

May 2003

Exact moment calculations for genetic models with migration, mutation, and drift

Rongwei Fu

University of Connecticut

Alan E. Gelfand

University of Connecticut

Kent E. Holsinger

University of Connecticut, kent.holsinger@uconn.edu

Follow this and additional works at: http://digitalcommons.uconn.edu/eeb_articles

Recommended Citation

Fu, Rongwei ; Gelfand, Alan E.; and Holsinger, Kent E., "Exact moment calculations for genetic models with migration, mutation, and drift" (2003). *EEB Articles*. 3.

http://digitalcommons.uconn.edu/eeb_articles/3

Exact moment calculations for genetic models with migration, mutation, and drift¹

Rongwei Fu*, Alan E. Gelfand*, and Kent E. Holsinger†

*Department of Statistics, U-4120, University of Connecticut, Storrs, CT 06296-4120

†Department of Ecology & Evolutionary Biology, U-3043, University of Connecticut, Storrs, CT 06269-3043

Abstract

Using properties of moment stationarity we develop exact expressions for the mean and covariance of allele frequencies at a single locus for a set of populations subject to drift, mutation, and migration. Some general results can be obtained even for arbitrary mutation and migration matrices, for example: (1) Under quite general conditions, the mean vector depends only on mutation rates, not on migration rates or the number of populations. (2) Allele frequencies covary among all pairs of populations connected by migration. As a result, the drift, mutation, migration process is not ergodic when any finite number of populations is exchanging genes. In addition, we provide closed form expressions for the mean and covariance of allele frequencies in Wright's finite-island model of migration under several simple models of mutation, and we show that the correlation in allele frequencies among populations can be very large for realistic rates of mutation unless an enormous number of populations are exchanging genes. As a result, the traditional diffusion approximation provides a poor approximation of the stationary distribution of allele frequencies among populations. Finally, we discuss some implications of our results for measures of population structure based on Wright's F -statistics.

1 Introduction

Most plant and animal species consist of many populations between which genetic exchange is limited. Since Wright (1931) first illustrated that limited gene exchange can lead to non-adaptive divergence among populations as a result of genetic drift, analytical and simulation studies of how local population size, migration rate, and mutation rate interact to affect the genetic structure of local populations and the degree of differentiation among them have been a mainstay of theoretical population genetics (e.g., Malécot 1948; Kimura and Weiss 1964; Nei and Feldman 1972; Felsenstein 1975; Nagylaki 1976; Maruyama 1977; Crow and Aoki 1984; Slatkin 1991). Interest has centered both on predicting the consequences of known rates of mutation and migration and on making inferences about those rates from

¹*Theoretical Population Biology* (in press, final revision 8 August 2002)

the pattern of genetic variation found in samples from natural populations (e.g., Slatkin and Barton 1989; Cockerham and Weir 1993; Kuhner et al. 1995; Beerli and Felsenstein 1999).

Wright's (1931) initial analyses of the process were heuristic and informal. Malécot's (1948) analysis, using probabilities of identity-by-descent, was the first formal analysis of the problem. Since then several other authors have used moment stationarity or stationarity of identity-by-descent measures to investigate similar problems (e.g., Kimura and Weiss 1964, stepping-stone models; Crow and Kimura 1970, single-population models of drift and mutation, p. 440; Cockerham and Weir 1987, finite-island models; Rousset 2001). Kimura (1964) introduced a second modeling framework in which a partial differential equation is used to approximate the Markov chain describing the Wright-Fisher drift process with migration and mutation. Analysis of this model for one locus with two alleles showed that the stationary distribution attained is a beta distribution with parameters that Wright intuited correctly (see Crow and Kimura 1970 and Ewens 1979 for extensive reviews of results obtained from diffusion models). More recently, analyses of models derived from Kingman's (1982a, 1982b) coalescent have received much attention (e.g., Takahata 1989; Notohara 1990, 2000, 2001; Slatkin 1991; Bahlo and Griffiths 2000, 2001).

While both diffusion and coalescent models have led to great insights, they rely on approximations. Diffusion models use a Markov process with a continuous state space to approximate a Markov chain with a discrete state space. The discrete-time model from which the coalescent process is derived as a continuous limit assumes that the population size is large enough that only one coalescent event can occur in any generation and that a migration event and a coalescent event cannot happen simultaneously. In this paper we use moment stationarity to derive exact expressions for the mean and covariance structure of discrete-time models that include arbitrary patterns and rates of mutation, arbitrary numbers of alleles, arbitrary patterns and rates of migration, arbitrary numbers of populations, and arbitrary (including variable) local population sizes. We illustrate how familiar analytical solutions emerge for two simple models (the finite-island and the one-dimensional stepping stone), and show how those results are affected for a model including stochastic variation in local population sizes (assuming that a stationary distribution of local population size exists).

One particularly striking result emerges from these analyses: The mutation, migration, drift process is not ergodic when any finite number of populations is exchanging genes. Moreover, when mutation is rare, the correlation in allele frequency among populations at any one time does not approach zero unless thousands or tens of thousands of populations are exchanging genes. Under these circumstances our simulations illustrate that the diffusion approximation provides a very poor approximation to the stationary distribution of allele frequencies. We conclude the paper by discussing the implications of these results for measures of population structure based on Wright's F -statistics.

2 Theoretical Results and Examples

Our approach is to model allele frequencies at the population level. Focusing on a single locus, assume that we have A allele types, b_1, b_2, \dots, b_A and k populations indexed by i . Let $\mathbf{V}_{A \times A}$, be a general mutation matrix, i.e. \mathbf{V}_{rs} is the probability of mutation from allele type b_r to allele type b_s . So \mathbf{V} is row stochastic. Let $\mathbf{M}_{k \times k}$ be a general (backward) migration matrix, i.e. $\mathbf{M}_{ij} = \overleftarrow{m}_{ij}$ is the probability that the allele in population i came from population j (compare Nagylaki 1982; Rousset 1999, 2001). So \mathbf{M} is also row stochastic. Let $\mathbf{p}_i^{(t)}$ be the $A \times 1$ vector of allele frequencies in population i at generation t ($\mathbf{1}'\mathbf{p}_i^{(t)} = 1$). Then let $\mathbf{P}^{(t)}$ be the $k \times A$ matrix whose i^{th} row is $(\mathbf{p}_i^{(t)})'$. So $\mathbf{P}^{(t)}$ is, as well, row stochastic.

We propose a first order stationary Markov transition model to provide the generation to generation transition in allele frequencies. In particular we specify the distribution of $\mathbf{P}^{(t+1)}$ given $\mathbf{P}^{(t)}$ as follows: let

$$\mathbf{P}^{*(t)} = \mathbf{M}\mathbf{P}^{(t)}\mathbf{V} \quad (1)$$

That is, $\mathbf{P}^{*(t)}$ is a deterministic function of $\mathbf{P}^{(t)}$. $\mathbf{P}^{*(t)}$ is, evidently, row stochastic. The rows of $\mathbf{P}^{*(t)}$, $(\mathbf{p}_i^{*(t)})'$ provide the allele frequency vectors for the stochastic part of specification. We note that (1) is invariant to whether mutation precedes migration or vice versa. That is, in either case the contribution to $p_{ir}^{*(t)}$ from $p_{js}^{(t)}$ is $\overleftarrow{m}_{ij}\nu_{sr}$.

Suppose population i is of size N_i , where N_i is the number of alleles in the i^{th} population. With N_i diploid individuals, N_i is replaced by $2N_i$. In analytical solutions common N_i across i is often assumed (see, for example, Crow and Kimura 1970; Crow and Aoki 1984; see Rousset 2001 for a more general treatment). We allow N_i to vary across populations. In fact, at the end of this section, we allow N_i to vary randomly across generations. Then, given $\mathbf{P}^{*(t)}$ and N_i , the stochastic specification assumes that the $\mathbf{p}_i^{(t+1)}$ are conditionally independent and

$$2N_i\mathbf{p}_i^{(t+1)} \sim \text{Mult}(2N_i, \mathbf{p}_i^{*(t)}) \quad (2)$$

Through (1) and (2) we pass from $\mathbf{P}^{(t)} \rightarrow \mathbf{P}^{*(t)} \rightarrow \mathbf{P}^{(t+1)}$.

In the preceding formulation we assembled the $\mathbf{p}_i^{(t)}$ into a $k \times A$ matrix $\mathbf{P}^{(t)}$. An alternative is to concatenate the $\mathbf{p}_i^{(t)}$ to a $kA \times 1$ vector $\mathbf{p}^{(t)}$. Using the foregoing notation it follows that

$$\mathbf{p}^{*(t)} = (\mathbf{M} \otimes \mathbf{V}')\mathbf{p}^{(t)} \quad (3)$$

where \otimes denotes the Kronecker product. If $\mathbf{B}(\mathbf{M}, \mathbf{V}) \equiv \mathbf{M} \otimes \mathbf{V}'$ we can, for convenience, write $\mathbf{p}^{*(t)} = \mathbf{B}(\mathbf{M}, \mathbf{V})\mathbf{p}^{(t)}$. \mathbf{B} is naturally partitioned into $A \times A$ blocks which we denote by \mathbf{B}_{ij} . In fact, from (3), $\mathbf{B}_{ij} = \overleftarrow{m}_{ij}\mathbf{V}'$. So, in general, the \mathbf{B}_{ij} are not symmetric.

