A Measure of Population Subdivision Based on Microsatellite Allele Frequencies

Montgomery Slatkin

Department of Integrative Biology, University of California, Berkeley, California 94720

Manuscript received June 19, 1994
Accepted for publication

MICROSATELLITE loci, loci that vary in the number of repeats of a simple DNA sequence, are becoming commonly used in the analysis of natural populations. Microsatellite loci are often highly polymorphic and relatively easy to survey and hence offer the hope of greater understanding of population structure. The question is how to make the best use of allele frequencies at microsatellite loci. This paper, like the accompanying paper by GOLDSTEIN *et al.* (1995), discusses how information about the mutation process at microsatellite loci can suggest statistics that are more appropriate for the analysis of microsatellite loci than are existing statistics. In this paper, I will introduce a new statistic analogous to WRIGHT's (1951) F_{ST} that can be used to estimate effective migration rates or times since population divergence. This statistic is closely related to the distance measures introduced by GOLDSTEIN *et al.* (1995).

F_{ST} AND COALESCENCE TIMES

I have shown previously that, for neutral loci under the infinite alleles model of mutation, there is a close relationship between F_{ST} and the average coalescence times within and between populations (SLATKIN 1991). That relationship provides a guide to what is needed for microsatellite loci so I will briefly review that theory here. If we have samples from d populations, we can, ignoring sampling considerations, compute F_{ST} by finding the average probability of identity in state of two alleles in each population, f_0, and the average probability of identity in state of two alleles chosen at random from all the populations together, ($\overline{f}$):

$$F_{ST} = \frac{f_0 - \overline{f}}{1 - \overline{f}}, \tag{1}$$

which is equivalent to NEI's (1973) definition for G_{ST}. Under the symmetric K alleles mutation model with mutation rate μ, a slight generalization of the method in SLATKIN (1991) shows

$$f_0 \approx 1 - \frac{K-1}{K}\mu\overline{t}_0 \tag{2a}$$

and

$$\overline{f} \approx 1 - \frac{K-1}{K}\mu\overline{t} \tag{2b}$$

in the limit of small values of μ, where $\overline{t}_0$ is the average coalescence time of two copies of a gene drawn from the same population and $\overline{t}$ is the average coalescence time of two copies of a gene drawn from the collection of populations. The result of SLATKIN (1991) for the infinite allele model is obtained by taking the limit of large K. Substituting (2) into (1), we have

$$F_{ST} = \frac{\overline{t} - \overline{t}_0}{\overline{t}} \tag{3}$$

in the limit of low mutation rates. This formula is convenient because both μ and K, which are in general unknown, cancel. It is also convenient because the values of $\overline{t}_0$ and $\overline{t}$ can be predicted for a variety of demographic models describing populations both at equilibrium and not at equilibrium (SLATKIN 1991, 1993). For example, in an island model with d populations, the equilibrium values are

$$\overline{t}_0 = 2Nd, \tag{4a}$$

$$\overline{t}_1 = 2Nd + \frac{(d-1)}{2m} \tag{4b}$$

and hence

$$\overline{t} = 2Nd + \frac{(d-1)^2}{2md} \tag{4c}$$

where N is the size of each population, m is the migration rate and $\overline{t}_1$ is the average coalescence time of two copies drawn from different populations (SLATKIN 1991). Substituting (4) into (3), we find

$$F_{ST} = \frac{1}{1 + \dfrac{4Nmd^2}{(d-1)^2}} \tag{5}$$

which is the result obtained by TAKAHATA (1983) and CROW and AOKI (1984) from direct analysis of the recursion equations for this model.

