Published in final edited form as:

Theor Popul Biol. 2011 August ; 80(1): 29–37. doi:10.1016/j.tpb.2011.04.004.

Genetic Diversity of Microsatellite Loci in Hierarchically Structured Populations

Seongho Song^{*,1}, Dipak K. Dey^{**}, and Kent E. Holsinger^{***}

^{*}Department of Mathematical Sciences University of Cincinnati, ML:210025 Cincinnati, OH 45221-0025

^{**}Department of Statistics University of Connecticut, U-4120 Storrs, CT 06269-4120

^{***}Department of Ecology & Evolutionary Biology University of Connecticut, U-3043 Storrs, CT 06269-3043

Abstract

Microsatellite loci are widely used for investigating patterns of genetic variation within and among populations. Those patterns are in turn determined by population sizes, migration rates, and mutation rates. We provide exact expressions for the first two moments of the allele frequency distribution in a stochastic model appropriate for studying microsatellite evolution with migration, mutation, and drift under the assumption that the range of allele sizes is bounded. Using these results we study the behavior of several measures related to Wright's F_{ST} , including Slatkin's R_{ST} . Our analytical approximations for F_{ST} and R_{ST} show that familiar relationships between $N_e m$ and F_{ST} or R_{ST} hold when migration and mutation rates are small. Using the exact expressions for F_{ST} and R_{ST} , our numerical results show that when migration and mutation rates are large, these relationships no longer hold. Our numerical results also show that the diversity measures most closely related to F_{ST} depend on mutation rates, mutational models (stepwise *versus* two-phase), migration rates, and population sizes. Surprisingly, R_{ST} is relatively insensitive to mutation rates and mutational models. The differing behaviors of R_{ST} and F_{ST} suggest that properties of the among-population distribution of allele frequencies may allow the roles of mutation and migration in producing patterns of diversity to be distinguished, a topic of continuing investigation.

Keywords

genetic diversity; genetic drift; microsatellite loci; mutation; migration; R_{ST} ; F_{ST}

1. Introduction

In the last decade microsatellite markers have become a standard tool for genetic analysis. Because of the relative ease with which they can be isolated and the large allelic diversity commonly present at each locus, they are widely used in the construction of genetic maps (e.g., Chistiakov et al., 2005; Kai et al., 2005), the identification of quantitative trait loci (e.g., Allan et al., 2005; Minvielle et al., 2005), and the analysis of gene flow and other

¹Corresponding author, Fax:+1-513-556-4047, seongho.song@uc.edu. Corresponding author: Dipak K. Dey Department of Statistics, U-4120 University of Connecticut Storrs, CT 06269-4120 Phone: (860) 486-4196 Fax: (860) 486-4113 dey@stat.uconn.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

evolutionary processes (e.g., Hall and Willis, 2005; Kretzer et al., 2005). In particular, many evolutionary applications use measures of population divergence derived from microsatellite markers as an indicator of the evolutionary distance among populations or the degree of evolutionary connection among them (e.g., $(\delta\mu)^2$: Goldstein et al., 1995; R_{ST} Slatkin, 1995). These measures are inspired by Wright's (1951) observation that the proportion of genetic diversity due to among-population differentiation can be a useful index of the degree to which populations are evolutionarily connected by gene flow. For more than fifty years Wright's F -statistics have been the most widely used index for describing the genetic structure of populations.

Useful as they are, Wright's F -statistics are implicitly based on the assumption that all alleles are mutationally equidistant from one another. Indeed, the widely adopted Weir and Cockerham framework for evolutionary inference from F -statistics (Weir and Cockerham, 1984; Weir and Hill, 2002) is based on co-ancestry or probabilities of identity by descent. Thus, an “infinite alleles” model of mutation underlies typical approaches to evolutionary inference from Wright's F -statistics (compare Rousset, 1996). As Slatkin (1995) and Goldstein et al. (1995) pointed out, however, an infinite alleles model of mutation may not be appropriate for microsatellite loci. The combination of high mutation rates (see, for example, Xu et al., 2005) and predominantly stepwise mutation among adjacent allele classes (see, for example, Calabrese et al., 2001) implies that alleles of the same size may have different mutational histories (i.e., homoplasy, see Estoup et al., 2002) and that alleles of similar size will tend to covary in frequency.

Several authors have studied the evolutionary dynamics of microsatellite loci either in models of isolated populations (for example, Feldman et al., 1997) or in models where migration occurs according to a finite-island model (for example, Rousset, 1996). For a stepwise mutation model (Ohta and Kimura, 1973; Wehrhahn, 1975), Rousset shows that R_{ST} has the familiar relationship with the migration rate and population size when migration and mutation are rare. Specifically,

$$R_{ST} = \frac{1}{4Nm\alpha + 1}, \quad (1)$$

where N is the local population size, m is the (backwards) migration rate, $\alpha = k/(k-1)$, and k is the number of populations. Similarly, simulations in Feldman et al. (1997) show that $(\delta\mu)^2$ increases roughly linearly with time since divergence for isolated populations. Nonetheless, the relationship in (1) is only approximate. The magnitude of among population differentiation tends to be smaller than predicted from (1) because of mutation-induced homoplasy (Rousset, 1996).

In this paper, we use the modeling framework introduced in Fu et al. (2003) to provide the allele frequency distribution in a stochastic evolutionary model appropriate for investigating the evolutionary dynamics of microsatellite loci, under the assumption that the range of allele sizes is bounded (compare Feldman et al., 1997; Pollock et al., 1998). We use our results to study the behavior of R_{ST} and of two measures more directly related to Wright's F_{ST} as functions of local population size, migration rate, and mutation rate both for the usual stepwise model of mutation and for a more realistic two-phase model proposed by Di Rienzo et al. (1994) and studied numerically by Rousset (1996). Although all of the results we present assume that the drift-mutation-migration process has reached stationarity, the close relationship between R_{ST} and coalescence times (Slatkin 1995) suggests that R_{ST} may be a useful index of gene flow among populations, so long as we remember that “gene flow” may refer either to recent common ancestry, continuing migration, or any combination of

these two. Finally, we discuss implications of our results for analysis and interpretation of data derived from microsatellite loci.

2. Process Model and Results

Fu et al. (2003) described a general stochastic framework for the study of drift, migration, and mutation. Here we consider models developed within that framework that are designed to illuminate the evolutionary dynamics of microsatellite alleles in which the range of allele sizes is bounded (compare Feldman et al., 1997; Pollock et al., 1998). Specifically, we focus on a single locus with A alleles, $b_1, b_2, \dots, b_A$. In the context of microsatellite variation, allele b_{j+1} has one more repeat unit than allele b_j . Correspondingly, allele b_1 corresponds to the allele with the smallest number of repeat units and b_A to the allele with the largest number of repeat units. Let $\mathbf{V}_{A \times A}$ be a general mutation matrix with elements, v_{rs} , the probability of mutation from allele type b_r to allele type b_s , $r = 1, \dots, A$ and $s = 1, \dots, A$.