2.1 Stationary equations for means

The Markov chain defined through (1) and (2) has a finite state space. If all entries of \mathbf{V}^t are nonzero for some $t \geq 1$, then this chain has no absorbing states. In fact, it is aperiodic and irreducible and thus has a unique stationary distribution (compare Ewens 1979). We now develop first and second moments of this stationary distribution. At stationarity, the distribution for $\mathbf{P}^{(t)}$ is the same for all t . Thus, let $\mathbf{U}_{k \times A}$ denote the stationary mean matrix, i.e. $E(\mathbf{P}^{(t)}|\mathbf{M}, \mathbf{V}, \mathbf{N})$ where \mathbf{N} is the vector of population sizes. Also, let $\mathbf{\Sigma}$, $kA \times kA$ be the stationary covariance matrix, i.e., $\mathbf{\Sigma}_{\mathbf{P}^{(t)}|\mathbf{M}, \mathbf{V}, \mathbf{N}}$. The stationary mean matrix is calculated iteratively. Specifically, the law of iterated expectations says that for random variables X and Y , provided expectations exist, $E(X) = E(E(X|Y))$ (see, for example, Casella and Berger 1990, p. 154.). Thus,

$$E(\mathbf{P}^{(t+1)}|\mathbf{M}, \mathbf{V}, \mathbf{N}) = E\left(E(\mathbf{P}^{(t+1)}|\mathbf{M}, \mathbf{V}, \mathbf{N}, \mathbf{P}^{*(t)})\right) .$$

Because $E(\mathbf{P}^{(t+1)}|\mathbf{M}, \mathbf{V}, \mathbf{N}, \mathbf{P}^{*(t)}) = \mathbf{P}^{*(t)}$ and $E(\mathbf{P}^{*(t)}) = \mathbf{M}E(\mathbf{P}^{(t)}|\mathbf{M}, \mathbf{V}, \mathbf{N})\mathbf{V}$, it follows immediately that

$$E(\mathbf{P}^{(t+1)}|\mathbf{M}, \mathbf{V}, \mathbf{N}) = \mathbf{M}E(\mathbf{P}^{(t)}|\mathbf{M}, \mathbf{V}, \mathbf{N})\mathbf{V} .$$

Thus, at stationarity

$$\mathbf{U} = \mathbf{M}\mathbf{U}\mathbf{V} . \quad (4)$$

Notice that, as expected since the deterministic processes affecting allele frequencies do not depend on population size, \mathbf{U} does not depend on \mathbf{N} . \mathbf{U} is also row stochastic. Expression (4) is attractive in its simplicity but explicit solution of this linear system is not possible.

Assume that $\mathbf{U}_{k \times A} = \mathbf{1}_{k \times 1}\mathbf{u}_{\mathbf{A}1 \times A}$. Then, since \mathbf{M} is row stochastic, (4) can be rewritten as

$$\mathbf{1}_{k \times 1}\mathbf{u}_{\mathbf{A}} = \mathbf{1}_{k \times 1}\mathbf{u}_{\mathbf{A}}\mathbf{V} ,$$

i.e.,

$$\mathbf{1}_{k \times 1}\mathbf{u}_{\mathbf{A}}(\mathbf{I} - \mathbf{V}) = \mathbf{0}_{k \times A} . \quad (5)$$

Obviously, $\mathbf{u}_{\mathbf{A}}$ satisfying

$$\mathbf{u}_{\mathbf{A}}(\mathbf{I} - \mathbf{V}) = \mathbf{0}_{1 \times A} \quad (6)$$

is a solution of (5). Moreover, $\mathbf{u}_{\mathbf{A}}$ is the left eigenvector of \mathbf{V} corresponding to an eigenvalue of 1, and the Perron-Frobenius theorem (Gantmacher 1960, p. 53) guarantees that $\mathbf{u}_{\mathbf{A}}$ is unique (up to normalization to $\mathbf{1}'\mathbf{u}_{\mathbf{A}} = 1$) with all positive elements whenever \mathbf{V} is irreducible. For example, under a constant mutation rate matrix, i.e., $\nu_{ll} = 1 - u$, $\nu_{ll'} = u/(A - 1)$, we immediately can obtain $\mathbf{u}_{\mathbf{A}} = \frac{1}{A}\mathbf{1}_{A \times 1}$. When there are only two alleles ($A = 2$), $u_1 = \frac{\nu_{21}}{1 - (\nu_{11} - \nu_{21})}$. In more customary notation, if $\nu_{12} = \nu_1$, $\nu_{21} = \nu_2$, we obtain $u_1 = \frac{\nu_2}{(\nu_1 + \nu_2)}$.

Thus, all populations have the same stationary mean vector when \mathbf{V} is irreducible and this vector does not depend upon \mathbf{M} or even upon the number of populations. In other words, the vector describing expected allele frequencies is the same in all populations and

does not depend on the pattern of migration, the fraction of any population composed of migrants, or the number of populations among which migration is happening. Indeed, it does not depend on whether any or all of the populations are completely isolated. The stationary vector of allele frequencies in each population depends only on the pattern and magnitude of mutation rates. This is in distinct contrast to patterns of identity by descent, in which both migration and mutation rates are involved in the stationary equations (Rousset 2001).

In the absence of mutation, $\mathbf{V} = \mathbf{I}$ and $\mathbf{U}_A = \mathbf{1}_{k \times 1} \mathbf{u}'_A$ solves (6) for any \mathbf{u}_A . This is expected. In the absence of mutation, the chain in (1) and (2) has absorbing states, e.g. $p_{il}^{(t)} = 1$ for some l and all i . It has no stationary distribution. In analyzing this case in the literature, one either studies fixation time properties of the model or specifies a fixed population allele frequency vector as a constant input to all populations as in the diffusion approximation (see, for example, Crow and Kimura 1970 and the related discussion in the section below on comparison of exact results with the diffusion approximations).

We note that, passing to stationarity, the identical results can be obtained using (3). Specifically, let $\mathbf{u} = E(\mathbf{p}^{(t)} | \mathbf{M}, \mathbf{V}, \mathbf{N})$. Then we obtain the general linear system equivalent to (4), i.e.,

$$\mathbf{u} = \mathbf{B}(\mathbf{M}, \mathbf{V})\mathbf{u} \quad . \quad (7)$$

Analogous to (6), it can be shown that the solution to this equation is given by $\mathbf{u} = \mathbf{1}_{k \times 1} \otimes \mathbf{u}_{A \times 1}$ satisfying $(\mathbf{I} - \mathbf{V}')\mathbf{u}_A = \mathbf{0}$.

2.2 Stationary equations for covariances

Turning to stationary covariances, it is most convenient to work with (3). Using iterated expectations, which for variance says that for random variables X and Y $\text{Var}(X) = \text{Var}(E(X|Y)) + E(\text{Var}(X|Y))$ provided that the expectations exist (see Casella and Berger 1990, p158),

$$\Sigma_{\mathbf{p}^{(t+1)} | \mathbf{M}, \mathbf{V}, \mathbf{N}} = \Sigma_{E(\mathbf{p}^{(t+1)} | \mathbf{p}^{(t)}, \mathbf{M}, \mathbf{V}, \mathbf{N})} + E(\Sigma_{\mathbf{p}^{(t+1)} | \mathbf{p}^{(t)}, \mathbf{M}, \mathbf{V}, \mathbf{N}}). \quad (8)$$

Since $E(\mathbf{p}^{(t+1)} | \mathbf{p}^{(t)}, \mathbf{M}, \mathbf{V}, \mathbf{N}) = \mathbf{p}^{*(t)}$ and $\mathbf{p}^{*(t)}$ is a linear transformation of $\mathbf{p}^{(t)}$, i.e., $\mathbf{p}^{*(t)} = \mathbf{B}\mathbf{p}^{(t)}$, we have $\Sigma_{E(\mathbf{p}^{(t+1)} | \mathbf{p}^{(t)}, \mathbf{M}, \mathbf{V}, \mathbf{N})} = \mathbf{B}\Sigma_{\mathbf{p}^{(t)} | \mathbf{M}, \mathbf{V}, \mathbf{N}}\mathbf{B}'$. Furthermore, the conditional independence of the $\mathbf{p}_i^{(t+1)}$ given $\mathbf{p}^{(t)}$, \mathbf{M} , \mathbf{V} and \mathbf{N} along with the multinomial distribution in (2) implies that $\Sigma_{\mathbf{p}^{(t+1)} | \mathbf{p}^{(t)}, \mathbf{M}, \mathbf{V}, \mathbf{N}}$ is block diagonal with i^{th} block

$$\Delta_i = \frac{1}{2N_i} \{ \text{Diag}(\mathbf{p}_i^{*(t)}) - \mathbf{p}_i^{*(t)}(\mathbf{p}_i^{*(t)})' \} \quad ,$$

where $\text{Diag}(\mathbf{p}_i^{*(t)})$ is the $A \times A$ diagonal matrix with l^{th} diagonal entry equal to $p_{il}^{*(t)}$. Using (3),

$$E(\Delta_i) = \frac{1}{2N_i} \{ \text{Diag} \left(\sum \mathbf{B}_{ij} E(\mathbf{p}_j^{(t)} | \mathbf{M}, \mathbf{V}, \mathbf{N}) \right) \}$$

$$\begin{aligned}
& - \sum_j \sum_{j'} \mathbf{B}_{ij} E(\mathbf{p}_j^{(t)} | \mathbf{M}, \mathbf{V}, \mathbf{N}) E(\mathbf{p}_{j'}^{(t)} | \mathbf{M}, \mathbf{V}, \mathbf{N})' \mathbf{B}'_{ij'} \\
& - \sum_j \sum_{j'} \mathbf{B}_{ij} \Sigma_{\mathbf{p}_j^{(t)}, \mathbf{p}_{j'}^{(t)} | \mathbf{M}, \mathbf{V}, \mathbf{N}} \mathbf{B}'_{ij'}.
\end{aligned}$$

Again, passing to stationarity, $E(\Delta_i)$ becomes

$$\frac{1}{2N_i} \{ \text{Diag}(\mathbf{u}_i) - \mathbf{u}_i \mathbf{u}_i' - (\mathbf{B} \Sigma \mathbf{B}')'_{ii} \}$$

since $\mathbf{u}_i = \sum \mathbf{B}_{ij} \mathbf{u}_j$ and thus (8) becomes

$$\begin{aligned}
\Sigma_{ii} &= \left(1 - \frac{1}{2N_i}\right) (\mathbf{B} \Sigma \mathbf{B}')_{ii} + \frac{1}{2N_i} (\text{Diag}(\mathbf{u}_i) - \mathbf{u}_i \mathbf{u}_i') \\
\Sigma_{ij} &= (\mathbf{B} \Sigma \mathbf{B}')_{ij}.
\end{aligned} \tag{9}$$

The equations in (9) provide a linear system to solve for the entries in Σ given \mathbf{M} , \mathbf{V} , \mathbf{N} and \mathbf{U} , but this system is not full rank, e.g., Σ_{ii} is symmetric and singular. As k increases, solving (9) when Σ_{ii} and Σ_{ij} vary with i and j becomes analytically untractable, although numerical solution of (9) may still be feasible for small k . Nonetheless, some general results are possible.