A STATISTIC FOR MICROSATELLITES

The reason that the above approach does not apply to microsatellite loci is that the mutation process at

Corresponding author. Montgomery Slatkin, Department of Integrative Biology, University of California, Berkeley, California 94720.
E-mail: monty@kaline.berkeley.edu

those loci does not conform to the *K*-allele model with low mutation rates. The assumption of low mutation rates is reasonable for allozyme loci but probably not for many microsatellite loci where rates may exceed 10^{-3} per generation (WEBER and WONG 1993). Furthermore, the *K*-allele model makes the assumption that the mutation process erases any memory of the prior allelic state, so excess genetic similarity between populations, as measured by F_{ST}, can be attributed to migration or historical association. With microsatellite loci, there is abundant evidence that the size of a new mutant allele depends on the size of the allele that mutated. Direct studies of mutations in human families have found that almost all mutants differ from their ancestor by one or two repeat units (*e.g.*, WEBER and WONG 1993). It is still an open question as to whether all mutations at microsatellite loci involve changes of only one or two repeat units, as assumed by VALDES *et al.* (1993), SHRIVER *et al.* (1993), and GOLDSTEIN *et al.* (1995) or whether mutations of larger effect occur occasionally, as suggested by DI RIENZO *et al.* (1994). In either case, mutation rates are high and the mutational process does not erase information about the ancestral state, so the assumptions made in using F_{ST} to estimate Nm or other demographic parameters are not satisfied.

We can use what is known about the mutation process to suggest another statistic that is more appropriate for microsatellite loci. We can assume a generalized stepwise mutation process in which the probability of a mutation is μ per generation and, when a mutation occurs, the increment in allele size is a random variable with mean 0 and variance σ_m^2 independently of allele size. The actual distribution of increments will not be important. The one-step mutation model is a special case with $\sigma_m^2 = 1$, and the two-phase model introduced by DI RIENZO *et al.* (1994) is another special case, which assumes that the distribution of changes in allele size under mutation is symmetric about 0.

First consider two copies of the locus, with allele sizes a_1 and a_2 measured in the number of repeat units, and assume that the time in the past at which they have a common ancestor (the coalescence time, t) is known. The two copies are then separated by a branch of a gene genealogy of total length $2t$. During that time, the number of mutations that occur is a random variable drawn from a Poisson distribution with mean $2\mu t$. Let α be the number of mutations that have occurred and let x_n be the increment in repeat number of the nth mutational event. Then

$$a_1 - a_2 = \sum_{n=1}^{\alpha} x_n. \tag{6}$$

where each x_n is drawn independently from a distribution with mean 0 and variance σ_m^2. Taking the expectation over the distribution of the x_n, $E(a_1 - a_2) = 0$ and

$$E[(a_1 - a_2)]^2 = \alpha\sigma_m^2 \tag{7}$$

and then taking the expectation of α,

$$E[(a_1 - a_2)^2] = 2\mu t\sigma_m^2. \tag{8}$$

Now imagine that, as in the previous section, we sample n individuals from each of d_s populations ($d_s \leq d$), and let a_{ij} be the allele size of the ith copy ($i = 1, \ldots, 2n$) in the jth population ($j = 1, \ldots, d_s$). We find the average sum of squares of the differences in allele size within each population to be

$$S_W = \frac{1}{d_s}\sum_{j=1}^{d_s}\frac{2}{2n(2n-1)}\sum_{i<i'}(a_{ij} - a_{i'j})^2, \tag{9a}$$

which is equivalent to D_0 of GOLDSTEIN *et al.* (1995). To estimate the average squared difference between all pairs of copies we define the between-population component, S_B, to be

$$S_B = \frac{2}{(2n)^2 d_s(d_s-1)}\sum_{j<j'}\sum_{i<i'}(a_{ij} - a_{i'j'})^2, \tag{9b}$$

which is equivalent to D_1 of GOLDSTEIN *et al.* (1995), to obtain

$$\bar{S} = \frac{2n-1}{2nd_s - 1}S_W + \frac{2n(d_s-1)}{2nd_s - 1}S_B. \tag{10}$$

The coefficients in (10) are the probabilities of choosing two different copies of the locus from the same population and two copies from different populations.

In practice, it may be easier to compute S_W and $\bar{S}$ directly from the variances of allele sizes. It is straightforward to show that S_W is twice the average of the estimated variances of allele size within each population and that $\bar{S}$ is twice the estimated variance in allele size in the collection of populations together, where the estimated variances are obtained using unbiased estimators.