2.1. Main Results

Assume that there are k populations indexed by i . Let $\mathbf{M}_{k \times k}$ be a general (backward) migration matrix, i.e., $\mathbf{M} = \left(\left(\bar{m}_{ij} \right) \right)$ where $\bar{m}_{ij}$ is the probability that the allele in population i came from population j (compare Nagylaki, 1982; Rousset, 1999, 2001). Let $\mathbf{p}_i^{(t)}$ be the $A \times 1$ vector of allele frequencies in population i at generation t , ($\mathbf{1}' \mathbf{p}_i^{(t)} = 1$). Concatenate the $\mathbf{p}_i^{(t)}$ to a $kA \times 1$ vector $\mathbf{p}^{(t)}$ and define

$$\mathbf{p}^{*(t)} = (\mathbf{M} \otimes \mathbf{V}') \mathbf{p}^{(t)} \quad (2)$$

where $\otimes$ denotes the Kronecker product between two matrices. Let $\mathbf{B} = \mathbf{B}(\mathbf{M}, \mathbf{V}')$ be $\mathbf{M} \otimes \mathbf{V}'$ then we can write $\mathbf{p}^{*(t)} = \mathbf{B}(\mathbf{M}, \mathbf{V}') \mathbf{p}^{(t)}$ for convenience. Let N_i be the number of individuals in the i^{th} population and let $\mathbf{N}$ be the vector of population sizes. For diploid organisms, the number of allele copies is $2N_i$. Given $\mathbf{p}^{*(t)}$ and N_i , the $\mathbf{p}_i^{(t+1)}$ are conditionally independent with

$$2N_i \mathbf{p}_i^{(t+1)} \sim \mathcal{M}(2N_i, \mathbf{p}_i^{*(t)}), \quad (3)$$

where $\mathcal{M}$ denotes a multinomial distribution. Through (2) and (3), we pass from $\mathbf{p}^{(t)} \rightarrow \mathbf{p}^{*(t)} \rightarrow \mathbf{p}^{(t+1)}$.

Fu et al. (2003) showed that the stationary mean satisfies

$$\mathbf{u} = \mathbf{B}(\mathbf{M}, \mathbf{V}') \mathbf{u} \quad (4)$$

where $\mathbf{u} = E(\mathbf{p}^{(t)} | \mathbf{M}, \mathbf{V}, \mathbf{N})$ as t tends to infinity. Additional analysis shows that the stationary mean is identical across populations and corresponds to the left eigenvector of $\mathbf{V}$ associated with the leading eigenvalue of unity.

If we assume that $N_i = N$ and that the entries in $\mathbf{M}$ do not depend on population indices, then a common distribution for all $\mathbf{p}_i^{(t)}$ will arise, regardless of $\mathbf{V}$. Thus, we can describe the stationary covariance structure for the entire set of k populations in terms of a single covariance matrix within populations, Σ_{11} , and another between populations, Σ_{12} . Given

that the entries in $\mathbf{M}$ do not depend on population indices, we denote the diagonal elements of $\mathbf{M}$ by $(1 - m)$ and the off-diagonal elements as $m/(k - 1)$. This migration model corresponds to the finite-island model studied by Crow and Aoki (1984) and Cockerham and Weir (1987).

The stationary equations for covariances are

$$(\mathbf{B}\Sigma\mathbf{B}')_{11} = \mathbf{V}' \{ (1 - r_k) \Sigma_{11} + r_k \Sigma_{12} \} \mathbf{V}$$

and

$$(\mathbf{B}\Sigma\mathbf{B}')_{12} = \mathbf{V}' \left\{ \frac{r_k}{k-1} \Sigma_{11} + \left(1 - \frac{r_k}{k-1} \right) \Sigma_{12} \right\} \mathbf{V},$$

where $r_k = r(m, k) = 2m - m^2 k / (k - 1)$. Solving for Σ_{11} and Σ_{12}

$$\Sigma_{11} = \left(1 - \frac{1}{2N} \right) \mathbf{V}' \{ (1 - r_k) \Sigma_{11} + r_k \Sigma_{12} \} \mathbf{V} + \frac{1}{2N} \left(\frac{1}{A} \mathbf{I}_A - \frac{1}{A^2} \mathbf{1}_A \mathbf{1}'_A \right) \quad (5)$$

and

$$\Sigma_{12} = \mathbf{V}' \left\{ \frac{r_k}{k-1} \Sigma_{11} + \left(1 - \frac{r_k}{k-1} \right) \Sigma_{12} \right\} \mathbf{V}, \quad (6)$$

(compare Fu et al. 2003). Further analysis shows that

$$\Sigma_{11} = \mathbf{Q} \Phi_{11} \mathbf{Q}' \left(\frac{1}{A} \mathbf{I}_A - \frac{1}{A^2} \mathbf{1}_A \mathbf{1}'_A \right) \quad (7)$$

and

$$\Sigma_{12} = \mathbf{Q} \Phi_{12} \mathbf{Q}' \left(\frac{1}{A} \mathbf{I}_A - \frac{1}{A^2} \mathbf{1}_A \mathbf{1}'_A \right), \quad (8)$$

where $\Phi_{11} = \text{diag}(\phi_{11,1}, \dots, \phi_{11,A})$ and $\Phi_{12} = \text{diag}(\phi_{12,1}, \dots, \phi_{12,A})$ with

$$\phi_{11,j} = \frac{\frac{1}{2N} \left\{ 1 - \left(1 - \frac{r_k}{k-1} \right) \lambda_j^2 \right\}}{\left[1 - \left(1 - \frac{1}{2N} \right) \lambda_j^2 \right] \left[1 - \left(1 - \frac{k}{k-1} r_k \right) \lambda_j^2 \right] - \frac{r_k}{2N} \lambda_j^2}$$

and

$$\phi_{12,j} = \frac{\frac{1}{2N} \frac{r_k}{k-1} \lambda_j^2}{\left[1 - \left(1 - \frac{1}{2N} \right) \lambda_j^2 \right] \left[1 - \left(1 - \frac{k}{k-1} r_k \right) \lambda_j^2 \right] - \frac{r_k}{2N} \lambda_j^2}$$

for $j = 2, \dots, A$ (see Appendix A for details). Further, we note that, as k tends to infinity, $\phi_{12,j}$, for all $j = 2, \dots, A$ become zero.

2.2. Stepwise Mutation Model for microsatellite loci

Microsatellite loci consist of short (2–6) nucleotide sequences repeated as many as 100 times (Tautz, 1993). Differences among alleles correspond to differences in the number of repeat units. Because mutations occur predominantly through mispairing or slippage, the stepwise mutation model originally developed for study of charge-state variation in isozyme alleles (Ohta and Kimura, 1973; Wehrhahn, 1975) is a widely used approximation to the mutational process. The mutation matrix, $\mathbf{V}$ in this case is of size $A \times A$ and is given by

$$\mathbf{V} = \begin{pmatrix} 1 - \frac{\mu}{2} & \frac{\mu}{2} & 0 & 0 & \cdots & 0 & 0 & 0 \\ \frac{\mu}{2} & 1 - \mu & \frac{\mu}{2} & 0 & \cdots & 0 & 0 & 0 \\ 0 & \frac{\mu}{2} & 1 - \mu & \frac{\mu}{2} & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & \frac{\mu}{2} & 1 - \mu & \frac{\mu}{2} \\ 0 & 0 & 0 & 0 & \cdots & 0 & \frac{\mu}{2} & 1 - \frac{\mu}{2} \end{pmatrix},$$

where the alleles sizes lie in the discrete space, $(1, 2, \dots, A)$, with allele size 1 corresponding to the smallest number of repeat units and allele size A corresponding to the largest number of repeat units.

The eigenvalues of $\mathbf{V}$ are $\lambda_j = 1 - \mu + \mu \cos\left(\frac{(j-1)\pi}{A}\right)$ for $j = 1, \dots, A$ and the corresponding eigenvectors are free from the mutation rate, μ , which are $\mathbf{q}_j = \mathbf{q}_j^* / \|\mathbf{q}_j^*\|$ for $j = 1, \dots, A$, where $\mathbf{q}_j^* = (q_{j1}^*, \dots, q_{jA}^*)'$ with $q_{jl}^* = \cos((2l-1)\alpha_j) / \cos(\alpha_j)$ for $l = 1, 2, \dots, A$ and $\alpha_j = \frac{(j-1)\pi}{2A}$. Details of the analysis are provided Appendix B.