Suppose that the populations are partitioned with respect to migration, i.e., that the migration matrix is reducible and populations in one partition have no connection through migration with populations in any other partition. Then, grouping populations appropriately, the migration matrix can be written in block diagonal form. In particular with L partitions, we can write

$$\mathbf{M} = \begin{pmatrix} \mathbf{M}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{M}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{M}_L \end{pmatrix}.$$

But then \mathbf{B} becomes block diagonal with l^{th} block equal to $\mathbf{M}_l \otimes \mathbf{V}'$. Suppose we look at the (l, l') block of Σ . This block denotes the stationary covariance between populations in partition l and populations in partition l' . Denoting the block by $\Sigma^{(ll')}$ we find $\Sigma^{(ll')} = \mathbf{B}_l \Sigma^{(ll')} \mathbf{B}'_{l'}$. \mathbf{B}_l and $\mathbf{B}'_{l'}$ are full rank. If they are symmetric, i.e., if \mathbf{V} and the \mathbf{M}_l are symmetric, and are not identity matrices, this equality implies $\Sigma^{(ll')} = \mathbf{0}$.

To make further analytical progress we ask when the marginal distribution for $\mathbf{p}_i^{(t)}$ will be the same for all i at stationarity. If so, $\Sigma_{ii} = \Sigma_{11}$ for all i , $\Sigma_{ij} = \Sigma_{12}$ for all i and j . We can simplify (9) to solving only for Σ_{11} and Σ_{12} regardless of k . A sufficient condition is that, if all $N_i = N$ and the entries in \mathbf{M} do not depend on i and j , i.e., \bar{m}_{ij} is constant, then a common distribution for all $\mathbf{p}_i^{(t)}$ will arise, regardless of \mathbf{V} . So, \mathbf{M} must be a constant migration rate matrix. We denote the entries in such an \mathbf{M} by $\mathbf{M}_{ii} = 1 - m$, $\mathbf{M}_{ij} = \frac{m}{k-1}$.

This corresponds to the finite-island model studied by Crow and Aoki (1984) and Cockerham and Weir (1987).

2.3 Results for the finite island model

Laborious calculation reveals that

$$\begin{aligned} (\mathbf{B}\Sigma\mathbf{B})_{11} &= \mathbf{V}' \{(1 - r(m, k))\Sigma_{11} + r(m, k)\Sigma_{12}\} \mathbf{V} \\ (\mathbf{B}\Sigma\mathbf{B})_{12} &= \mathbf{V}' \left\{ \frac{r(m, k)}{k-1} \Sigma_{11} + \left(1 - \frac{r(m, k)}{k-1}\right) \Sigma_{12} \right\} \mathbf{V} \end{aligned}$$

where $r(m, k) = 2m - \frac{m^2 k}{k-1}$. Finally,

$$\begin{aligned} \Sigma_{11} &= \left(1 - \frac{1}{2N}\right) \mathbf{V}' \{(1 - r(m, k))\Sigma_{11} + r(m, k)\Sigma_{12}\} \mathbf{V} \\ &\quad + \frac{1}{2N} (\text{Diag}(\mathbf{u}_1) - \mathbf{u}_1 \mathbf{u}_1') \\ \Sigma_{12} &= \mathbf{V}' \left\{ \frac{r(m, k)}{k-1} \Sigma_{11} + \left(1 - \frac{r(m, k)}{k-1}\right) \Sigma_{12} \right\} \mathbf{V}. \end{aligned} \tag{10}$$

An immediate conclusion from (10) is that for any finite k , $\mathbf{p}_i^{(t)}$ and $\mathbf{p}_j^{(t)}$ are correlated. However, as $k \rightarrow \infty$, $r(m, k) \rightarrow 2m - m^2$ and thus $\frac{r(m, k)}{k-1} \rightarrow 0$. Hence, in the limit, we obtain $\Sigma_{12} = \mathbf{V}' \Sigma_{12} \mathbf{V}$. So, as long as $\mathbf{V} \neq \mathbf{I}$, this equality can hold only if $\Sigma_{12} = 0$. Thus, $\Sigma_{12} \rightarrow 0$ as $k \rightarrow \infty$, implying that $\mathbf{p}_i^{(t)}$ and $\mathbf{p}_j^{(t)}$ are asymptotically uncorrelated.

In general, we need consider only the upper left $a \times a$ submatrix of Σ_{11} and Σ_{12} . In fact, when $A = 2$, Σ_{11} reduces to say $\sigma^{2(k)}$ and Σ_{12} reduces to, say $\rho^{(k)} \sigma^{2(k)}$. That is, for a given k , $\rho^{(k)}$ is the correlation between $\mathbf{p}_i^{(t)}$ and $\mathbf{p}_j^{(t)}$. Hence (10) produces two equations in two unknowns. The solution is:

$$\begin{aligned} \sigma^{2(k)} &= \frac{u_1 - u_1^2}{2N - (2N - 1)(\nu_{11} - \nu_{21})^2 \left(1 - r(m, k) + r(m, k)\rho^{(k)}\right)} \\ \rho^{(k)} &= \frac{\frac{r(m, k)}{k-1} (\nu_{11} - \nu_{21})^2}{1 - (\nu_{11} - \nu_{21})^2 \left(1 - \frac{r(m, k)}{k-1}\right)} \end{aligned} \tag{11}$$

where $u_1 = \frac{\nu_{21}}{\nu_{12} + \nu_{21}}$. So as $k \rightarrow \infty$, $\sigma^{2(k)} \rightarrow \sigma^{2(\infty)} = \frac{u_1 - u_1^2}{2N - (2N - 1)(\nu_{11} - \nu_{21})^2 (1 - m)^2}$. Again, $\rho^{(\infty)} = 0$.

For an arbitrary number of alleles A with a constant mutation rate (i.e., $\nu_{ii} = 1 - \nu$, $\nu_{ij} = \nu/(A - 1)$) in addition to a constant migration rate, each component of $\mathbf{p}_i^{(t)}$ has the same marginal distribution. Also, the correlation between any pair of components is the same. Hence Σ_{11} has a constant diagonal say $\sigma^{2(k)}$ and all off-diagonal entries are say $\rho_1^{(k)} \sigma^{2(k)}$. Similarly, Σ_{12} will have a constant diagonal say $\rho_2^{(k)} \sigma^{2(k)}$ and all off-diagonal entries are say $\rho_3^{(k)} \sigma^{2(k)}$. Hence given k , $\rho_1^{(k)}$ is the correlation between $p_{il}^{(t)}$ and $p_{il'}^{(t)}$, $\rho_2^{(k)}$ is the correlation between $p_{il}^{(t)}$ and $p_{jl}^{(t)}$, and $\rho_3^{(k)}$ is the correlation between $p_{il}^{(t)}$ and $p_{jl'}^{(t)}$. Solving explicitly

yields:

$$\begin{aligned}
\sigma^{2(k)} &= \frac{(A-1)/A^2}{2N - (2N-1) \left(1 - \frac{A}{A-1}\nu\right)^2 \left(1 - r(m, k) + r(m, k)\rho_2^{(k)}\right)} \\
\rho_1^{(k)} &= -\frac{1}{A-1} \\
\rho_2^{(k)} &= \frac{\left(1 - \frac{A}{A-1}\nu\right)^2 \frac{r(m, k)}{k-1}}{1 - \left(1 - \frac{A}{A-1}\nu\right)^2 \left(1 - \frac{r(m, k)}{k-1}\right)} \\
\rho_3^{(k)} &= \rho_1^{(k)} \rho_2^{(k)}.
\end{aligned} \tag{12}$$

As $k \rightarrow \infty$, $\sigma^{2(k)} \rightarrow \sigma^{2(\infty)} = \frac{(A-1)/A^2}{2N - (2N-1)\left(1 - \frac{A}{A-1}\nu\right)^2 (1-m)^2}$, $\rho_2^{(k)} \rightarrow \rho_2^{(\infty)} = 0$ and $\rho_3^{(k)} \rightarrow \rho_3^{(\infty)} = 0$.

Another example where we can explicitly solve (10) arises when $\nu_{rs} = \nu_s$ for all $r \neq s$. If ν_{rs} is free of r , the interpretation is that the rate of mutation to b_s does not depend upon the current allele type. This corresponds to Kingman's (1980) "house of cards" model of mutation. In this setting, the variance and correlation arising in Σ_{11} are:

$$\begin{aligned}
\sigma_i^{2(k)} &= \frac{\sum_{j \neq i}^A \nu_j \nu_i / \left(\sum_{j=1}^A \nu_j\right)^2}{2N - (2N-1) \left(1 - \sum_{j=1}^A \nu_j\right)^2 \left(1 - r(m, k) + r(m, k)\rho^{(k)}\right)} \\
\rho_{ii'}^{(k)} &= -\frac{\nu_i \nu_{i'}}{\sqrt{\left(\sum_{j \neq i}^A \nu_j \nu_i\right) \left(\sum_{j \neq i'}^A \nu_j \nu_{i'}\right)}}.
\end{aligned} \tag{13}$$

Interestingly, $\Sigma_{12} = \rho^{(k)} \Sigma_{11}$ where

$$\rho^{(k)} = \frac{\left(1 - \sum_{j=1}^A \nu_j\right)^2 \frac{r(m, k)}{k-1}}{1 - \left(1 - \sum_{j=1}^A \nu_j\right)^2 \left(1 - \frac{r(m, k)}{k-1}\right)}. \tag{14}$$