From (8), we can find the expectation of S_W and $\bar{S}$ under this model of mutation:

$$E(S_W) = 2\mu\bar{t}_0\sigma_m^2 \tag{11a}$$

and

$$E(\bar{S}) = 2\mu\bar{t}\sigma_m^2, \tag{11b}$$

where, as above, $\bar{t}_0$ and $\bar{t}$ are the average pairwise coalescence times within populations and in the group of populations sampled. Both (10) and (11) depend on the parameters of the mutation model, just as f_0 and $\bar{f}$ do in (2) for the infinite alleles model. But in a ratio analogous to (1), the parameters of the mutation model cancel and we have

$$\frac{E(\bar{S}) - E(S_W)}{E(\bar{S})} = \frac{\bar{t} - \bar{t}_0}{\bar{t}} \tag{12}$$

which is exactly the result for F_{ST} under the *K*-allele

mutation model. Thus these results suggest that the ratio

$$R_{ST} = \frac{\overline{S} - S_W}{\overline{S}} \tag{13}$$

has the same properties for microsatellite loci that follow the mutation model used here as does F_{ST} for the K-alleles mutation model, and the notation is chosen to emphasize that similarity.

Because S_W and $\overline{S}$ are proportional to the within-population and total variances, R_{ST} is just the fraction of the total variance of allele size that is between populations. Thus, the value of R_{ST} is similar to θ defined by WEIR and COCKERHAM (1984), which is also a between-population component of variance. The difference here is that allele sizes are taken into account whereas in WEIR and COCKERHAM's analysis only identity or nonidentity of allelic state enters. For microsatellites, as in WEIR and COCKERHAM's approach, the analysis of variance can be performed at any level. One could, *e.g.*, define R_{IS} to be the within-individual component of variance in allele size and then use the formal theory of the analysis of variance to determine whether there is a significant within-individual component, thus testing for evidence of nonrandom mating. Or one could test for a significant hierarchical structure of a population.

When more than one locus is examined, the question arises of how to combine information across loci. For F_{ST}, WEIR and COCKERHAM (1984) recommend averaging the numerator and denominator in expressions equivalent to (1) and then taking the ratio. As GOLDSTEIN *et al.* (1995) point out, estimates of their D_0 and D_1 (and consequently S_W and $\overline{S}$ can be obtained by averaging across loci, because the expected values are proportional to time. In (11), the average of $\mu\sigma_m^2$ would appear, as noted by GOLDSTEIN *et al.*, but that average would still cancel when the ratio is taken. Hence, I will follow the procedure of GOLDSTEIN *et al.* (1995) and first compute S_W and $\overline{S}$ by averaging across loci and then taking the ratio to estimate R_{ST}.

DEMOGRAPHIC MODELS

In the simulations described below, I consider two demographic models. The first is the d-island model at equilibrium, for which the analytic theory presented above applies. For that model, both R_{ST} and F_{ST} lead to estimates of the product Nm, obtained by solving (5) and the equivalent expression for R_{ST}. To estimate Nm using samples from a relatively small number of populations, we have to take sampling considerations into account. In (11a), $\overline{t}$ is the average coalescence time of pairs of alleles in the sample, and hence is

$$\overline{t} = \frac{1}{d_s}\,\overline{t}_0 + \frac{d_s - 1}{d_s}\,\overline{t}_1, \tag{14a}$$

where $\overline{t}_0$ and $\overline{t}_1$ are given in (4). Therefore, the expected value of R_{ST} is

$$R_{ST} = \frac{1}{1 + 4Nm\left(\frac{d}{d-1}\right)\left(\frac{d_s}{d_s - 1}\right)}. \tag{14b}$$

Although it is reasonable to assume that d is large, it is often the case that d_s is small, so the estimate of Nm is obtained from

$$M_R = \frac{d_s - 1}{4d_s}\left(\frac{1}{R_{ST}} - 1\right). \tag{15a}$$

where M is the estimate of Nm and the subscript R indicates that the estimate was based on R_{ST}. In particular, if $d_s = 2$, as in the simulation results presented below and when testing for isolation by distance, an important factor of 2 enters (SLATKIN 1993). If F_{ST} is estimated using WEIR and COCKERHAM's $\hat{\theta}$ (as is done below), then the sampling considerations are already incorporated in the estimator and hence we can use

$$M_F = \frac{1}{4}\left(\frac{1}{F_{ST}} - 1\right). \tag{15b}$$

If instead NEI's G_{ST} is used then the factor of $(d_s - 1)/d_s$ would enter in the expression for M_F as well.