Now we see that $\mathbf{q}_1 = \frac{1}{\sqrt{A}} \mathbf{1}_A$ and $\mathbf{q}_j' \mathbf{1}_A = 0$ for all $j = 2, \dots, A$. Furthermore, it turns out that

$$\begin{aligned} \Sigma_{11} &= \mathbf{Q} \Phi_{11} \mathbf{Q}' \left(\frac{1}{A} \mathbf{I}_A - \frac{1}{A^2} \mathbf{1}_A \mathbf{1}_A' \right) \\ &= \sum_{j=1}^A \phi_{11,j} \mathbf{q}_j \mathbf{q}_j' \left(\frac{1}{A} \mathbf{I}_A - \frac{1}{A^2} \mathbf{1}_A \mathbf{1}_A' \right) \\ &= \frac{1}{A} \sum_{j=1}^A \phi_{11,j} \mathbf{q}_j \mathbf{q}_j' - \frac{1}{A^2} \sum_{j=1}^A \phi_{11,j} \mathbf{q}_j \mathbf{q}_j' \mathbf{1}_A \mathbf{1}_A' \\ &= \frac{1}{A} \phi_{11,1} \mathbf{q}_1 \mathbf{q}_1' \left(\mathbf{I}_A - \frac{1}{A} \mathbf{J}_A \right) + \frac{1}{A} \sum_{j=2}^A \phi_{11,j} \mathbf{q}_j \mathbf{q}_j' \\ &= \frac{1}{A^2} \phi_{11,1} \mathbf{J}_A \left(\mathbf{I}_A - \frac{1}{A} \mathbf{J}_A \right) + \frac{1}{A} \sum_{j=2}^A \phi_{11,j} \mathbf{q}_j \mathbf{q}_j' \\ &= \frac{1}{A} \sum_{j=2}^A \phi_{11,j} \mathbf{q}_j \mathbf{q}_j' \end{aligned}$$

and similarly

$$\begin{aligned}\Sigma_{12} &= \mathbf{Q}\Phi_{12}\mathbf{Q}'\left(\frac{1}{A}\mathbf{I}_A - \frac{1}{A^2}\mathbf{1}_A\mathbf{1}_A'\right) \\ &= \frac{1}{A}\sum_{j=2}^A \phi_{12,j}\mathbf{q}_j\mathbf{q}_j'\end{aligned}$$

Notice that Σ_{11} and Σ_{12} do not depend on the largest eigenvalue, $\lambda_1 = 1$ or its corresponding eigenvector, $\mathbf{q}_1 = \frac{1}{\sqrt{A}}\mathbf{1}_A$.

3. F_{ST} Analysis for Microsatellite Loci

Wright (1951) and Malécot (1948) introduced F -statistics to describe hierarchical structure in genetic data for one locus with two alleles, defining F_{ST} as a scaled variance

$$F_{ST} = \frac{\sigma_p^2}{\mu_p(1 - \mu_p)}, \quad (9)$$

where μ_p is the mean allele frequency across populations and σ_p^2 is the variance in allele frequency among populations. Equivalently, F_{ST} can be regarded as the intraclass correlation coefficient between pairs of alleles arising from a random-effects model of population sampling, as in the widely adopted Weir and Cockerham framework for population structure analysis (Cockerham, 1969; Weir and Cockerham, 1984; Weir, 1996; Weir and Hill, 2002). Fu et al. (2003) and Song et al. (2004) point out that in an evolutionary context there are two statistics related to equation (9) that might be of interest:

$$\theta^{(I)} = \frac{\sigma_{p^{(t)}}^2}{\mu_p(1 - \mu_p)},$$

which corresponds to the scaled *temporal* variance in allele frequency, and

$$\theta(p_1^{(t)}, \dots, p_k^{(t)}) = E \left(\frac{(1/k) \sum (p_i^{(t)} - \mu_{p^{(t)}})^2}{\mu_{p^{(t)}}(1 - \mu_{p^{(t)}})} \right)$$

with $\mu_{p^{(t)}} = (1/k) \sum p_i^{(t)}$, which corresponds to the scaled *geographical* variance in allele frequency. When the number of populations exchanging genes is even moderately large, say 10 or more,

$$\theta^{(II)} = \frac{E \left((1/k) \sum (p_i^{(t)} - \mu_{p^{(t)}})^2 \right)}{E \left(\mu_{p^{(t)}}(1 - \mu_{p^{(t)}}) \right)}$$

provides a satisfactory approximation to $\theta(p_1^{(t)}, \dots, p_k^{(t)})$ by the Central Limit Theorem and the Slutsky's theorem in probability theory.

In addition to using the analytical results above to study the behavior of $\theta^{(I)}$ and $\theta^{(II)}$ as a function of mutation rates, migration rates, and population size, we will also consider the behavior of R_{ST} , an analogue of F_{ST} that is sensitive not only to allele frequency differences among populations but also to repeat-size differences among those alleles. Slatkin (1995) introduced R_{ST} , defining it as

$$R_{ST} = \frac{\bar{S} - S_w}{\bar{S}}$$

where

$$\bar{S} = \frac{2N-1}{2Nk-1} S_w + \frac{2N(k-1)}{2Nk-1} S_B.$$

S_w is the average sum of squares of the differences in allele size within each population, which is equivalent to D_0 of Goldstein et al. (1995), and S_B is the average sum of squares of the differences in allele size between populations, which is equivalent to D_1 of Goldstein et al. (1995). Because S_w and $\bar{S}$ are proportional to the within-population and total variances, R_{ST} is just the proportion of the total allele-size variance accounted for by differences among populations. Therefore, R_{ST} has an interpretation similar to that of Weir and Cockerham's (1984) θ , which is also defined as a ratio of among-population to total variances. Moreover, Slatkin (1995) points out that for a stepwise mutation model R_{ST} is related to the excess

coalescence time for alleles found in different populations. Specifically, $R_{ST} \approx (\bar{t} - t_w) / \bar{t}$, where $\bar{t}$ is the average coalescence time for alleles drawn at random without respect to population and t_w is the average coalescence time for allele drawn at random within populations.

3.1. Asymptotic results for θ statistics

Suppose that $(\mathbf{p}_1^{(t)}, \dots, \mathbf{p}_k^{(t)})$ arise under (2) and (3). At stationarity, it is shown that

$$\theta^{(I)} = \frac{\text{tr}(\Sigma_{11})/A}{\frac{1}{A}\left(1 - \frac{1}{A}\right)} \quad (10)$$

and

$$\theta^{(II)} = \frac{1}{A} \sum_{j=1}^A \frac{\frac{k-1}{k} (\sigma_{11,j}^2 - \sigma_{12,j}^2)}{\frac{1}{A} \left(1 - \frac{1}{A}\right) - \frac{1}{k} (\sigma_{11,j}^2 + (k-1) \sigma_{12,j}^2)}, \quad (11)$$

where $\sigma_{11,j}^2$ and $\sigma_{12,j}^2$ are the j^{th} diagonal element of Σ_{11} and Σ_{12} , respectively, and $\text{tr}(\cdot)$ denotes the trace of a matrix. Notice that when k is moderate to large

$$\theta^{(II)} = \sum_{j=1}^A \frac{(\sigma_{11,j}^2 - \sigma_{12,j}^2)/A}{\frac{1}{A} \left(1 - \frac{1}{A}\right) - \sigma_{12,j}^2}.$$

Thus, $\theta^{(I)} > \theta^{(II)}$ unless $\sigma_{12,j}^2 = 0$. Since $\sigma_{12,j}^2 \rightarrow 0$ as $k \rightarrow \infty$, $\theta^{(II)} \rightarrow \theta^{(I)}$ as $k \rightarrow \infty$. Moreover, as k tends to infinity, $r_k = 2m - m^2$ and

$$\begin{aligned}\theta^{(I)} &= \frac{1}{A-1} \sum_{j=2}^A \phi_{11,j} \\ &= \frac{1}{A-1} \sum_{j=2}^A \left[2N - (2N-1)(1-r_k) \lambda_j^2 \right]^{-1}.\end{aligned}$$

Further, we observe that, once we ignore the terms $O(Nm^2)$ and $O(N\mu^2)$,

$$\theta^{(I)} = \frac{1}{A-1} \sum_{j=2}^A \psi(j), \quad (12)$$

where $\psi(j) = [1 - 2m + 4Nm + 2(2N-1)(1-2m)\mu(1 - \cos \alpha_j)]^{-1}$ for $j = 2, \dots, A$. Thus, $\theta^{(I)}$ is the simple average of $\psi(j)$. Moreover, if $O(N\mu)$ is negligible and the migration is negligible with respect to Nm , we find that $\theta^{(I)} \approx (1 + 4Nm)^{-1}$. Further, we observe that $\theta^{(I)}$ has the lower bound, $[1 - 2m + 4Nm + 4(2N-1)(1-2m)\mu]^{-1}$ and the upper bound, $(1 + 4Nm)^{-1}$.