2.4 Results for the one-dimensional stepping stone

As a final illustration, we again let $A = 2$ but consider a nonconstant migration matrix. In fact, we look at the one-dimensional, circular stepping stone model proposed originally in Kimura and Weiss (1964) and Weiss and Kimura (1965). For k populations, \mathbf{M} takes the form of $m_{ii} = 1 - m$, $\overleftarrow{m}_{i, i+1} = \overleftarrow{m}_{i, i-1} = m/2$, $\overleftarrow{m}_{ij} = 0$ otherwise. Note that, we have chosen to align the diagonal of this migration matrix \mathbf{M} with that of the constant migration rate matrix. Hence, here m is twice the value used in Kimura and Weiss (1964). Again it is clear that if all $N_i = N$, then all $p_i^{(t)}$ have the same marginal distribution. Indeed, from the discussion above (6), the stationary mean is $\frac{\nu_{21}}{1 - (\nu_{11} - \nu_{21})}$. In (9), Σ_{ii} is a constant $\sigma^{2(k)}$, while $\Sigma_{ii'} = \rho_{|i-i'|}^{(k)} \sigma^{2(k)}$. A recursion for $\rho_l^{(k)}$ introduces $\rho_{l'}^{(k)}$, $l' = l - 1, l - 2, l + 1$ and $l + 2$. In the linear stepping stone model, assuming $k = \infty$, Kimura and Weiss achieved an approximate solution starting from ρ_l of the form λ^l . Maruyama (1977) shows that exact calculations

k	$\nu = 5 \times 10^{-6}$			$\nu = 5 \times 10^{-3}$		
	$2N$			$2N$		
	10^2	10^4	10^6	10^2	10^4	10^6
50	0.2476 (2.273×10^{-2})	0.1255 (2.497×10^{-4})	2.496×10^{-3} (2.500×10^{-6})	0.0839 (2.232×10^{-2})	1.245×10^{-3} (2.449×10^{-4})	1.251×10^{-5} (2.451×10^{-6})
100	0.2453 (2.273×10^{-2})	0.0852 (2.497×10^{-4})	1.285×10^{-3} (2.500×10^{-6})	0.0839 (2.232×10^{-2})	1.245×10^{-3} (2.449×10^{-4})	1.251×10^{-5} (2.451×10^{-6})
300	0.2389 (2.273×10^{-2})	0.0441 (2.497×10^{-4})	5.337×10^{-4} (2.500×10^{-6})	0.0839 (2.232×10^{-2})	1.245×10^{-3} (2.449×10^{-4})	1.251×10^{-5} (2.451×10^{-6})
500	0.2364 (2.273×10^{-2})	0.0367 (2.497×10^{-4})	4.296×10^{-4} (2.500×10^{-6})	0.0839 (2.232×10^{-2})	1.245×10^{-3} (2.449×10^{-4})	1.251×10^{-5} (2.451×10^{-6})

Table I: Stationary variance for the two allele circular stepping stone model for $m = 0.05$ and selected values of ν , k and N . Values in parentheses are from the single-population diffusion approximation.

based on probabilities of identity by descent converge to those of Kimura and Weiss (1964) as $k \rightarrow \infty$. Table I provides values of $\sigma^{2(k)}$ for $\nu_{12} = \nu_{21} = 5 \times 10^{-6}$ and 5×10^{-3} , $m = 0.05$ and $k = 50, 100, 300, 500$, $N = 10^2, 10^4, 10^6$. The parenthetic entries in the table will be discussed below in the section on accuracy of the diffusion approximation. Figure 1 shows the decline in correlation between $p_i^{(t)}$ and $p_{i'}^{(t)}$ as a function of $|i - i'|$. Correlation between nearby populations is always quite strong and only for the largest k does it eventually become 0 at $k/2$. Notice that in contrast to the results provided in Kimura and Weiss (1964) and Weiss and Kimura (1965), the results presented here are for the correlation of allele frequencies between populations, not for the correlation of allele frequencies within individuals, which is derived from a continuous approximation to the discrete process modeled here (compare Cox and Durrett 2001).

2.5 Variable population sizes

We conclude this section with some results on extending the stationary Markov transition model to allow N_i to vary across generations. Let $\mathbf{N}^{(t)'} = (N_1^{(t)}, \dots, N_k^{(t)})$ and suppose trajectories of $\mathbf{N}^{(t)}$ vs t arise from a stationary Markov chain with parametric transition distribution $q(\mathbf{N}^{(t+1)} | \mathbf{N}^{(t)}, \boldsymbol{\theta})$ (compare Karlin 1968). Suppose $\mathbf{N}^{(t)}$ has a stationary distribution and, further, that under this distribution $E(N_i^{(t)})^{-1}$ exists. Then the transition from $(\mathbf{P}^{(t)}, \mathbf{N}^{(t)}) \rightarrow (\mathbf{P}^{(t+1)}, \mathbf{N}^{(t+1)})$ proceeds to update $\mathbf{N}^{(t+1)}$ given $\mathbf{N}^{(t)}$ and $\boldsymbol{\theta}$ and then $\mathbf{P}^{(t+1)}$ given $\mathbf{P}^{(t)}$, $\mathbf{N}^{(t+1)}$, \mathbf{M} and \mathbf{V} . Figure 2 shows a graphical model for this first order dynamic Markov specification. If $\mathbf{N}^{(t)}$ is degenerate at say \mathbf{N} then we have the original model defined through (1) and (2). Hence, with a stationary chain for $\mathbf{N}^{(t)}$, overall we have a stationary chain for $(\mathbf{P}^{(t)}, \mathbf{N}^{(t)})$. Again, we will have a stationary distribution for $(\mathbf{P}^{(t)}, \mathbf{N}^{(t)})$, and interest focuses on the marginal stationary distribution for $\mathbf{P}^{(t)}$.

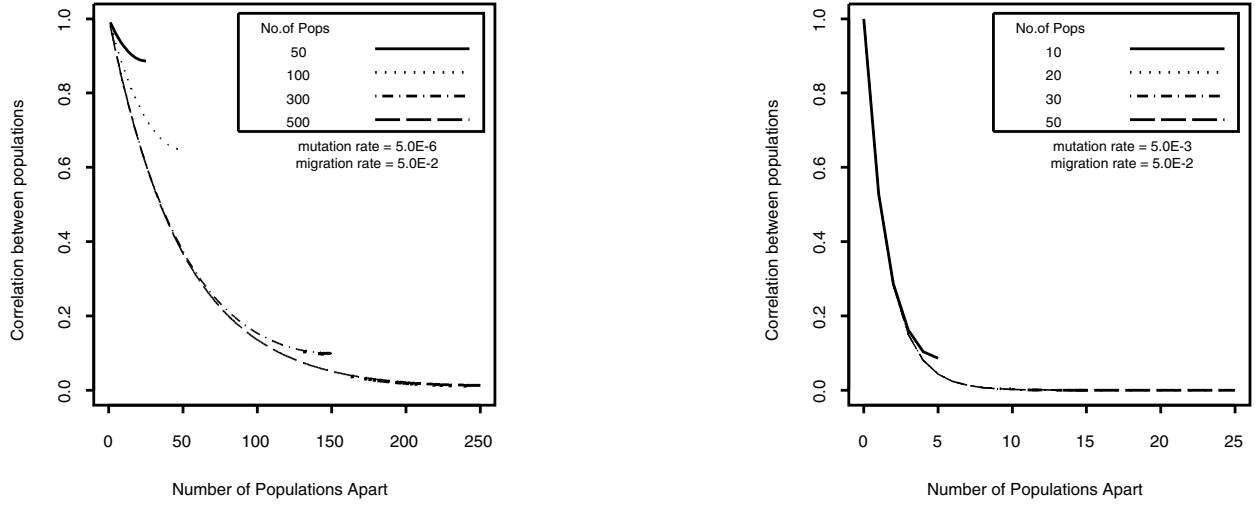


Figure 1: Correlation as a function of distance apart around the circle for the two - allele circular stepping stone model.

$$\begin{array}{ccccccccccc}
 \mathbf{N}^{(0)} & \rightarrow & \mathbf{N}^{(1)} & \rightarrow & \mathbf{N}^{(2)} & \rightarrow & \dots & \rightarrow & \mathbf{N}^{(t)} & \rightarrow & \mathbf{N}^{(t+1)} \\
 & & \downarrow & & \downarrow & & & & \downarrow & & \downarrow \\
 \mathbf{P}^{(0)} & \rightarrow & \mathbf{P}^{(1)} & \rightarrow & \mathbf{P}^{(2)} & \rightarrow & \dots & \rightarrow & \mathbf{P}^{(t)} & \rightarrow & \mathbf{P}^{(t+1)}
 \end{array}$$

Figure 2: A dynamic Markov model for generational transition.

Since equation (1) is free of \mathbf{N} , expression (4) again emerges as the stationary equation for \mathbf{U} . All of the subsequent discussion to (4) is applicable here. In particular, the mean vector of allele frequencies does not depend on either the pattern or magnitude of migration or on the distribution of population sizes. Turning to the stationary covariance, again writing $\mathbf{P}^{(t)}$ as $\mathbf{p}^{(t)}$, the analogue of (8) is, suppressing \mathbf{M} and \mathbf{V} ,

$$\Sigma_{\mathbf{p}^{(t+1)}} = \Sigma_{E(\mathbf{p}^{(t+1)}|\mathbf{p}^{(t)},\mathbf{N}^{(t+1)})} + E(\Sigma_{\mathbf{p}^{(t+1)}|\mathbf{p}^{(t)},\mathbf{N}^{(t+1)}}). \quad (15)$$

As below (8), the first term on the right side of (15) is $\mathbf{B}\Sigma_{\mathbf{p}^{(t)}}\mathbf{B}'$. For the second term on the right side, again the $\mathbf{p}_i^{(t+1)}$ are conditionally independent given $\mathbf{p}_i^{(t)}$ and $\mathbf{N}^{(t+1)}$. It is straightforward to show, using calculations similar to those leading to (9), that (15) yields

$$\begin{aligned}
 \Sigma_{ii} &= (1 - E(2N_i^{(t)})^{-1})(\mathbf{B}\Sigma\mathbf{B})'_{ii} + E(2N_i^{(t)})^{-1}(\text{Diag}(\mathbf{u}_i) + \mathbf{u}_i\mathbf{u}_i') \\
 \Sigma_{ij} &= (\mathbf{B}\Sigma\mathbf{B})'_{ij}.
 \end{aligned} \quad (16)$$

A stochastic model for $\mathbf{N}^{(t+1)}|\mathbf{N}^{(t)}$ would probably view the population sizes $N_i^{(t)}$ as

exchangeable, and might even regard them as independent. If the $N_i^{(t)}$ are regarded as exchangeable, the stationary marginal distribution for each one is the same, i.e. $E(N_i^{(t)})^{-1}$ is a constant, say $1/N_0$. Hence, the analyses for equation (9) are all still applicable with $1/N_0$ replacing $1/N$.