The second demographic model is one of two populations, now completely isolated, that are descended from a single ancestral population at some time t in the past. The ancestral population and both descendent populations are of size N and there is no subsequent gene flow. This is a special case of the "radiation model" that I have discussed previously (SLATKIN 1993). It is easy to compute the average coalescence times, $\overline{t}_0$ and $\overline{t}$ needed to predict F_{ST} and R_{ST}. Clearly $\overline{t}_0 = 2N$ because each population separately can be regarded as an isolated population of size N. And $\overline{t}_1$, the average coalescence time of one copy of the locus drawn from each of the two populations is just $\tau + 2N$, because there is no chance of coalescence until generation τ in the past (TAJIMA 1983). Hence,

$$\overline{t} = 2N + \frac{\tau}{2} \tag{16}$$

and, under the assumptions made above,

$$F_{ST} = R_{ST} = \frac{T}{T + 4} \tag{17}$$

in expectation, where $T = \tau/N$. Equation 17 yields two estimators of T for this model:

$$T_F = \frac{4F_{ST}}{1 - F_{ST}} \tag{18}$$

and

$$T_R = \frac{4R_{ST}}{1 - R_{ST}}. \quad (19)$$

SIMULATION PROGRAM

To determine whether R_{ST} is actually better suited than F_{ST} for analyzing microsatellite data, I carried out a simulation study of the two demographic models described above. I used a program that I had previously written to model neutral loci in a subdivided population (SLATKIN 1993). To that program, I added the two-phase mutation model of DI RIENZO *et al.* (1994). In that model, when a mutation occurs, the probability that the increment to allele size ± 1 is p, and the probability that it is a random variable drawn from a specified distribution, g (with variance σ_g^2) is $1 - p$. In both cases, the probabilities of an increase or decrease in allele size are equal. With $p = 1$, the two phase model reduces to the one-step model, but with $p < 1$, there is the possibility of mutations of much larger effect. DI RIENZO *et al.* (1994) found evidence that $p < 1$ for 8 of 10 loci examined in a sample of Sardinians. For the two-phase model, $\sigma_m^2 = p + (1 - p)\sigma_g^2$. In the simulations, I assumed $\sigma_g^2 = 50$ and adjusted the value of μ so that $2N\mu\sigma_m^2 = 10$ for three different values of p, with $N = 10{,}000$ in all cases. Thus, *e.g.*, $N\mu = 10$ for $p = 1$ and $N\mu = 0.755$ for $p = 0.75$. The resulting choices of parameter values were in the range of parameter values found to be appropriate for several microsatellite loci in the Sardinian population examined by DI RIENZO *et al.* (1994).

In the simulations, I assumed that samples of 50 individuals were drawn from each of two populations. In each replicate, the allelic state of each copy of the locus of each individual was found by first finding the gene genealogy and then assigning allelic states by working upward from the root. The results from one replicate represented hypothetical data for a single locus. One set of replicates consisted of 100 replicates representing a hypothetical (large) data set and the data were combined across loci in the way described above, using WEIR and COCKERHAM's (1984) $\hat{\theta}$ statistic to estimate F_{ST}, and by computing the averages of S_W and $\bar{S}$ to estimate R_{ST}. From F_{ST} and R_{ST}, estimates of M were computed from (14) and (15) for the island model and estimates of T from (18) and (19) for the special case of the radiation model. For each set of parameter values used in the simulations, 10 such sets of replicates were run. The results presented in Table 1 show the average estimates of M or T and their SDs over the 10 sets of data. A relatively large number of loci was used for each data set because my purpose here is to illustrate expected behavior of these statistics rather than to explore their sampling properties.