3.2. Asymptotic results for R_{ST}

Rousset (1996) pointed out that with stepwise mutation and unbounded allele sizes, $R_{ST} \approx (1 + 4\alpha Nm)^{-1}$ where $\alpha = k/(k-1)$. Our approach provides similar results. Specifically, when the terms $O(Nm^2)$ and $O(N\mu^2)$ are negligible,

$$R_{ST} \approx \sum_{j=2}^A w_j \cdot \psi(j),$$

where

$$w_j = - \frac{\sum_{l=1}^A \sum_{l'=1}^A (l-l')^2 q_{jl} q_{jl'}}{\sum_{l'=1}^A (l-l')^2 / A}$$

and $\psi(j)$ is as before. Thus, R_{ST} is the weighted average of $\psi(j)$, where the weights depend only on the differences in allele sizes and the number of alleles since

$$\sum_{j=2}^A q_{jl} q'_{jl} = \begin{cases} 1 - \frac{1}{A}, & l=l' \\ -\frac{1}{A}, & l \neq l' \end{cases}.$$

If m and μ are negligible and k tends to infinity, then

$$\begin{aligned}
 R_{ST} &\approx \frac{1}{1+4Nm} \cdot \frac{1}{A} \sum_{j=2}^A \frac{-\sum_{l=1}^A \sum_{l'=1}^A (l-l')^2 q_{jl} q_{jl'}}{\sum_{l=1}^A \sum_{l'=1}^A (l-l')^2 / A^2} \\
 &= \frac{1}{1+4Nm} \cdot \sum_{j=2}^A w_j \\
 &= (1+4Nm)^{-1},
 \end{aligned}$$

showing that in the limit of a large number of populations, small mutation rates, and small migration rates $\theta^{(I)}$, $\theta^{(II)}$, and R_{ST} have equivalent values.

3.3. Exact results from numerical studies:

While the asymptotic results just presented provide some insight into patterns of population differentiation expected at microsatellite loci, they are limited in two respects. First, the degree to which asymptotic results apply when the number of populations is moderate or small and when mutation or migration rates are moderate is unknown. Second, they depend on a highly simplified model of mutation, namely the stepwise mutation model. In this section we use the exact results in (4)–(8) to study the behavior of $\theta^{(I)}$, $\theta^{(II)}$, and R_{ST} over a broad range of mutation rates and migration rates. In addition, we explore the sensitivity of these parameters to the details of the mutational process by comparing results from the stepwise mutation model with an extreme case of the two-phase model suggested by Di Rienzo et al. (1994). In the two-phase model mutations may increase or decrease microsatellite size by more than one repeat. Specifically, with probability ϕ , mutation increases or decreases allele size difference by one repeat, and with probability $1 - \phi$ it increases or decreases allele size difference by j repeats, where j follows some probability distribution. Di Rienzo et al. (1994) considered a truncated geometric distribution where $Pr(j) \propto \alpha^j$ for $j \geq 1$. We restrict our attention to the case where $\phi = 0$, which results in more multistep mutations than any other choice of ϕ .

Figure 1 displays the behavior of $\theta^{(I)}$ and $\theta^{(II)}$ as a function of μ and m for two combinations of k and A and $2N = 100$. As expected, both parameters decrease towards zero as the migration rate increases. Similarly, both parameters decrease towards zero as the mutation rate increases, because mutation-induced homoplasy causes similarity among populations when mutation rates are high, unless mutation matrices are population dependent. The high values of $\theta^{(I)}$ with small k , high m , and low μ may be initially surprising, but notice that $\theta^{(I)}$ depends on the variance of allele frequencies over time, not over populations. Under these conditions populations will be nearly fixed for one allele or another at all times, causing the variance of allele frequency over time and $\theta^{(I)}$ to be near their maxima. Thus, populations with similar allele frequencies at high mutation rate loci may be similar either because they exchange alleles frequently or because the mutation rate is large enough to swamp the effects of genetic drift (or because the populations have only recently diverged from one another; Felsenstein, 1982).

Tables 1 and 2 provide more details on the behavior of $\theta^{(I)}$ and $\theta^{(II)}$ for two extremes of local population size, $2N = 100$ and $2N = 10,000$, and for a variety of realistic migration and mutation parameters. Several additional observations emerge from examining these tables. First, equation (12) and the discussion that follow suggest that the expected value of $\theta^{(I)}$ at stationarity should not depend on mutation rate when $N\mu$ is small. Evaluation of the exact expression (10), on the other hand, shows that $\theta^{(I)}$ is strongly dependent on mutation rates in the range of 10^{-2} to 10^{-4} , which may be characteristic of microsatellite loci. Second, although the values of both $\theta^{(I)}$ and $\theta^{(II)}$ are influenced by the particular mutational model chosen, differences between values associated with the stepwise mutation model typically differ from those associated with the two-phase model by only a few percent, and in no case

are the differences greater than about 10%. Third, $\theta^{(I)}$ and $\theta^{(II)}$ are only weakly dependent on the size range (number) of alleles. Together these observations suggest that $\theta^{(I)}$ and $\theta^{(II)}$ are strongly influenced by the overall rate of mutation, but only weakly influenced by details of the mutational process. Finally, notice that values of $\theta^{(II)}$ are smaller than corresponding values of $\theta^{(I)}$, and that the differences can be substantial when k is small.

As the results in Tables 3 and 4 show, there is one striking difference between the behavior of R_{ST} as a function of local population sizes, migration rate, and mutational parameters and the behavior of $\theta^{(I)}$ and $\theta^{(II)}$ as functions of those same parameters: R_{ST} is not only relatively insensitive to the choice of mutational model (stepwise *versus* two-phase), it is also relatively insensitive to the overall rate of mutation. Moreover, the expected value of R_{ST} at stationarity is relatively close to the value predicted for a finite-island model when the range of allele sizes is unbounded (Rousset, 1996). As Table 5 shows, however, the relatively small differences in R_{ST} may mask larger differences in the value of Nma that would be inferred from them (compare Rousset, 1996), especially when the number of populations exchanging genes is small.

4. Discussion

The results presented above lead to several important conclusions regarding evolutionary analysis of microsatellite data. First, our results show that R_{ST} is sensitive to demographic parameters that determine the importance of gene flow (local population size, migration rate, and the number of populations in a metapopulation), but it is relatively insensitive to mutational parameters (mutation rates and stepwise *versus* two-phase mutational models). Thus, it provides a useful index of the degree to which populations are genetically isolated from one another.

Second, our results reinforce previous observations that the amount of genetic differentiation among contemporaneous populations is substantially less than the amount of genetic variation expected within any one population over evolutionary time (compare Fu et al. 2003; Holsinger in press). Because populations connected through gene flow tend to “drift” together, allele frequencies among contemporaneous populations are correlated with one another. Methods that ignore this correlation may substantially underestimate the extent of stochastic variation in allele frequencies (compare Song et al., 2004; Fu et al., 2005; Holsinger, in press). The effect of the among-population correlation is particularly pronounced when the number of populations exchanging genes (not the number of populations from which samples are available) is small.