Recall that when population size varies from one generation to the next, the inbreeding effective size of the population (i.e., the size of an ideal population in which identity by descent accumulates at the same rate as this one) is approximately equal to the harmonic mean of the $N_i^{(t)}$ (Ewens 1979; Karlin 1968). To estimate $1/N_0$ with a sample of population size N_1, N_2, \dots, N_k , we would use $k^{-1} \sum (1/N_i)$. That is, the harmonic mean of population sizes would also provide an appropriate measure for the effect of genetic drift on the covariance structure of allele frequencies within and among populations in finite island models of migration.

3 Comparison of Exact results with the Diffusion Approximation

The familiar diffusion approximation (Kimura 1964; Crow and Kimura 1970) in the case of one locus with two alleles focuses on one population and assumes that at each generation the fraction of immigration is m and that the allele frequency in migrants is a constant say, p_0 . Assuming mutation *before* migration,

$$p^{*(t)} = (1 - m)(\nu_{11}p^{(t)} + \nu_{21}(1 - p^{(t)})) + mp_0 \quad . \quad (17)$$

Then, as in (2)

$$2Np^{(t+1)} \sim \text{Bi}(2N, p^{*(t)}) \quad . \quad (18)$$

To draw some simple parallels, with migration *before* mutation, we obtain

$$p^{*(t)} = (1 - m)(\nu_{11}p^{(t)} + \nu_{21}(1 - p^{(t)})) + m(\nu_{11}p_0 + \nu_{21}(1 - p_0)) \quad . \quad (19)$$

Under equation (1) we obtain for population i

$$p_i^{*(t)} = (1 - m)(\nu_{11}p_i^{(t)} + \nu_{21}(1 - p_i^{(t)})) + m \left(\nu_{11} \frac{\sum_{j \neq i} p_j^{(t)}}{k - 1} + \nu_{21} \frac{\sum_{j \neq i} (1 - p_j^{(t)})}{k - 1} \right) \quad . \quad (20)$$

So, if $\lim_{k \rightarrow \infty} \frac{\sum_{j \neq i} p_j^{(t)}}{k - 1} = p_0$ almost surely, then (20) tends to (19) almost surely. But also, for any given k , the variance of $p^{*(t)}$ under (20) is greater than variance of $p^{*(t)}$ under (19) due to the randomness of the $p_j^{(t)}$'s.

Under equation (17), the diffusion approximation produces a stationary distribution for

$p^{(t)}$ which is $\text{Be}(4N(\nu_{21} + mp_0), 4N(\nu_{12} + m(1 - p_0)))$ and hence, stationary variance

$$\sigma^2 = \frac{(\nu_{21} + mp_0)(\nu_{12} + m(1 - p_0))}{(\nu_{21} + \nu_{12} + m)^2(4N(\nu_{21} + \nu_{12} + m) + 1)} \quad (21)$$

It is a matter of algebra to show that when $p_0 = \nu_{21}/(\nu_{12} + \nu_{21})$, $\sigma^{2(k)}$ in (11) agrees with (21), ignoring non-first order terms in ν_{12} , ν_{21} and m .

In applying these results to likelihood or Bayesian inference from population samples, it is customary to assume that the $p_i^{(t)}$ are independent and that joint density for $(p_1^{(t)}, \dots, p_k^{(t)})$ is a product of k Beta densities (see, for example, Barton et al. 1983; Wehrhahn and Powell 1987; Slatkin and Barton 1989; Wehrhahn 1989; Rannala and Hartigan 1995; Holsinger 1999). The diffusion result implicitly assumes that terms containing a product of migration and mutation rates are ignored (as well as higher order mutation and migration rate terms). It also assumes that the conditional variance of $p^{(t+1)}$ given $p^{(t)}$ is $p^{(t)}(1 - p^{(t)})/2N$ rather than $p^{*(t)}(1 - p^{*(t)})/2N$.

For the remainder of this section we investigate the accuracy of the Beta approximation for the model under (1) and (2) when $A = 2$. We do this in three ways. First, to study the effect of the independence assumption, we use (11) to calculate the decay in correlation as k increases. Second, we look at the variance in (11) and compare it with (21). Finally, we look at density estimates for $p_i^{(t)}$ under (1) and (2) and compare them with the corresponding density estimates under the Beta approximation. We note immediately that the stationary mean of $p_i^{(t)}$ is $\nu_{21}/(\nu_{12} + \nu_{21})$ while the Beta approximation yields $(\nu_{21} + mp_0)/(\nu_{12} + \nu_{21} + m) \approx p_0$ if m is much larger than ν_{21} and ν_{12} as is usually assumed in practice. Hence, the Beta approximation is suitably centered only if $p_0 = \nu_{21}/(\nu_{12} + \nu_{21})$.

Figure 3 plots $\rho^{(k)}$ vs. $\log_{10} k$ for $\nu_{12} = \nu_{21} = 5 \times 10^{-6}$ and 5×10^{-3} with $m = 0.05$. Note that for the smaller mutation rate we require 10^6 populations before $\rho^{(k)}$ is essentially 0.

Assuming no mutation in (17) and using the same logic that leads to (10), Crow and Kimura (1970) obtain the stationary variance approximation

$$\sigma^2 = \left(1 - (1 - m)^2 \left(1 - \frac{1}{2N}\right)\right)^{-1} \frac{p_0(1 - p_0)}{2N} \quad (22)$$

Straightforward algebra shows that (22) is greater than (21) at $\nu_{12} = 0$ and $\nu_{21} = 0$, suggesting that the variance of the Beta approximation will underestimate the true population variance. Since the variance of $p_i^{(t)}$ under (20) exceeds that of $p_i^{(t)}$ under (19), assuming no mutation, we can show explicitly that the stationary variance for $p_i^{(t)}$ under (20) is larger than (22) and, in fact, decreases to (22) as $k \rightarrow \infty$. We omit details. With mutation we expect $\sigma^{2(k)}$ in (11) to exceed (19). Figure 4 plots $\sigma^{2(k)}$ vs. $\log_{10} k$ for different $\nu_{12} = \nu_{21}$, m , and N combinations. Indeed, $\sigma^{2(k)}$ in (11) exceeds that in (21) at $p_0 = 1/2$ and, for small mutation rate, the relative error of the former to the latter could be considerable unless k is large.

In the circular stepping-stone model described above we have a common equilibrium

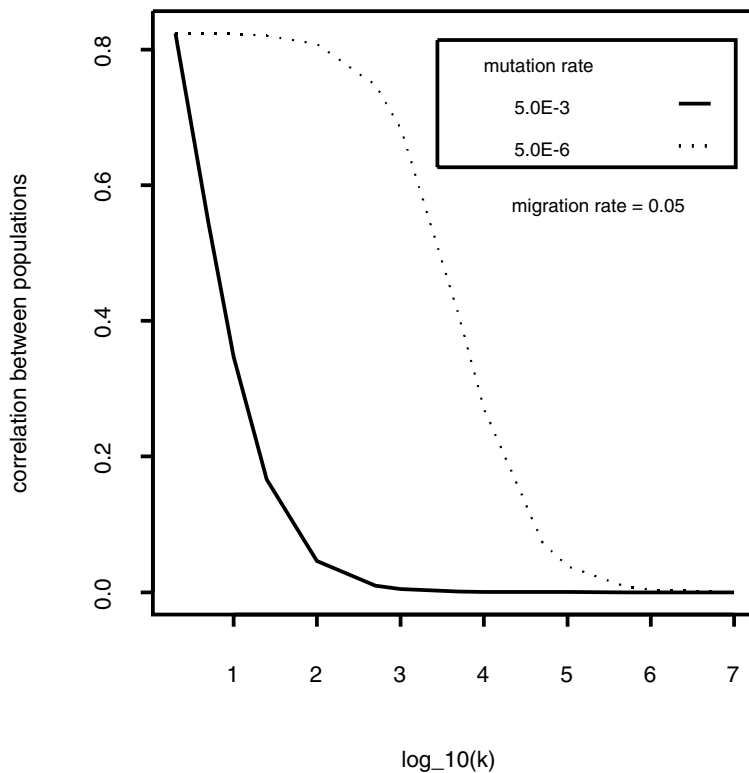


Figure 3: For the two-allele model, a plot of correlation vs. $\log_{10}(k)$ using (11) for two choices of mutation rate with a given migration rate.

distribution for all $p_i^{(t)}$. In fact,

$$p_i^{*(t)} = (1 - m)(\nu_{11}p_i^{(t)} + \nu_{21}(1 - p_i^{(t)})) + m \left(\nu_{11} \frac{p_{i-1}^{(t)} + p_{i+1}^{(t)}}{2} + \nu_{21} \frac{(1 - p_{i-1}^{(t)}) + (1 - p_{i+1}^{(t)})}{2} \right) \quad (23)$$

Thus we have the same fraction of immigrants into each population as in the constant migration rate model (20). The difference is that in averaging over contributions from $k - 1$ other populations rather than just two, the variance of $p_i^{*(t)}$ in (20) will be smaller than that in (23). Hence we expect the variance of $p_i^{(t)}$ under the stepping stone model to exceed $\sigma^{2(k)}$ in (11) and thus to exceed (21). The parenthetic entries in Table I, which apply to $p_0 = 1/2$, support this assertion.

