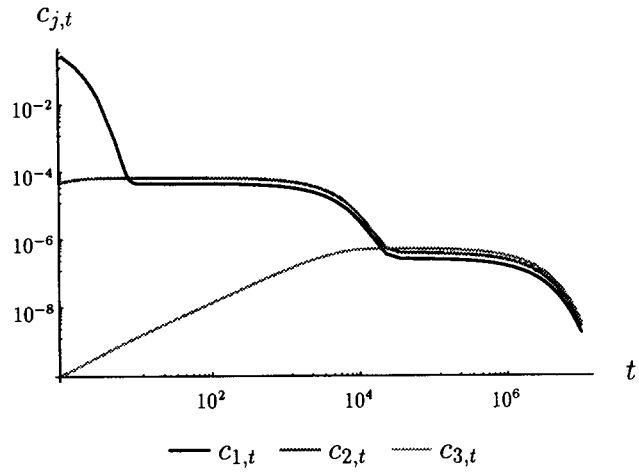

FIGURE 2.—Distributions of coalescence times. Note the log or log-log scale. The probabilities $c_{j,t}$ of coalescence time t and their cumulative distributions are given for the different pairs of genes in the case $Nm = 1$, $\sigma = 0.5$, $n = 100$ and $2N = 20000$.

may cause deviation from the ratios of coalescence times (SLATKIN 1995): high mutation rates and the nature of the mutations. When the probability of identity decreases linearly with time the Fs are identical to the ratios of coalescence times, but when multiple mutations occur the probabilities of identity are no longer linearly decreasing with time. Figure 3 shows that the deviation from linearity is more pronounced in the IAM than in the SMM, and even more in the KAM, as assessed by the convexity of the curves. As a consequence F_{ST} deviates more from C_{ST} in the IAM than under a stepwise process, and F_{ST} is a poorer measure of population subdivision if mutation is not stepwise. For a given value of the mutation rate, F_{ST} is closer to C_{ST} in the one-step than in the two-phase SMM (Table 1). Hence F_{ST} is closer to C_{ST} in the two-phase model with $2Nu = 0.755$ than in the one-step model with $2Nu = 10$ only because the mutation rate is lower, not because the mutation process is less stepwise.

Genes that are identical in state but not identical by descent (homoplasy) can be produced by the k allele or stepwise mutation processes. $\dot{M}$ is lower in the SMMs

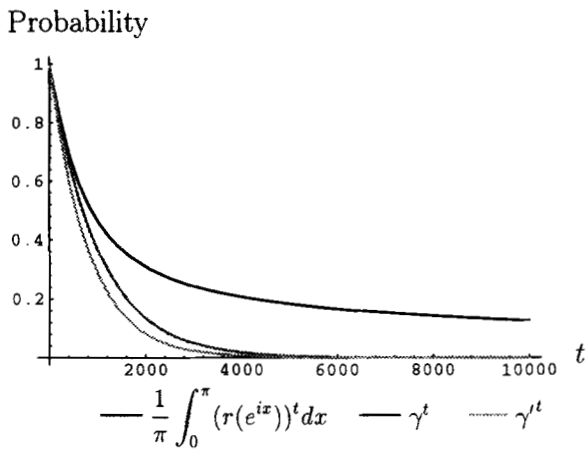

FIGURE 3.—Comparison of probabilities of identity. For the present comparison $[1 + (k-1)\gamma'^t]/k$ is equivalent to γ'^t. $u = {}^1/_{2000}$, $k = 5$, and $r(e^{ix}) = [1 - u(1 - \cos(x))]^2$ (one-step mutation model).

than in the IAM, while it is larger in the KAM. This shows that there is no simple effect of homoplasy on F_{ST}.

Estimation and relationship to SLATKIN's R_{ST}: If unbiased estimates $\hat{S}_2$ and $\hat{S}_3$ are obtained then a simple estimator of ρ_{ST} is $\hat{\rho}_{ST} \equiv (\hat{S}_3 - \hat{S}_2)/\hat{S}_3$. Likewise ρ_{IS} can be estimated by $\hat{\rho}_{IS} \equiv (\hat{S}_2 - \hat{S}_1)/\hat{S}_2$ (R. STREIFF, unpublished DEA thesis, Montpellier). SLATKIN (1995) considered the weighted sum

$$\hat{\bar{S}} = \frac{2s_s - 1}{2s_s n_s - 1}\hat{S}_2 + \frac{2s_s(n_s - 1)}{2s_s n_s - 1}\hat{S}_3, \qquad (16)$$

where n_s is the sample number and s_s the sample size. Thus

$$R_{ST} = \frac{\hat{\bar{S}} - \hat{S}_2}{\hat{\bar{S}}} = \frac{(1-c)\hat{\rho}_{ST}}{1 - c\hat{\rho}_{ST}}, \qquad (17)$$

where $c \equiv (2s_s - 1)/(2s_s n_s - 1)$. The relation between R_{ST} and $\hat{\rho}_{ST}$ is similar to that between "$\hat{G}_{CA}$" and "$\hat{\beta}$" in COCKERHAM and WEIR (1993).

MICHALAKIS and EXCOFFIER (1996) have defined an analysis of variance (ANOVA) estimator of ρ_{ST} by analogy to the definition of the estimator $\hat{\theta}$ of F_{ST} (WEIR and COCKERHAM 1984). There are several possible ANOVA estimators of ρ_{ST} that are all based on unbiased estimates of the S_j but differ by the weights given to samples of different sizes. They should provide different estimates only when sample sizes are variable as for the ANOVA estimators of F_{ST} (COCKERHAM 1973). Their small sample properties remain to be investigated.