Finally, our results show that while R_{ST} is relatively insensitive to mutational parameters, measures of among-population genetic differentiation that depend only on allele frequency, namely $\theta^{(I)}$ and $\theta^{(II)}$, depend quite sensitively on the overall mutation rate. This observation suggests that by taking into account the special mutational properties of microsatellite data, we may be able to develop inferential methods that allow us to make separate estimates of the contribution of mutation and migration to similarities and differences among populations that are geographically structured. Clearly, coalescent methods like those described in Beerli and Felsenstein (2001) allow such distinctions, but the differing properties of R_{ST} and $\theta^{(I)}/\theta^{(II)}$ suggest that it may be possible to estimate $N_e\mu$ and N_em directly from R_{ST} and θ , a topic of continuing investigation.

Of course, all of the results we present in this paper depend on the assumption that populations have reached stationarity with respect to mutation, migration, and drift. In real populations the assumption of stationarity will never be satisfied. In many cases it may not even be approximately correct. Nonetheless, the relationship between R_{ST} and coalescence

times (Slatkin, 1995) suggests that it remains a useful index of population differentiation and gene flow for microsatellite loci, provided we remember that “gene flow” may reflect either continuing migration of individuals among distinct populations or recent divergence of those populations from one another or any combination of those two.

Acknowledgments

This research was supported in part by a grant from the U.S. National Institutes of Health, 1 R01 GM068449-01A1.

Appendix A

We begin with the observation that the stationary covariances are given by

$$\Sigma_{11} = \left(1 - \frac{1}{2N}\right) \mathbf{V}' \{ (1 - r_k) \Sigma_{11} + r_k \Sigma_{12} \} \mathbf{V} + \frac{1}{2N} \left(\frac{1}{A} \mathbf{I}_A - \frac{1}{A^2} \mathbf{1}_A \mathbf{1}_A' \right)$$

and

$$\Sigma_{12} = \mathbf{V}' \left\{ \frac{r_k}{k-1} \Sigma_{11} + \left(1 - \frac{r_k}{k-1}\right) \Sigma_{12} \right\} \mathbf{V},$$

where Σ_{11} , Σ_{12} , and $\mathbf{V}$ are symmetric matrices. Rearranging we obtain

$$\Sigma_{11} = \left(1 - \frac{1}{2N}\right) \left\{ (1 - r_k) \mathbf{V}^2 \Sigma_{11} + r_k \mathbf{V}^2 \Sigma_{12} \right\} + \frac{1}{2N} \left(\frac{1}{A} \mathbf{I}_A - \frac{1}{A^2} \mathbf{1}_A \mathbf{1}_A' \right)$$

and

$$\Sigma_{12} = \frac{r_k}{k-1} \mathbf{V}^2 \Sigma_{11} + \left(1 - \frac{r_k}{k-1}\right) \mathbf{V}^2 \Sigma_{12}.$$

Thus,

$$\begin{aligned} \Sigma_{11} &= \frac{1}{2N} \left[\mathbf{I} - \left(1 - \frac{r_k}{k-1}\right) \mathbf{V}^2 \right] \mathbf{D}_1^{-1} \left(\frac{1}{A} \mathbf{I}_A - \frac{1}{A^2} \mathbf{1}_A \mathbf{1}_A' \right) \\ &= \frac{1}{2N} \mathbf{Q} \left[\mathbf{I} - \left(1 - \frac{r_k}{k-1}\right) \Lambda^2 \right] \mathbf{D}_1^{-1} \mathbf{Q}' \left(\frac{1}{A} \mathbf{I}_A - \frac{1}{A^2} \mathbf{1}_A \mathbf{1}_A' \right) \end{aligned}$$

and

$$\begin{aligned} \Sigma_{12} &= \frac{1}{2N} \frac{r_k}{k-1} \mathbf{V}^2 \mathbf{D}_1^{-1} \left(\frac{1}{A} \mathbf{I}_A - \frac{1}{A^2} \mathbf{1}_A \mathbf{1}_A' \right) \\ &= \frac{1}{2N} \frac{r_k}{k-1} \mathbf{Q} \Lambda^2 \mathbf{D}_1^{-1} \mathbf{Q}' \left(\frac{1}{A} \mathbf{I}_A - \frac{1}{A^2} \mathbf{1}_A \mathbf{1}_A' \right), \end{aligned}$$

where

$$\mathbf{D}_1 = \left[\mathbf{I} - \left(1 - \frac{1}{2N}\right) \Lambda^2 \right] \left[\mathbf{I} - \left(1 - \frac{k}{k-1} r_k\right) \Lambda^2 \right] - \frac{r_k}{2N} \Lambda^2,$$

$\Lambda = \text{diag}\{\lambda_1, \dots, \lambda_A\}$ and $\mathbf{Q}_{A \times A} = (\mathbf{q}_1, \dots, \mathbf{q}_A)$ with λ_j is the j^{th} eigenvalue of $\mathbf{V}$, with $\mathbf{q}_j$, the $A \times 1$ corresponding eigenvector. Further algebraic simplification yields equations (7) and (8).

Appendix B

To establish the results in the text it is sufficient to show that $\lambda_j = 1 - \mu + \mu \cos\left(\frac{(j-1)\pi}{A}\right)$ and $\mathbf{q}_j = \mathbf{q}_j^* / \|\mathbf{q}_j^*\|$ satisfy the characteristic function of $\mathbf{V}$. Following Gregory and Karney (1969) and Barnett (1990) it follows that this condition is equivalent to

$$q_{j1} + q_{j2} - 2\cos(2\alpha) q_{j1} = 0, \quad (\text{B.1})$$

$$q_{jk} + q_{j,k+2} - 2\cos(2\alpha) q_{j,k+1} = 0, \quad (\text{B.2})$$

for $k = 1, \dots, A-2$ and

$$q_{j,A-1} + q_{jA} - 2\cos(2\alpha) q_{jA} = 0, \quad (\text{B.3})$$

for $j = 1, \dots, A$. First, it is trivial for $j = 1$. Next, we verify (B.1), (B.2) and (B.3) for $j = 2, 3, \dots, A$.

$$\begin{aligned} \text{l.h.s. of (B.1)} &= \sin(2\alpha_j) + 2\cos(3\alpha_j) \sin(\alpha_j) - 2\cos(2\alpha_j) \sin(2\alpha_j) \\ &= \sin(2\alpha_j) + \sin(\alpha_j - 3\alpha_j) + \sin(\alpha_j + 3\alpha_j) - \sin(4\alpha_j) \\ &= \sin(2\alpha_j) + \sin(-2\alpha_j) \\ &= 0. \end{aligned}$$

Now, for $k = 1$, we have

$$\begin{aligned} \text{l.h.s. of (B.2)} &= \sin(2\alpha_j) + 2\cos(5\alpha_j) \sin(\alpha_j) - 4\cos(2\alpha_j) \cos(3\alpha_j) \sin(\alpha_j) \\ &= \sin(2\alpha_j) - \sin(4\alpha_j) + \sin(6\alpha_j) + 2\cos(2\alpha_j) (\sin(2\alpha_j) - \sin(4\alpha_j)) \\ &= \sin(2\alpha_j) - \sin(4\alpha_j) + \sin(6\alpha_j) + \sin(0) + \sin(4\alpha_j) - \sin(2\alpha_j) - \sin(6\alpha_j) \\ &= 0. \end{aligned}$$

The final steps in (B.2) and the identity in (B.3) follow from the standard trigonometric identity

$$2 \cos z_1 \cos z_2 = \cos(z_1 - z_2) + \cos(z_1 + z_2).$$

References

- Allan M, Eisen G, Pomp D. Genomic mapping of direct and correlated responses to long-term selection for rapid weight gain in mice. *Genetics*. 2005 105.041319.
- Barnett, S. *Matrices: Methods and Applications*. Oxford University Press; Oxford: 1990.
- Berli P, Felsenstein J. Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. *Proceedings of the National Academy of Sciences of the United States of America*. 2001; 98(8):4563–4568. [PubMed: 11287657]