RESULTS AND DISCUSSION

The results from the simulation study are shown in Table 1. In interpreting these results, it is useful to note two points. First, the value of p in the mutation model determines how similar the mutation process is to the infinite alleles model and how high the mutation rate is relative to the migration rate. With $p = 1$ (the one-step model) the mutation process is most likely to produce the same allelic state more than once and the mutation rate is also the highest ($N\mu = 10$). Smaller values of p indicate that a higher proportion of the mutations are of large effect. Hence there is less chance for the same allele size to be produced twice by independent mutations, and the mutation rate is smaller ($N\mu = 0.755$ for $p = 0.75$). Although this range of mutation rates seems quite high compared with the usual values assumed for allozyme loci, they appear to be realistic for microsatellites and necessary to maintain variances of allele size at equilibrium in the range of observed values (VALDES *et al.* 1993). The second point to note is that the parameters of the demographic model determine the time scales of interest and hence the time scales in which mutation can act. The value of R_{ST} is determined by extra mutations that accumulate within each population. The opportunity for having such mutations depends on the difference between $\bar{t}$ and $\bar{t}_0$. When that difference is large, as is the case with small values of Nm or large values of τ/N, there is ample time for mutations to accumulate and reflect the true demographic structure, whereas if Nm is large or τ/N is small, there is relatively little time. In the latter case genetic drift is likely to be more important than mutation.

With these points in mind, the results are easy to interpret. Estimates based on F_{ST} show too much genetic similarity, particularly when there is a relatively large difference in average coalescence times (large τ/N or small Nm), whereas estimates using R_{ST} seem to be unbiased or to have little bias. Also, the performance of F_{ST} improves as p decreases, because the mutation rate is lower and because the mutation model is closer to the infinite alleles model. The performance of F_{ST} also improves when the difference in average coalescence times is relatively small (small τ/N and large Nm), because then genetic drift is the dominant process in creating local differentiation and mutation plays little role.

It is worth noting that the values of M_F are not as high as they would be under an infinite alleles model with the same mutation rate, so the bias of M_F is not attributable simply to the higher mutation rate assumed. To see this, we use the result that, under the infinite alleles model, $F_{ST} \approx 1/(1 + 4Nm + 4N\mu)$ when μ is not negligible compared with m (CROW and AOKI 1984), so $M_F = Nm + N\mu$ under the infinite alleles

TABLE 1

Estimates of demographic parameters in simulated data sets

	$p = 1.0$		$p = 0.9$		$p = 0.75$	
Nm	M_F	M_R	M_F	M_R	M_F	M_R
	A. Estimates of $M = Nm$ is an island model with 10 populations at equilibrium					
0.1	2.680 ± 0.101	0.122 ± 0.047	1.084 ± 0.056	0.114 ± 0.028	0.719 ± 0.032	0.120 ± 0.054
1.0	5.195 ± 0.195	0.983 ± 0.143	2.405 ± 0.147	1.154 ± 0.264	1.736 ± 0.099	1.225 ± 0.360
10.0	14.54 ± 0.479	12.23 ± 3.29	10.39 ± 0.516	13.38 ± 2.92	9.787 ± 0.292	11.85 ± 5.13
	$p = 1.0$		$p = 0.95$		$p = 0.75$	
τ/N	T_F	T_R	T_F	T_R	T_F	T_R
	B. Estimates of $T = \tau/N$ in the radiation model with two populations					
0.1	0.108 ± 0.007	0.092 ± 0.019	0.159 ± 0.010	0.088 ± 0.020	0.194 ± 0.020	0.104 ± 0.036
0.5	0.200 ± 0.015	0.479 ± 0.050	0.382 ± 0.015	0.512 ± 0.098	0.694 ± 0.038	0.512 ± 0.117
1.0	0.252 ± 0.015	1.092 ± 0.132	0.479 ± 0.032	0.895 ± 0.161	1.049 ± 0.058	1.005 ± 0.325

Samples of 50 individuals from each of two populations were analyzed. In all cases, the population size, N, was 10,000, and the parameters of the two-phase mutation model of DI RIENZO *et al.* (1994) were chosen so that $2N\mu\sigma_m^2 = 10$, with the variance of the second phase, $\sigma_g^2 = 50$. The parameter p is the proportion of one-step mutations, with $p = 1$ corresponding to the one-step model used by GOLDSTEIN *et al.* (1995). The results shown represent the averages (±SD) of 10 estimates of each parameter (M or T). The subscript (F or R) indicates which statistic (F_{ST} or R_{ST}) was used to estimate the parameter.

model. But the values of M_F in Table 1 are all smaller than this value, indicating that the mutation process works in the opposite direction, presumably by reducing the number of possible allelic states.