Finally, returning to the finite island model, for choices of N , k , $\nu_{12} = \nu_{21}$, and m we ran the Markov transitions for 10^4 iterations ($\nu_{12} = \nu_{21} = 5 \times 10^{-3}$) and for 10^6 iterations ($\nu_{12} = \nu_{21} = 5 \times 10^{-6}$) to reach rough equilibrium. We then retained $p_i^{(t)}$ for every 100th ($\nu_{12} = \nu_{21} = 5 \times 10^{-3}$) or every 10⁵ th ($\nu_{12} = \nu_{21} = 5 \times 10^{-6}$) iteration. Finally, for illustration,

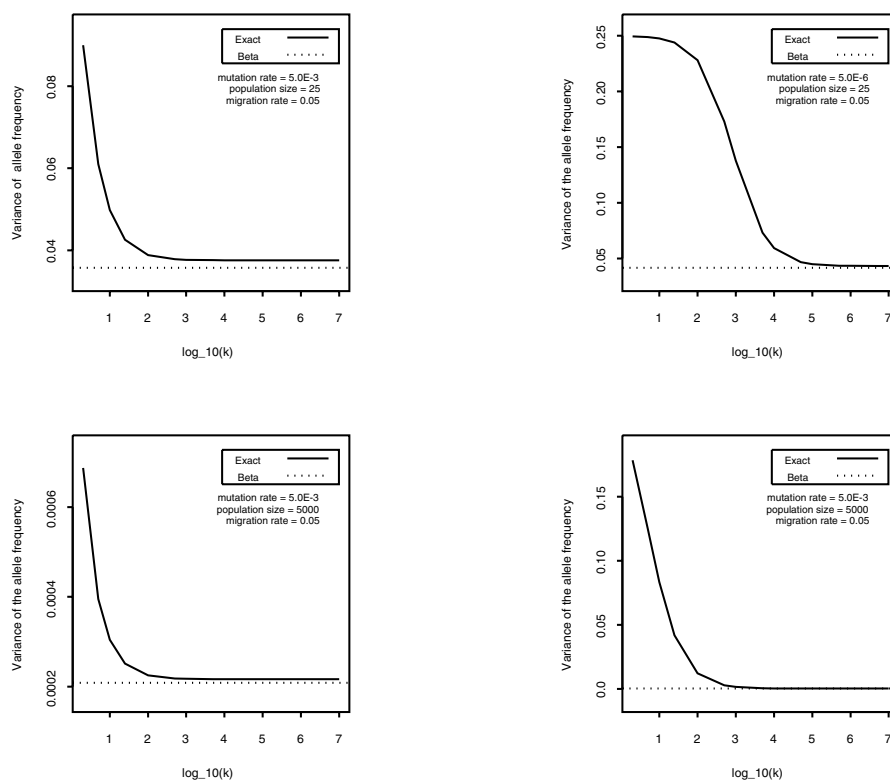


Figure 4: Plot of variance of allele frequency vs. $\log_{10}(k)$ for choices of mutation rate, migration rate and population size for the two allele model.

we created a kernel density estimate of the sample of $p_1^{(t)}$'s. While the distribution of $p_1^{(t)}$ is, in fact, discrete, for large N it is approximately continuous making such density estimates sensible. Figure 5a compares these density estimates with the Beta approximation when $N = 500$ or $2N = 1000$, Figure 5b makes this comparison with $N = 5000$ or $2N = 10000$. Again, the Beta approximation is too concentrated and, for small mutation rate, even the shape can be very far off.

6 Interpretation for F_{st} analysis

F -statistics as introduced by Wright (1951) and Malécot (1948) are widely used to describe hierarchical structure in genetic data. Wright (1969) defined F_{st} as “the correlation between random gametes within [populations], relative to gametes of the total [set of populations] (p.

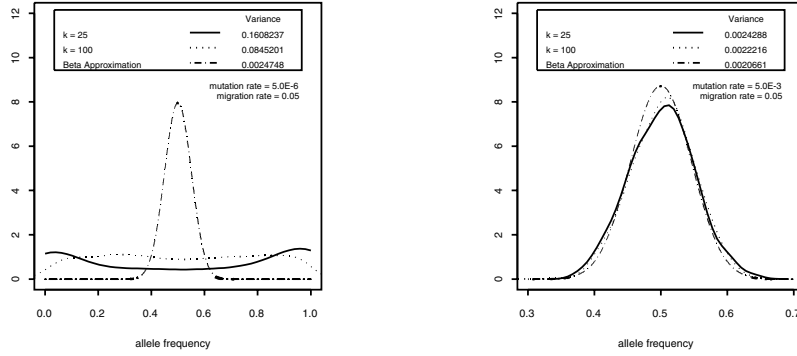


Figure 5a: Comparison of stationary density of allele frequency distribution based on the simulation with the beta distribution. $2N = 1000$.

294). For one locus with two alleles this is equivalent to an intraclass correlation coefficient

$$F_{st} = \frac{\sigma_p^2}{\mu_p(1 - \mu_p)} \quad , \quad (24)$$

where μ_p is the mean allele frequency across populations and σ_p^2 is the variance in allele frequency among populations. Thus, for a finite set of populations evolving according to (1) and (2) a natural analog of (24) is

$$\theta(p_1^{(t)}, \dots, p_k^{(t)}) = \frac{\sum (p_i^{(t)} - \mu_{p^{(t)}})^2 / k}{\mu_{p^{(t)}}(1 - \mu_{p^{(t)}})} \quad , \quad (25)$$

where $\mu_{p^{(t)}} = (1/k) \sum p_i^{(t)}$. $\theta(p_1^{(t)}, \dots, p_k^{(t)})$ is, of course, both random and unobservable. We use θ (suppressing $p_1^{(t)}, \dots, p_k^{(t)}$) to represent the parameter of interest to emphasize that it is a stochastic process analog of Cockerham's random-effects model for θ -statistics (Cockerham 1969; Weir and Cockerham 1984; Weir 1996; Weir and Hill 2002). Empirical population

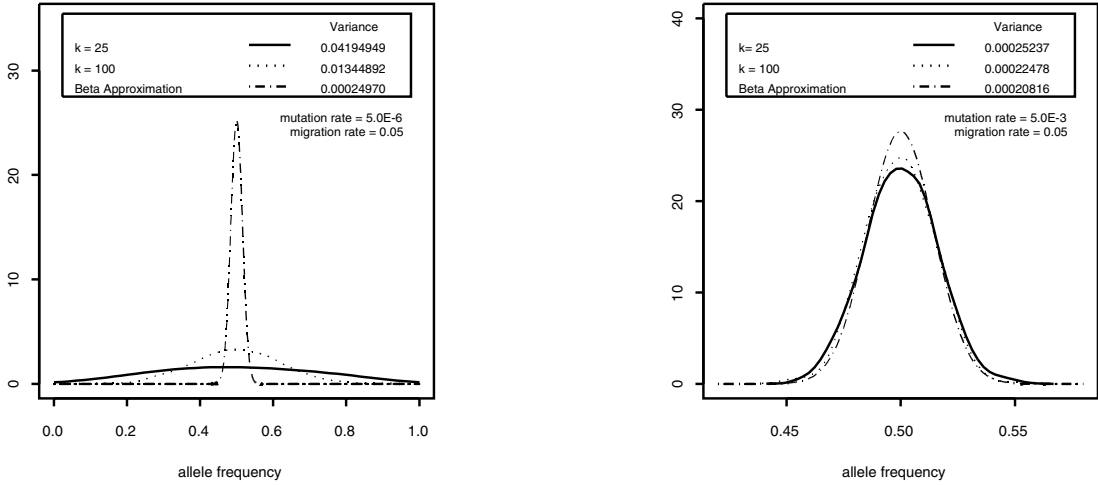


Figure 5b: Comparison of stationary density of allele frequency distribution based on the simulation with the beta distribution. $2N = 10000$.

geneticists are interested in using estimates of $p_i^{(t)}$, i.e., \hat{p}_i , to estimate θ .

If the $p_i^{(t)}$ were i.i.d, then we would replace (25) by $\theta^{(I)} = \sigma_{p_i^{(t)}}^2 / \mu_{p^{(t)}}(1 - \mu_{p^{(t)}})$ and view $\theta^{(I)}$ as the parameter to be estimated. If we let Num denote the numerator on the right side of (25), $Denom$ the denominator then we might also consider the parameter $\theta^{(II)} = E(Num)/E(Denom)$ or perhaps $\theta^{(III)} = E(Num/Denom)$. $\theta^{(III)}$ corresponds directly with the definition in (25), i.e., were it observable (25) would be an unbiased estimator of $\theta^{(III)}$. $\theta^{(II)}$ is clearly not equal to $\theta^{(III)}$, but as $k \rightarrow \infty$, standard convergence results show that $\theta^{(II)} \rightarrow \theta^{(III)}$. Note, however, that while $\theta^{(II)}$ and $\theta^{(III)}$ depend on k , $\theta^{(I)}$ does not (compare Cockerham and Weir 1987).

In the simple cases considered above, the finite-island model and the one-dimensional stepping stone model, the $p_i^{(t)}$ are identically distributed but not independent. From Figure 3, for plausible migration rates in the finite-island model we may require k very large to feel comfortable with the assumption that the $p_i^{(t)}$'s are roughly uncorrelated. Expressed in different words, (25) considers variability in allele frequency across populations while $\theta^{(I)}$ looks at variability across generations. Only if the $p_i^{(t)}$ are i.i.d will the former approximate the latter, and the discussion following equation (9) makes it clear that the $p_i^{(t)}$ will not be i.i.d. in general. Because Wright's F -statistics and Cockerham's θ -statistics are used primarily to describe genetic structure within a single collection of samples rather than to describe variation in allele frequencies across time within a single population, interest will usually center on $\theta^{(III)}$. Indeed, estimates of F_{st} based on Cockerham's (1969) analysis of variance approach to partitioning genetic diversity (e.g., Weir and Cockerham 1984; Cockerham and Weir 1987; Weir 1996; Excoffier 2001; Weir and Hill 2002) explicitly justify using

the method-of-moments analog of $\theta^{(II)}$ by noting that it provides an unbiased estimator of $\theta^{(III)}$, in the sense that method-of-moments estimators for the numerator and denominator are unbiased. Importantly, under (1) and (2), for small to moderate k we show that $\theta^{(I)}$, $\theta^{(II)}$ and $\theta^{(III)}$ can be quite different and that they can also be quite different from corresponding values obtained using a product Beta distribution for the $p_i^{(t)}$.

Suppose that $(p_1^{(t)}, \dots, p_k^{(t)})$ arise under (1) and (2). At stationarity, standard calculation reveals that

$$E\left(\frac{\sum(p_i^{(t)} - \mu_{p^{(t)}})^2}{k}\right) = \frac{k-1}{k}\sigma^{2(k)}(1 - \rho^{(k)}) \quad (26)$$

with $E(\mu_{p^{(t)}}) = \nu_{21}/(\nu_{12} + \nu_{21})$ and $\sigma^{2(k)}$ and $\rho^{(k)}$ as in (11). Also

$$E(\mu_{p^{(t)}}(1 - \mu_{p^{(t)}})) = E(\mu_{p^{(t)}})(1 - E(\mu_{p^{(t)}})) - \frac{1}{k}\sigma^{2(k)}(1 + (k-1)\rho^{(k)}). \quad (27)$$

If the $p_i^{(t)}$ are i.i.d., (26) simplifies to $\frac{(k-1)}{k}\sigma_{p^{(t)}}^2$ and under the approximate Beta distribution $\sigma_{p^{(t)}}^2 = \sigma^2$ in (21). Similarly, (27) simplifies to $E(\mu_{p^{(t)}})(1 - E(\mu_{p^{(t)}})) - \sigma_{p^{(t)}}^2/k$.