- Calabrese PP, Durrett RT, Aquadro CF. Dynamics of microsatellite divergence under stepwise mutation and proportional slippage point mutation models. *Genetics*. 2001; 159:839–852. [PubMed: 11606557]
- Chistiakov DA, Hellemans B, Haley CS, Law AS, Tsigeneopoulos CS, Kotoulas G, Bertotto D, Libertini A, Volckaert FAM. A microsatellite linkage map of the European seabass *Dicentrarchus labrax* L. *Genetics*. 2005 104.039719.
- Cockerham CC. Variance of gene frequencies. *Evolution*. 1969; 23:72–84.
- Cockerham CC, Weir BS. Correlations, descent measures: Drift with migration and mutation. *Proceedings of the National Academy of Sciences USA*. 1987; 84:8512–8514.
- Crow JF, Aoki K. Group selection for a polygenic behavioral trait: estimating the degree of population subdivision. *Proceedings of the National Academy of Sciences USA*. 1984; 81:6073–6077.
- Di Rienzo A, Peterson AC, Garza JC, Valdes AM, Slatkin M, Freimer NB. Mutational processes of simple-sequence repeat loci in human populations. *Proceedings of the National Academy of Sciences USA*. 1994; 91:3166–3170.
- Estoup A, Jarne P, Cornuet J-M. Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. *Molecular Ecology*. 2002; 11:1591–1604. [PubMed: 12207711]
- Feldman MW, Bergman A, Pollock DD, Goldstein DB. Microsatellite genetic distances with range constraints: analytic description and problems of estimation. *Genetics*. 1997; 145:207–216. [PubMed: 9017402]
- Felsenstein J. How can we infer geography and history from gene frequencies? *Journal of Theoretical Biology*. 1982; 96:9–20. [PubMed: 7109659]
- Fu R, Gelfand AE, Holsinger KE. Exact moment calculations for genetic models with migration, mutation, and drift. *Theoretical Population Biology*. 2003; 63:231–243.
- Goldstein DB, Linares AR, Cavalli-Sforza LL, Feldman MW. An evaluation of genetic distances for use with microsatellite loci. *Genetics*. 1995; 139:463–471. [PubMed: 7705647]
- Gregory, RT.; Karney, DL. *A Collection of Matrices for Testing Computational Algorithms*. Wiley-Interscience; 1969. 1969
- Hall MC, Willis JH. Transmission ratio distortion in intraspecific hybrids of *Mimulus guttatus*: implications for genomic divergence. *Genetics*. 2005; 170:375–386. [PubMed: 15781698]
- Holsinger, KE. Bayesian hierarchical models in geographical genetics. In: Clark, JS.; Gelfand, AE., editors. *Applications of Computational Statistics in the Environmental Sciences*. Oxford University Press; New York, NY: in press
- Kretzer AM, Dunham S, Molina R, Spatafora JW. Patterns of vegetative growth and gene flow in *Rhizopogon vinicolor* and *R. vesiculosus* (Boletales, Basidiomycota). *Molecular Ecology*. 2005; 14:2259–2268. [PubMed: 15969712]
- Kai W, Kikuchi K, Fujita M, Suetake H, Fujiwara A, Yoshiura Y, Ototake M, Venkatesh B, Miyaki K, Suzuki Y. A genetic linkage map for the tiger pufferfish, *Takifugu rubripes*. *Genetics*. 2005 105.042051.
- Malécot, G. *Les Mathématiques de l'Hérédité*. Masson et Cie; Paris: 1948.
- Minvielle F, Kayang BB, Inoue-Murayama M, Miwa M, Vignal A, Gourichon D, Neau A, Monvoisin JL, Ito S. Microsatellite mapping of QTL affecting growth, feed consumption, egg production, tonic immobility and body temperature of Japanese quail. *BMC Genomics*. 2005; 6:87. [PubMed: 15941487]
- Nagylaki T. Geographical invariance in population genetics. *Journal of Theoretical Biology*. 1983; 99:159–172. [PubMed: 7169795]
- Ohta T, Kimura M. A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. *Genetical Research*. 1973; 22:201–204. [PubMed: 4777279]
- Pollock DD, Bergman A, Feldman MW, Goldstein DB. Microsatellite behavior with range constraints: parameter estimation and improved distances for use in phylogenetic reconstruction. *Theoretical Population Biology*. 1998; 53:256–271. [PubMed: 9679321]

- Rousset F. Equilibrium values of measures of population subdivision for stepwise mutation processes. *Genetics*. 1996; 142:1357–1362. [PubMed: 8846911]
- Rousset F. Genetic differentiation in populations with different classes of individuals. *Theoretical Population Biology*. 1999; 55:297–308.
- Rousset, F. Inferences from spatial population genetics. In: Balding, DJ.; Bishop, M.; Cannings, C., editors. *Handbook of Statistical Genetics*. John Wiley & Sons; Chichester: 2001. p. 239-269.
- Slatkin M. A measure of population subdivision based on microsatellite allele frequencies. *Genetics*. 1995; 139:457–462. [PubMed: 7705646]
- Song S, Dey DK, Holsinger KE. Hierarchical models with migration, mutation, and drift: implications for genetic inference. *Evolution*. 2006; 60:1–12. [PubMed: 16568626]
- Tautz, D. Note on the definition and nomenclature of tandemly repetitive DNA sequences. In: Pena, SDJ.; Eplen, JT.; Jeffreys, AJ., editors. *DNA Fingerprinting: State of the Science*. Birkhauser Verlag; Basel: 1993. p. 21-28.
- Wehrhahn CF. The evolution of selectively similar electrophoretically detectable alleles in finite natural populations. *Genetics*. 1975; 80:375–394. [PubMed: 1132689]
- Weir, BS. *Genetic Data Analysis II*. Sinauer Associates; Sunderland, MA: 1996.
- Weir BS, Cockerham CC. Estimating F -statistics for the analysis of population structure. *Evolution*. 1984; 38:1358–1370.
- Weir BS, Hill WG. Estimating F -statistics. *Annual Reviews of Genetics*. 2002; 36:721–750.
- Wright S. The genetical Structure of populations. *Annals of Eugenics*. 1951; 15:323–354.
- Xu H, Chakraborty R, Fu Y-X. Mutation rate variation at human dinucleotide microsatellites. *Genetics*. 2005; 170:305–312. [PubMed: 15716501]

Highlights

We provide exact expressions for the moments of the allele frequency distribution in a stochastic model appropriate for studying microsatellite evolution with migration, mutation, and drift under the assumption that the range of allele sizes is bounded.

We study the behavior of several measures related to Wright's F_{st} , including Slatkin's R_{st} . Our results show that familiar relationships between N_m and F_{st} or R_{st} hold when migration and mutation rates are small.

Figure 1.
Plots of $\theta^{(I)}$ and $\theta^{(II)}$ v.s. m and μ with $2N = 100$.

Behavior of $\theta^{(l)}$ and $\theta^{(l)}$ as a function of k , A , m , and mutational parameters when $2N = 100$. The subscript SMM refers to results for the stepwise mutation model. The subscript 10 refers to results from the two-phase model with $\phi = 0$ and $\alpha = 0.1$. The subscript 90 refers to results from the two-phase model with $\phi = 0$ and $\alpha = 0.9$.