The importance of the differences in average coalescence times is also evident in the SDs of the estimators. The coefficient of variation for M_F and T_F is usually lower than for M_R and T_R and that is particularly true for $Nm = 10$ and $\tau/N = 0.1$. The difference in average coalescence times is sufficiently short that the values of R_{ST} obtained depend on the occurrence of relatively few mutations, even with the large data sets used in these simulations. Hence the coefficients of variation in M_R and T_R are a factor of 2 or more larger than for M_F and T_F. The lower coefficient of variation for estimates based on F_{ST} is offset by the possiblity of bias.

The generally better performance of R_{ST} relative to F_{ST} is attributable to the fact that R_{ST} was designed to fit the generalized stepwise mutation model used in the simulations. If that model is not appropriate for microsatellite loci, then the performance of R_{ST} would suffer accordingly. The two key assumptions of the mutation model are that there be no constraints on allele size and that the properties of the mutation process not depend on allele size. There is already evidence of constraints on allele size based on interspecies comparisons (BOWCOCK *et al.* 1994; GOLDSTEIN *et al.* 1995). But as discussed by GOLDSTEIN *et al.* (1995) those constraints may not be important for the relatively short time scales of interest when estimating demographic parameters for a single species. For minisatellites, there is also evidence that the mutation process itself depends on allele size. JEFFREYS *et al.* (1994) found the tendency for relatively small alleles to increase in size under mutation. For technical reasons, JEFFREYS *et al.* (1994) could not test for the reverse tendency for relatively large alleles. WEBER and WONG (1993), however, found no evidence of similar bias for microsatellites. Whatever mutation model is appropriate for microsatellites, there is clearly some memory to the mutation process, suggesting that F_{ST} will yield biased estimates of demographic parameters except in cases where the time scale of interest is sufficiently short that mutation plays little role. If a typical mutation rate at a microsatellite locus is 10^{-3}, then F_{ST} can be used if it is known from other information that the time scales of interest are tens or hundreds of generations. But the value of F_{ST} itself cannot be used to determine the time scale of interest because the simulation results show that it will always indicate genetic similarity, even when that is not justified. The simulation results suggest that the difference between estimates of demographic parameters based on F_{ST} and R_{ST} could provide more information, although the extent of the difference depends on the value of p, the proportion of one-step mutations, and that parameter is currently not known. The results of DI RIENZO *et al.* (1994) suggest that p is close to 1 for most loci, and that is consistent with direct observations of mutations that have not found changes in allele size of more than one or two repeat units (WEBER and WONG 1993).

CONCLUSIONS

We can conclude that under the assumptions of the generalized stepwise model of mutation at microsatel-

lite loci, the statistic R_{ST} will generally provide less biased estimates of demographic parameters for a population than will F_{ST}. That conclusion depends on what can be assumed about the mutation process at microsatellite loci, but currently available information supports the assumptions made in the simulations described here, at least for modeling differences among populations of the same species. The results also suggest that comparing estimates of demographic parameters obtained by using R_{ST} and F_{ST} might provide further information about the time scales of interest in the populations being examined.

I thank F. BONHOMME, D. COUVET, D. GOLDSTEIN, Y. MICHALAKIS and I. OLIVIERI for helpful discussions of this topic and comments on an earlier version of this paper. This research was supported in part by a grant from the National Institutes of Health. Much of the work for this paper was done while supported by the Centre Nationale de Recherche Scientifique (CNRS) and while visiting the laboratory of F. BOHOMME in Montpellier, France.

LITERATURE CITED

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

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

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.

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

JEFFREYS, A. J., K. TAMASKI, A. MCLEOD, D. G. MONCKTON, D. L. NEIL, *et al.*, 1994 Complex gene conversion events in germline mutation at huma minisatellites. Nature Genetics **6:** 136–145.

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

SHRIVER, M. D., L. JIN, R. CHAKRABORTY and E. BOERWINKLE, 1993 VNTR allele frequency distributions under the stepwise mutation model. 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 nonequilibrium populations. Evolution **47:** 264–279.

TAJIMA, F., 1983 Evolutionary relationship of DNA sequences in finite populations. Genetics **105:** 437–460.

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

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

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

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

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

Communicating editor: W. J. EWENS