Table II provides comparisons among $\theta^{(I)}$, $\theta^{(II)}$ and $\theta^{(III)}$ and between the model in (1) and (2) and the product Beta approximation for $2N = 100$, $\nu_{12} = \nu_{21} = 5.0 \times 10^{-6}$, $m = 0.01$, $k = \{25, 100\}$ and for four combinations of $N = \{1000, 10000\}$, $k = \{25, 100\}$ with $\nu_{12} = \nu_{21} = 5.0 \times 10^{-6}$, and $m = 5 \times 10^{-2}$. The results are striking. While $\theta^{(I)}$, $\theta^{(II)}$, and $\theta^{(III)}$ differ little from one another in a product Beta model, they are strikingly different from one another in the exact model. Notice especially that $\theta^{(I)}$ can be two orders of magnitude greater than $\theta^{(II)}$ and $\theta^{(III)}$ under the exact model. Not surprisingly $\theta^{(II)}$ and $\theta^{(III)}$ differ less from one another, except when the population size is small. With $2N = 100$, however, $\theta^{(III)}$ in the exact model is less than 80% of $\theta^{(II)}$, even for $k = 100$. These results suggest that when populations are small, estimates of F_{st} based on $\theta^{(II)}$ will substantially overestimate the amount of genetic differentiation among populations, regardless of whether they are derived using the method-of-moments, maximum-likelihood, or a Bayesian approach.

Finally, we simulated the distribution of (25) under the model in (1) and (2) and under the Beta approximation for $2N = 100$, $m = 0.01$ and selected values of mutation rate and number of populations. Figure 6 displays the resulting kernel density estimates for $\theta^{(III)}$ in both cases. For the larger mutation rate the distributions are quite similar while for the small rate and somewhat smaller k 's, the differences are substantial. Also noteworthy is that, in this latter case when $k = 25$, only 24% of F_{st} 's were different from 0/0, and when $k = 100$, only 70% were different from 0/0.

The reason why $\theta^{(III)}$ in the exact model departs so substantially from $\theta^{(II)}$ in the exact model and from $\theta^{(II)}$ and $\theta^{(III)}$ in the product beta model becomes apparent upon reflection. The expectation of a ratio depends on the covariance between numerator and denominator. $\theta^{(II)}$ in the exact model ignores the covariance between numerator and denominator, while the numerator and denominator are free to vary independently in the product beta model.

		Exact Model			Beta approximation			
m	$2N$	k	$\theta^{(I)}$	$\theta^{(II)}$	$\theta^{(III)}$	$\theta^{(I)}$	$\theta^{(II)}$	$\theta^{(III)}$
			$\frac{Var(p_i)}{E(p_i)(1-E(p_i))}$	$\frac{E(Num)}{E(Denom)}$	$E(\frac{Num}{Denom})$	$\frac{Var(p_i)}{E(p_i)(1-E(p_i))}$	$\frac{E(Num)}{E(Denom)}$	$E(\frac{Num}{Denom})$
			Mean	Mean	Mean S.D.	Mean	Mean	Mean S.D.
0.01	100	25	0.9539	0.3456	0.2149 2.360×10^{-2}	0.3331	0.3517	0.3237 3.606×10^{-3}
		100	0.8473	0.3388	0.2655 1.799×10^{-2}	0.3331	0.3376	0.3310 8.593×10^{-4}
0.05	1000	25	0.6679	1.018×10^{-2}	8.754×10^{-3} 3.924×10^{-3}	9.899×10^{-3}	1.032×10^{-2}	9.566×10^{-3} 2.761×10^{-3}
		100	0.3380	1.017×10^{-2}	9.987×10^{-3} 1.441×10^{-3}	9.899×10^{-3}	1.000×10^{-2}	9.759×10^{-3} 1.389×10^{-3}
0.05	10000	25	0.1673	1.026×10^{-3}	9.441×10^{-4} 2.726×10^{-4}	9.988×10^{-4}	1.040×10^{-3}	9.676×10^{-4} 2.933×10^{-4}
		100	0.0485	1.025×10^{-3}	1.010×10^{-3} 1.426×10^{-4}	9.988×10^{-4}	1.009×10^{-3}	9.850×10^{-4} 1.337×10^{-4}

Table II: Comparison of three definitions for $\theta(p_1, \dots, p_k)$ under model (1) and (2) and under the Beta approximation for the two allele cases with migration rate $= 5 \times 10^{-2}$ and mutation rate $= 5 \times 10^{-6}$. Entries for $E(\frac{Num}{Denom})$ are based on simulation results. The remainder are exact calculations.

In short, only $\theta^{(III)}$ in the exact model reflects the temporal covariance in allele frequencies expected with gene exchange among a finite number of populations. As illustrated in Figure 7a, the distribution of allele frequencies among 100 populations that are exchanging genes changes substantially over time, even after the drift, migration, mutation process has reached stationarity. When 1000 populations are exchanging genes, temporal changes in distribution of allele frequencies among populations are less dramatic, but they are still readily visible (see Figure 7b).

8 Discussion

Nearly all existing methods for inference of F -statistics depend either explicitly, in the case of the product beta assumption for likelihood and Bayesian methods, or implicitly, in the case of method-of-moments methods, on the assumption that allele frequencies vary independently across populations (see Weir and Hill 2002 for a description of how to relax the independence assumption for method-of-moments estimators). Our results show that, in contrast, allele frequencies covary whenever a finite number of populations is exchanging genes. The covariance arises because drift occurs at the level of the entire set of populations, resulting in changes in the mean allele frequency over time. Moreover, the correlation in allele frequencies among populations causes the single-population diffusion approximation to provide a poor description both of the allele frequency distribution among populations and

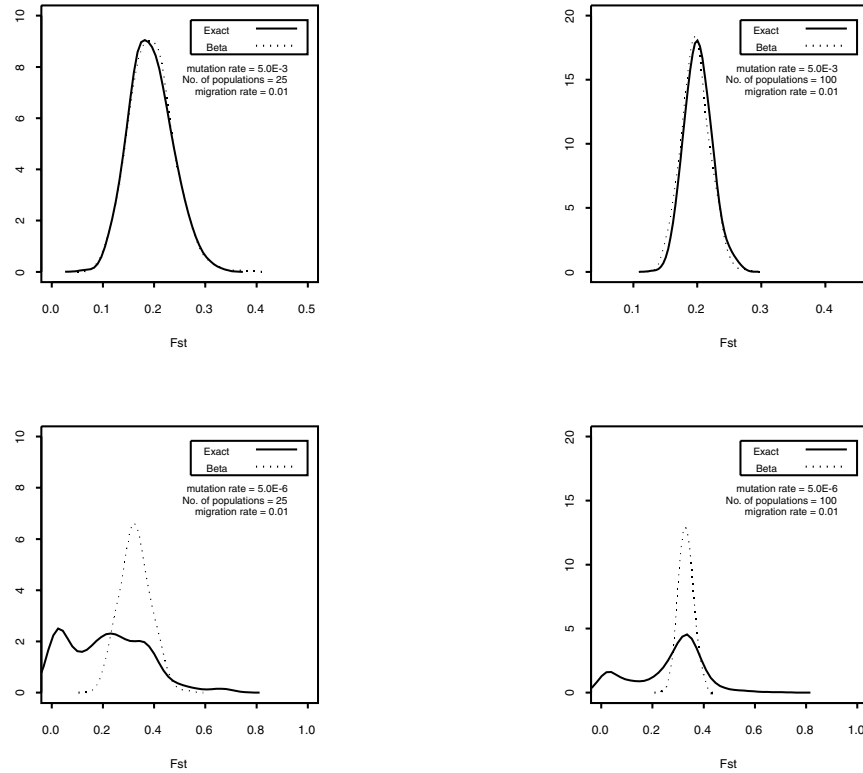


Figure 6: Comparison of the distribution $\theta^{(III)}$ based on simulation from model (1) and (2) and the beta approximation. $2N = 100$.

of the distribution of F_{st} unless mutation is relatively frequent or the number of populations exchanging genes is extremely large.

As suggested above, we regard $\theta^{(III)}$ as the appropriate parameter for inference from data. Given a sample including representatives from all k populations, Bayesian inference regarding $\theta(p_i^{(t)}, \dots, p_k^{(t)})$ could be derived from the posterior distribution of the set of \hat{p}_i (see, for example, Holsinger 1999), but empirical population geneticists rarely collect from all populations. Nonetheless, as Weir (1996) and others have emphasized, investigators are most often interested in inferences about population structure associated with all populations, not merely those that were actually sampled. Because the total number of available populations is rarely known with any degree of precision, an important task for future research is to determine what sampling strategies can help to ensure that a sample from $k' < k$ populations provides reliable information about $\theta(p_i^{(t)}, \dots, p_k^{(t)})$.