I thank P. JARNE, M. RAYMOND, M. SLATKIN and F. VIARD for discussion, and T. GUILLEMAUD for his patience. This work was supported by the Programme Environnement du Centre National de la Recherche Scientifique (GDR 11.05). This is paper 96-004 of the Institut des Sciences de l'Évolution.

LITERATURE CITED

BOWCOCK, A. M., A. RUIZ-LINARES, J. TOMFOHRDE, E. MINCH, and J. R. KIDD *et al.*, 1994 High resolution of human evolutionary trees with polymorphic microsatellites. Nature **368:** 455–457.

COCKERHAM, C. C., 1973 Analyses of gene frequencies. Genetics **74:** 679–700.

COCKERHAM, C. C., and B. S. WEIR, 1987 Correlations, descent measures: drift with migration and mutation. Proc. Natl. Acad. Sci. USA **84:** 8512–8514.

COCKERHAM, C. C., and B. S. WEIR, 1993 Estimation of gene flow from *F*-statistics. Evolution **47:** 855–863.

CROW, J. F., and K. AOKI, 1984 Group selection for a polygenic behavioural trait: estimating the degree of population subdivision. Proc. Natl. Acad. Sci. USA **81:** 6073–6077.

DEKA, R., M. D. SHRIVER, L. M. YU, L. JIN, C. E. ASTON *et al.*, 1994 Conservation of human chromosome 13 polymorphic microsatellite $(CA)_n$ repeats in chimpanzees. Genomics **22:** 226–230.

DI RIENZO, A., A. C. PETERSON, J. C. GARZA, A. M. VALDES, M. SLATKIN *et al.*, 1994 Mutational processes of simple-sequence repeat loci in human populations. Proc. Natl. Acad. Sci. USA **91:** 3166–3170.

ESTOUP, A., L. GARNERY, M. SOLIGNAC and J.-M. CORNUET, 1995a Microsatellite variation in honey bee (*Apis mellifera* L.) populations: hierarchical genetic structure and test of the infinite allele and stepwise mutation models. Genetics **140:** 679–695.

ESTOUP, A., C. TAILLIEZ, J.-M. CORNUET and M. SOLIGNAC, 1995b Size homoplasy and mutational processes of interrupted microsatellites in two bee species, *Apis mellifera* and *Bombus terrestris* (Apidae). Mol. Biol. Evol. **12:** 1074–1084.

GARZA, J. C., M. SLATKIN and N. B. FREIMER, 1995 Microsatellite allele frequencies in humans and chimpanzees, with implications for constraints on allele size. Mol. Biol. Evol. **12:** 594–603.

GOLDSTEIN, D. B., A. R. LINARES, L. L. CAVALLI-SFORZA and M. W. FELDMAN, 1995 An evaluation of genetic distances for use with microsatellite loci. Genetics **139:** 463–471.

HARTL, D. L., and A. G. CLARK, 1989 *Principles of Population Genetics.* Sinauer, Sunderland, MA.

MARUYAMA, K., and H. TACHIDA, 1992 Genetic variability and geographical structure in partially selfing populations. Jpn. J. Genet. **67:** 39–51.

MARUYAMA, T., 1970 Effective number of alleles in a subdivided population. Theor. Pop. Biol. **1:** 273–306.

MARUYAMA, T., 1977 *Stochastic Problems in Population Genetics.* Springer Verlag, Berlin.

MICHALAKIS, Y., and L. EXCOFFIER, 1996 A generic estimation of population subdivision using distances between alleles with special reference for microsatellite loci. Genetics **142:** 1061–1064.

MORAN, P. A. P., 1975 Wandering distributions and the electrophoretic profile. Theor. Pop. Biol. **8:** 318–330.

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

OHTA, T., and M. KIMURA, 1973 A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. Genet. Res. **22:** 201–204.

SHRIVER, M. D., L. JIN, R. CHAKRABORTY and E. BOERWINKLE, 1993 VNTR allele frequency distributions under the stepwise mutation model: a computer simulation approach. Genetics **134:** 983–993.

SLATKIN, M., 1991 Inbreeding coefficients and coalescence times. Genet. Res. **58:** 167–175.

SLATKIN, M., 1993 Isolation by distance in equilibrium and non-equilibrium populations. Evolution **47:** 264–279.

SLATKIN, M., 1995 A measure of population subdivision based on microsatellite allele frequencies. Genetics **139:** 457–462.

TACHIDA, H., 1985 Joint frequencies of alleles determined by separate formulations for the mating and mutation systems. Genetics **111:** 963–974.

VALDES, A. M., M. SLATKIN and N. B. FREIMER, 1993 Allele frequencies at microsatellite loci: the stepwise mutation model revisited. Genetics **133:** 737–749.

WEBER, J. L., and C. WONG, 1993 Mutation of human short tandem repeats. *Hum. Mol. Genet.* **2:** 1123–1128.

WEHRHAHN, C. F., 1975 The evolution of selectively similar electrophoretically detectable alleles in finite natural populations. Genetics **80:** 375–394.

WEIR, B. S., 1990 *Genetic Data Analysis.* Sinauer, Sunderland, MA.

WEIR, B. S., and C. C. COCKERHAM, 1984 Estimating *F*-statistics for the analysis of population structure. Evolution **38:** 1358–1370.

WOLFRAM, S., 1991 *Mathematica,* Ed. 2. Addison Wesley, Redwood City, CA.

Communicating editor: W. J. EWENS