Table 1

k	A	m	μ	$\theta_{SMM}^{(l)}$	$\theta_{10}^{(l)}$	$\theta_{90}^{(l)}$	$\theta_{SMM}^{(m)}$	$\theta_{10}^{(m)}$	$\theta_{90}^{(m)}$
5	10	0.1	0.01	0.20719	0.25370	0.18154	0.03055	0.03139	0.03150
			0.001	0.57706	0.67914	0.65222	0.03232	0.03244	0.03246
			0.0001	0.91358	0.94881	0.94804	0.03255	0.03256	0.03256
		0.01	0.01	0.29395	0.34922	0.29359	0.15761	0.18282	0.18464
			0.001	0.62695	0.70975	0.68321	0.22827	0.23543	0.23707
			0.0001	0.91647	0.94977	0.94883	0.24304	0.24397	0.24419
	50	0.001	0.01	0.39728	0.48408	0.44163	0.31972	0.39721	0.37635
			0.001	0.76294	0.81812	0.80356	0.63171	0.68162	0.69078
			0.0001	0.93523	0.95725	0.95532	0.74525	0.75367	0.75567
		0.1	0.01	0.23408	0.29433	0.21945	0.03046	0.03143	0.03154
			0.001	0.58506	0.70155	0.67689	0.03231	0.03244	0.03246
			0.0001	0.91392	0.95289	0.95220	0.03254	0.03256	0.03256
100	10	0.01	0.01	0.31682	0.38499	0.32580	0.15500	0.18443	0.18703
			0.001	0.63455	0.72958	0.70513	0.22743	0.23573	0.23732
			0.0001	0.91688	0.95376	0.95291	0.24293	0.24400	0.24421
		0.001	0.01	0.41302	0.51493	0.47196	0.31307	0.40536	0.39321
			0.001	0.76745	0.83014	0.81614	0.62620	0.68408	0.69331
			0.0001	0.93590	0.96049	0.95874	0.74424	0.75398	0.75592
	50	0.1	0.01	0.06174	0.06767	0.05600	0.04593	0.04740	0.04740
			0.001	0.15748	0.18610	0.12794	0.04908	0.04926	0.04928
			0.0001	0.45111	0.54384	0.49488	0.04945	0.04947	0.04948
		0.01	0.01	0.22611	0.26144	0.25009	0.21637	0.24937	0.24466
			0.001	0.37258	0.39648	0.36177	0.31132	0.31999	0.32059
			0.0001	0.57028	0.62966	0.58784	0.32964	0.33077	0.33099
100	0.001	0.01	0.39137	0.47652	0.43990	0.38762	0.47221	0.43679	0.43679
		0.001	0.72735	0.77196	0.76708	0.71809	0.76200	0.76183	0.76183
		0.0001	0.93590	0.96049	0.95874	0.74424	0.75398	0.75592	0.75592

k	A	m	μ	$\theta_{SMM}^{(j)}$	$\theta_{10}^{(j)}$	$\theta_{90}^{(j)}$	$\theta_{SMM}^{(j)}$	$\theta_{10}^{(j)}$	$\theta_{90}^{(j)}$
50		0.1	0.0001	0.84835	0.86139	0.84932	0.81713	0.82355	0.82435
			0.01	0.08378	0.09584	0.06601	0.04586	0.04751	0.04755
			0.001	0.18772	0.22744	0.15971	0.04906	0.04927	0.04929
			0.0001	0.46463	0.57335	0.52768	0.04945	0.04947	0.04948
			0.01	0.24108	0.28438	0.26194	0.21575	0.25343	0.25055
			0.001	0.39449	0.42630	0.38248	0.31048	0.32042	0.32122
			0.0001	0.58191	0.65268	0.61317	0.32951	0.33081	0.33103
			0.01	0.40289	0.50221	0.46659	0.39405	0.49237	0.46289
			0.001	0.73267	0.78320	0.77510	0.71524	0.76512	0.76649
			0.0001	0.85313	0.86904	0.85627	0.81644	0.82384	0.82469

Behavior of $\theta^{(I)}$ and $\theta^{(II)}$ as a function of k , A , m , and mutational parameters when $2N = 10,000$. Refer to Table 1 for an explanation of the subscripts.

Table 2

k	A	m	μ	$\theta_{SMM}^{(I)}$	$\theta_{10}^{(I)}$	$\theta_{90}^{(I)}$	$\theta_{SMM}^{(II)}$	$\theta_{10}^{(II)}$	$\theta_{90}^{(II)}$
5	10	0.1	0.01	0.00393	0.00499	0.00221	0.00031	0.00032	0.00032
			0.001	0.03273	0.04174	0.01884	0.00033	0.00033	0.00033
			0.0001	0.18458	0.23256	0.15770	0.00033	0.00033	0.00033
		0.01	0.01	0.00558	0.00696	0.00412	0.00198	0.00231	0.00224
	50	0.1	0.001	0.03521	0.04426	0.02148	0.00297	0.00307	0.00307
			0.0001	0.18658	0.23435	0.15973	0.00318	0.00319	0.00319
			0.01	0.01018	0.01293	0.00790	0.00665	0.00836	0.00603
			0.001	0.05002	0.06175	0.03949	0.01898	0.02219	0.02176
	100	0.1	0.0001	0.20440	0.25092	0.17870	0.02863	0.02961	0.02974
			0.01	0.01344	0.01782	0.00228	0.00031	0.00032	0.00032
			0.001	0.06203	0.07874	0.03543	0.00033	0.00033	0.00033
			0.0001	0.21222	0.27440	0.19694	0.00033	0.00033	0.00033
100	10	0.1	0.01	0.01506	0.01979	0.00433	0.00199	0.00236	0.00231
			0.001	0.06440	0.08114	0.03801	0.00297	0.00307	0.00308
			0.0001	0.21416	0.27608	0.19885	0.00317	0.00319	0.00319
		0.01	0.01	0.01999	0.02637	0.00977	0.00723	0.00727	0.00937
	50	0.1	0.001	0.07839	0.09796	0.05595	0.01880	0.02251	0.02230
			0.0001	0.23140	0.29167	0.21666	0.02852	0.02965	0.02979
			0.01	0.00066	0.00073	0.00059	0.00048	0.00049	0.00049
			0.001	0.00233	0.00286	0.00146	0.00051	0.00051	0.00051
	100	0.1	0.0001	0.01768	0.02256	0.00986	0.00051	0.00052	0.00052
			0.01	0.00300	0.00357	0.00330	0.00282	0.00333	0.00320
			0.001	0.00629	0.00699	0.00559	0.00448	0.00465	0.00465
		0.01	0.0001	0.02187	0.02673	0.01414	0.00485	0.00487	0.00487
100	10	0.1	0.01	0.00913	0.01150	0.00784	0.00895	0.01126	0.00774
			0.001	0.02877	0.03417	0.03170	0.02707	0.03197	0.03080
			0.0001	0.05855	0.06460	0.05296	0.04282	0.04442	0.04442
		0.01	0.01	0.00066	0.00073	0.00059	0.00048	0.00049	0.00049
	50	0.1	0.001	0.00233	0.00286	0.00146	0.00051	0.00051	0.00051
			0.0001	0.01768	0.02256	0.00986	0.00051	0.00052	0.00052
			0.01	0.00300	0.00357	0.00330	0.00282	0.00333	0.00320
			0.001	0.00629	0.00699	0.00559	0.00448	0.00465	0.00465
	100	0.1	0.0001	0.02187	0.02673	0.01414	0.00485	0.00487	0.00487
			0.01	0.00913	0.01150	0.00784	0.00895	0.01126	0.00774
			0.001	0.02877	0.03417	0.03170	0.02707	0.03197	0.03080
			0.0001	0.05855	0.06460	0.05296	0.04282	0.04442	0.04442

k	A	m	μ	$\theta_{SMM}^{(j)}$	$\theta_{10}^{(j)}$	$\theta_{90}^{(j)}$	$\theta_{SMM}^{(j)}$	$\theta_{10}^{(j)}$	$\theta_{90}^{(j)}$
50		0.1	0.01	0.00131	0.00168	0.00072	0.00048	0.00049	0.00049
			0.001	0.00786	0.01055	0.00260	0.00051	0.00051	0.00051
			0.0001	0.04122	0.05245	0.02050	0.00051	0.00052	0.00052
		0.01	0.01	0.00368	0.00462	0.00355	0.00285	0.00344	0.00332
			0.001	0.01177	0.01463	0.00675	0.00490	0.00467	0.00467
			0.0001	0.04527	0.05645	0.02470	0.00485	0.00488	0.00488
		0.001	0.01	0.01099	0.01431	0.00981	0.01020	0.01319	0.00959
			0.001	0.03419	0.04227	0.03389	0.02732	0.03292	0.03193
			0.0001	0.08060	0.09279	0.06297	0.04274	0.04454	0.04459

Behavior of R_{ST} under different mutational models when $2N = 100$. $R_{ST,L} = 1/(4Nm\alpha + 1)$, where $\alpha = k/(k - 1)$. The subscript SMM refers to the stepwise mutation model. The numerical subscripts, K , refer to the two-phase model with $\phi = 0$ and $\alpha = K/100$.