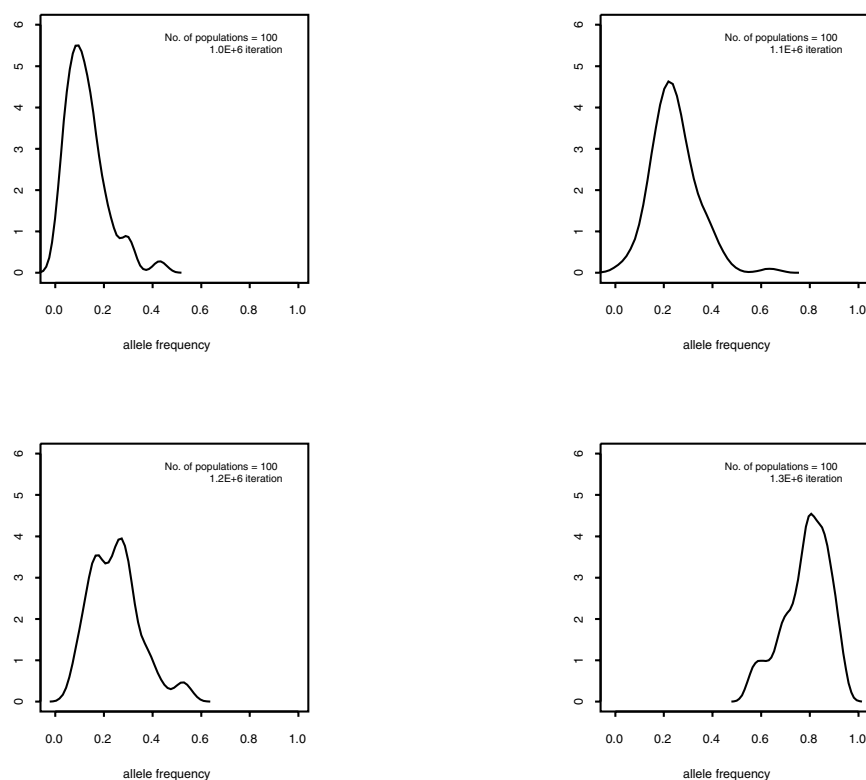


Figure 7a: Distribution of the allele frequency among populations at four different generations. $2N = 1000$, $\nu_{12} = \nu_{21} = 5.0E - 6$, and $m = 0.01$.

Acknowledgments

We are indebted to Dipak Dey for advice and discussion and to Bruce Weir and an anonymous reviewer for helpful comments on an earlier version of this paper.

References

- Bahlo, M. and R. C. Griffiths. 2000. Inference from gene trees in a subdivided population. *Theoretical Population Biology* 57:79–95.
- Bahlo, M. and R. C. Griffiths 2001. Coalescence time for two genes from a subdivided population. *Journal of Mathematical Biology* 43:397–410.
- Barton, N. H., R. B. Halliday, and G. B. Hewitt. 1983. Rare electrophoretic variants in a hybrid zone. *Heredity* 50:139–146.
- Berli, P. and J. Felsenstein. 1999. Maximum-likelihood estimation of migration rates and effective population numbers in two populations using a coalescent approach. *Genetics*

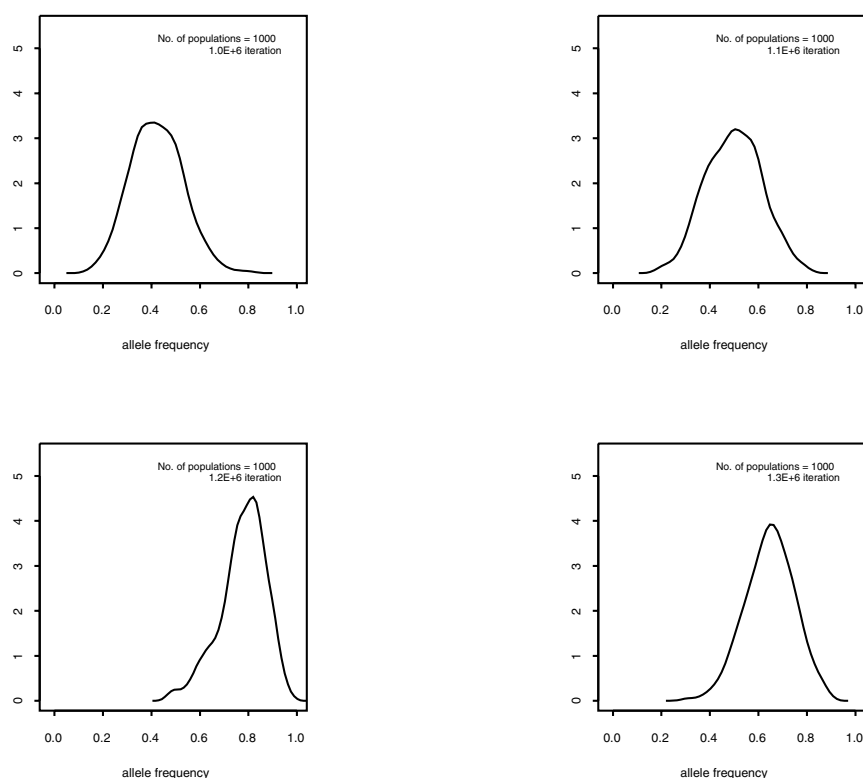


Figure 7b: Distribution of the allele frequency among populations at four different generations. $2N = 1000$, $\nu_{12} = \nu_{21} = 5.0E - 6$, and $m = 0.01$.

152:763–773.

Casella, G. and R. L. Berger. 1990. *Statistical Inference*. Duxbury Press, Belmont, CA.

Cockerham, C. C. 1969. Variance of gene frequencies. *Evolution* 23:72–84.

Cockerham, C. C., and B. S. Weir. 1987. Correlations, descent measures: drift with migration and mutation. *Proceedings of the National Academy of Sciences USA* 84:8512–8514.

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

Cox, J. T., and R. Durrett. 2001. *The stepping stone model: new formulas expose old myths*. <http://www.math.cornell.edu/durrett/SS/ss060401.ps>

Crow, J. F., and K. Aoki. 1984. Group selection for a polygenic behavioral trait: estimating the degree of population subdivision. *Proceedings of the National Academy of Sciences USA* 81:6073–6077.

Crow, J. F., and M. Kimura. 1970. *An Introduction to Population Genetics Theory*. Harper & Row, New York, NY.

- Excoffier, L. 2001. Analysis of population subdivision. In *Handbook of Statistical Genetics*, ed. D. J. Balding, M. Bishop, and C. Cannings, pp. 271–307. John Wiley & Sons, Chichester.
- Ewens, W. J. 1979. *Mathematical Population Genetics*. Springer-Verlag, New York, NY.
- Felsenstein, J. 1975. A pain in the torus: some difficulties with models of isolation by distance. *American Naturalist* 109:359–368.
- Gantmacher, F. R. 1960. *The Theory of Matrices*, vol. II. Chelsea Publishing Company, New York, NY.
- Holsinger, K. E. 1999. Analysis of genetic diversity in geographically structured populations: a Bayesian perspective. *Hereditas* 130:245–255.
- Karlin, S. 1968. Rates of approach to homozygosity for finite stochastic models with variable population size. *American Naturalist* 102:443–455.
- Kimura, M. 1964. Diffusion models in population genetics. *Journal of Applied Probability* 1:177–232.
- Kimura, M., and G. H. Weiss. 1964. The stepping stone model of population structure and the decrease of genetic correlation with distance. *Genetics* 49:561–576.
- Kingman, J. F. C. 1980. *Mathematics of Genetic Diversity*. CBMS-NSF Regional Conference Series in Applied Mathematics, no. 34. Society for Industrial and Applied Mathematics, Philadelphia, PA.
- Kingman, J. F. C. 1982a. The coalescent. *Stochastic Processes and their Applications* 13:235–248.
- Kingman, J. F. C. 1982b. On the genealogy of large populations. *Journal of Applied Probability* 19A:27–43.
- Kuhner, M. K., J. Yamato, and J. Felsenstein. 1995. Estimating effective population size and mutation rate from sequence data using Metropolis-Hastings sampling. *Genetics* 140:1421–1430.
- Malécot, G. 1948. *Les Mathématiques de l’Hérédité* Masson Cie, Paris.
- Maruyama, T. 1977. *Stochastic Problems in Population Genetics. Lecture Notes in Biomathematics No. 17*. Springer-Verlag, Berlin.
- Nagylaki, T. 1976. The decay of genetic variability in geographically structured populations. *Theoretical Population Biology* 10:70–82.
- Nagylaki, T. 1982. Geographical invariance in population genetics. *Journal of Theoretical Biology* 99:159–172.
- Nei, M. and M. W. Feldman. 1972. Identity of genes by descent within and between populations under mutation and migration pressures. *Theoretical Population Biology* 3:460–465.
- Notohara, M. 1990. The coalescent and the genealogical process in geographically structured population. *Journal of Mathematical Biology* 29:59–75.
- Notohara, M. 2000. A perturbation method for the structured coalescent with strong migration. *Journal of Applied Probability* 37:148–167.
- Notohara, M. 2001. The structured coalescent process with weak migration. *Journal of Applied Probability* 38:1–17.

- Rannala, B., and J. A. Hartigan. 1995. Identity by descent in island-mainland populations. *Genetics* 139:429–437.
- Rousset, F. 1999. Genetic differentiation in populations with different classes of individuals. *Theoretical Population Biology* 55:297–308.
- Rousset, F. 2001. Inferences from spatial population genetics. In *Handbook of Statistical Genetics*, ed. D. J. Balding, M. Bishop, and C. Cannings, pp. 239–269. John Wiley & Sons, Chichester.
- Slatkin, M. 1991. Inbreeding coefficients and coalescence times. *Genetical Research* 58:167–175.
- Slatkin, M., and N. H. Barton. 1989. A comparison of three indirect methods for estimating average levels of gene flow. *Evolution* 43:1349–1368.
- Takahata, N. 1989. Gene genealogy in three related populations: consistency probability between gene and population trees. *Genetics* 122:957–966.
- Wehrhahn, C. F. 1989. Proceedings of the ecological genetics workshop. *Genome* 31:1098–1099.
- Wehrhahn, C. F., and R. Powell. 1987. Electrophoretic variation, regional differences, and gene flow in the coho salmon (*Onchorhynchus kisutch*) of southern British Columbia. *Canadian Journal of Fisheries and Aquatic Sciences* 44:822–831.
- Weir, B. S. 1996. *Genetic Data Analysis II*. Sinauer Associates, Sunderland, MA.
- Weir, B. S. and C. C. Cockerham. 1984. Estimating F -statistics for the analysis of population structure. *Evolution* 38:1358–1370.
- Weir, B. W. and W. G. Hill. 2002. Estimating F -statistics. *Annual Review of Genetics* (in press).
- Weiss, G. H., and M. Kimura. 1965. A mathematical analysis of the stepping stone model of genetic correlation. *Journal of Applied Probability* 2:129–149.
- Wright, S. 1931. Evolution in Mendelian populations. *Genetics* 16:97–159.
- Wright, S. 1951. The genetical structure of populations. *Annals of Eugenics* 15:323–354.
- Wright, S. 1969. *Evolution and the Genetics of Populations. Volume 2, The Theory of Gene Frequencies*. University of Chicago Press, Chicago, IL.