Table 3

k	A	m	μ	$R_{ST,L}$	R_{ST}	$R_{ST,10}$	$R_{ST,50}$	$R_{ST,90}$
5	10	0.1	0.01	0.03846	0.03251	0.03254	0.03238	0.03210
			0.001		0.03262	0.03262	0.03261	0.03259
			0.0001		0.03263	0.03264	0.03263	0.03263
		0.01	0.01	0.28571	0.23721	0.23879	0.22814	0.21107
			0.001		0.24418	0.24433	0.24356	0.24167
		0.001	0.0001		0.24523	0.24523	0.24520	0.24503
			0.01	0.80000	0.69443	0.70790	0.61446	0.50037
			0.001		0.75457	0.75614	0.74634	0.72728
			0.0001		0.76277	0.76293	0.76232	0.76052
			0.01	0.03846	0.03262	0.03262	0.03261	0.03259
50	10	0.1	0.01		0.03263	0.03263	0.03263	0.03263
			0.001		0.03264	0.03264	0.03264	0.03264
		0.01	0.01	0.28571	0.24442	0.24455	0.24381	0.24268
			0.001		0.24518	0.24519	0.24510	0.24480
		0.001	0.0001		0.24536	0.24536	0.24536	0.24531
			0.01	0.80000	0.75816	0.75924	0.75384	0.74623
			0.001		0.76276	0.76289	0.76198	0.76016
			0.0001		0.76391	0.76392	0.76383	0.76346
		0.01	0.01	0.04717	0.04929	0.04933	0.04900	0.04844
			0.001		0.04948	0.04949	0.04945	0.04940
100	10	0.1	0.01		0.04950	0.04950	0.04950	0.04949
			0.0001		0.32106	0.32327	0.30641	0.28134
		0.01	0.01	0.33110	0.33099	0.33124	0.32947	0.32641
			0.001		0.33206	0.33209	0.33195	0.33165
		0.001	0.01	0.83193	0.76564	0.77831	0.68440	0.56754
			0.001		0.82450	0.82602	0.81454	0.79562
			0.0001		0.83116	0.83133	0.83025	0.82828
		0.001	0.01		0.04950	0.04950	0.04950	0.04949
			0.0001		0.32106	0.32327	0.30641	0.28134
			0.01	0.33110	0.33099	0.33124	0.32947	0.32641
			0.001		0.33206	0.33209	0.33195	0.33165

k	A	m	μ	$R_{ST,L}$	R_{ST}	$R_{ST,10}$	$R_{ST,50}$	$R_{ST,90}$
50		0.1	0.01	0.04717	0.04949	0.04950	0.04948	0.04946
			0.001		0.04950	0.04950	0.04950	0.04950
			0.0001		0.04950	0.04950	0.04950	0.04950
		0.01	0.01	0.33110	0.33162	0.33177	0.33101	0.32977
			0.001		0.33211	0.33213	0.33203	0.33188
			0.0001		0.33220	0.33220	0.33219	0.33215
		0.001	0.01	0.83193	0.82866	0.82961	0.82479	0.82097
			0.001		0.83154	0.83164	0.83113	0.83033
			0.0001		0.83191	0.83193	0.83185	0.83172

Table 4

Behavior of R_{ST} under different mutational models when $2N = 10,000$. $R_{ST,L} = 1/(4Nma + 1)$, where $a = k/(k - 1)$. The subscript SMM refers to the stepwise mutation model. The numerical subscripts, K , refer to the two-phase model with $\phi = 0$ and $\alpha = K/100$.

k	A	m	μ	$R_{ST,L}$	R_{ST}	$R_{ST,10}$	$R_{ST,50}$	$R_{ST,90}$
5	10	0.1	0.01	0.00040	0.00033	0.00033	0.00033	0.00033
			0.001		0.00033	0.00033	0.00033	0.00033
			0.0001		0.00033	0.00033	0.00033	0.00033
	0.01	0.01	0.00398	0.00308	0.00310	0.00291	0.00263	
			0.001		0.00319	0.00319	0.00317	0.00314
			0.0001		0.00320	0.00320	0.00320	0.00320
	0.001	0.01	0.03846	0.02230	0.02363	0.01539	0.00969	
			0.001		0.02979	0.03003	0.02820	0.02553
			0.0001		0.03088	0.03090	0.03072	0.03039
	50	0.1	0.01	0.00040	0.00033	0.00033	0.00033	0.00033
0.001				0.00033	0.00033	0.00033	0.00033	
0.0001				0.00033	0.00033	0.00033	0.00033	
0.01		0.01	0.00398	0.00320	0.00320	0.00319	0.00319	
			0.001		0.00320	0.00320	0.00320	0.00320
			0.0001		0.00320	0.00320	0.00320	0.00320
0.001		0.01	0.03846	0.03047	0.03062	0.02986	0.02961	
			0.001		0.03094	0.03096	0.03087	0.03074
			0.0001		0.03101	0.03101	0.03099	0.03097
100		10	0.1	0.01	0.00049	0.00051	0.00051	0.00051
	0.001				0.00052	0.00052	0.00051	0.00051
	0.0001				0.00052	0.00052	0.00052	0.00052
	0.01	0.01	0.00493	0.00466	0.00471	0.00435	0.00386	
			0.001		0.00488	0.00488	0.00484	0.00477
			0.0001		0.00490	0.00490	0.00490	0.00489
	0.001	0.01	0.04717	0.03170	0.03386	0.02087	0.01268	
			0.001		0.04453	0.04496	0.04165	0.03700
			0.0001		0.04650	0.04655	0.04618	0.04556

k	A	m	μ	$R_{ST,L}$	R_{ST}	$R_{ST,10}$	$R_{ST,50}$	$R_{ST,90}$
50		0.1	0.01	0.00049	0.00052	0.00052	0.00052	0.00051
			0.001		0.00052	0.00052	0.00052	0.00052
			0.0001		0.00052	0.00052	0.00052	0.00052
		0.01	0.01	0.00493	0.00489	0.00489	0.00488	0.00486
			0.001		0.00490	0.00490	0.00490	0.00490
			0.0001		0.00490	0.00490	0.00490	0.00490
		0.001	0.01	0.04716	0.04578	0.04606	0.04477	0.04308
			0.001		0.04664	0.04667	0.04651	0.04633
			0.0001		0.04673	0.04673	0.04672	0.04669

Table 5

Comparison of Nma that would be inferred from stationary values of $R_{ST}(\widehat{Nma})$ and Nma for several combinations of N , m , and k with $A = 10$, $\mu = 0.001$, and $\gamma = (0.1, 0.5, 0.9)$. SMM in subscripts refers to the stepwise mutation model. The numerical subscripts, K , refer to the two-phase model with $\phi = 0$ and $\alpha = K/100$.

$2N$	m	k	Nma	$\widehat{Nma}_{SMM}$	$\widehat{Nma}_{10}$	$\widehat{Nma}_{50}$	$\widehat{Nma}_{90}$
100	0.001	100	0.0505	0.0532	0.0527	0.0569	0.0642
		5	0.0625	0.0813	0.0806	0.0850	0.0937
	0.01	100	0.505	0.505	0.505	0.509	0.516
		5	0.625	0.787	0.773	0.776	0.784
	0.1	100	5.05	4.80	4.80	4.81	4.81
		5	6.25	7.41	7.41	7.42	7.42
10000	0.001	100	5.05	5.36	5.31	5.75	6.51
		5	6.25	8.14	8.08	8.62	9.54
	0.01	100	50.5	50.6	50.6	51.4	52.2
		5	62.5	78.1	78.1	78.6	79.4