Maximum-Likelihood Estimation of Migration Rates and Effective Population Numbers in Two Populations Using a Coalescent Approach

Peter Beerli and Joseph Felsenstein

Department of Genetics, University of Washington, Box 357360, Seattle, Washington 98195-7360

ABSTRACT

A new method for the estimation of migration rates and effective population sizes is described. It uses a maximum-likelihood framework based on coalescence theory. The parameters are estimated by Metropolis-Hastings importance sampling. In a two-population model this method estimates four parameters: the effective population size and the immigration rate for each population relative to the mutation rate. Summarizing over loci can be done by assuming either that the mutation rate is the same for all loci or that the mutation rates are gamma distributed among loci but the same for all sites of a locus. The estimates are as good as or better than those from an optimized F_{ST}-based measure. The program is available on the World Wide Web at http://evolution.genetics.washington.edu/lamarc.html/.

SEVERAL methods for the estimation of migration rates between subpopulations have been proposed. We can subdivide them into two very different approaches: (1) marking individuals and tracking their individual movements and then extrapolating these individual movements to migration rates; or (2) surveying genetic markers in the populations of interest and calculating a migration rate from allele frequencies or sequence differences. Of course, one should be aware that these two approaches do not estimate the same quantity: approach 1 estimates the actual "instantaneous" migration rate of individuals, whereas approach 2 reflects an average over a time period whose length is determined by the rate of mutation per generation of the locus under study or by the time scale of genetic drift. The genetic approach tends to gives a lower estimate than the individual migration rates approach because the method is looking at changes that become established in the subpopulation gene pool.

Current estimation methods for genetic data are methods such as those related to F_{ST} (*e.g.*, Slatkin 1991; Slatkin and Hudson 1991), the rare allele approach of Slatkin (1985), a maximum-likelihood method using gene frequency distributions (Rannala and Hartigan 1996; Tufto *et al.* 1996), and approaches based on coalescent theory (Kingman 1982a,b), such as the cladistic approach of Slatkin and Maddison (1989), the method outlined in Wakeley (1998), and maximum likelihood using coalescent theory (Nath and Griffiths 1993, 1996). We describe here a new method using a maximum-likelihood- and coalescent theory-based approach.

Our method integrates over all possible genealogies

and migration events. This should be superior to pairwise estimators such as those using F_{ST} or related statistics (*cf.* Felsenstein 1992b) and also more powerful than the cladistic approach of Slatkin and Maddison (1989), which needs to know the true genealogy. Our approach estimates similar parameters as the program of Bahlo and Griffiths (http://www.maths.monash.edu.au/~mbahlo/mpg/gtree.html/), which is based on the work of Griffiths and Tavaré (1994) and Nath and Griffiths (1996). It differs in the way we search the genealogy space, and our methods support mutation models for different types of data: the infinite allele model for electrophoretic markers, a one-step model for microsatellites, and a finite-sites model for nucleotide sequences. The sequence model is more useful than the infinite sites model, which forces the researcher to discard data when there are multiple mutations at the same sites.

MODEL

We propose a method to make a maximum-likelihood estimate of population parameters for geographically subdivided populations. The general outline of such estimates involves extending coalescence theory (Kingman 1982a,b) to include migration events. Migration models with coalescents were first developed by Takahata (1988) and Takahata and Slatkin (1990) for two gene copies and discussed more generally for n gene copies in two populations by Hudson (1990), with generalization to multiple populations and different models of migration by Notohara (1990). Nath and Griffiths (1993, 1996) used this migration-coalescent process for maximum-likelihood estimation of one parameter, the effective number of migrants $\bar{\gamma} = 4N_e m$, where N_e is the effective population size and m is the migration rate per generation.

Corresponding author: Peter Beerli, J255 Health Sciences, Department of Genetics, University of Washington, Box 357360, Seattle, WA 98195-7360. E-mail: beerli@genetics.washington.edu
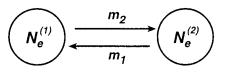
Figure 1.—Two-population model with four parameters: m_1 is the migration rate per generation from population 2 to 1, m_2 from population 1 to 2, and $N_e^{(i)}$ is the effective population size.

Our migration model consists of two populations. These have effective population sizes $N_e^{(1)}$ and $N_e^{(2)}$, which need not be equal. The populations exchange migrants at rates m_1 and m_2 per generation (Figure 1).

Because we observe only genetic differences and do not know the mutation rate μ, we must absorb μ into the parameters so that we use as our parameters

$$\Theta_i = 4N_e^{(i)}\mu,$$

and

$$\mathcal{M}_i = \frac{m_i}{\mu},$$

and occasionally write

$$\gamma_i = \Theta_i\mathcal{M}_i = 4N_e^{(i)}m_i.$$

We make the following assumptions: that diploid individuals in each population reproduce according to any standard population genetic model that converges to the same diffusion equation as a Wright-Fisher model, that populations have a constant population size through time so that they do not grow or shrink, and that the mutation rate μ is constant. These are also the assumptions used by Kingman (1982a) for coalescence theory. We use the convention with sequence data that μ is the rate of mutation per generation per site, whereas with microsatellite or enzyme electrophoretic data we take μ to be the rate of mutation per generation per locus. To emphasize this distinction between sites and loci, we use a capital Θ in the first case and a small θ for the latter ones. The mutation rate μ can vary among loci according to a gamma distribution. In addition we also need to assume that the two populations exchange migrants with constant rates per generation.

The likelihood formula for this family of methods (Felsenstein 1988, 1992a; Kuhner *et al.* 1995, 1998) is based on the product between the genealogy likelihood Prob($D|G$) (*e.g.*, Felsenstein 1973; Swofford *et al.* 1996) and the prior probability of the coalescent genealogy Prob($G|\mathcal{P}$) (Kingman 1982a,b)

$$L(\mathcal{P}) = \sum_G \text{Prob}(D|G)\,\text{Prob}(G|\mathcal{P}), \quad (1)$$

where the parameters $\mathcal{P}$ are

$$\mathcal{P} = [\Theta_1,\ \Theta_2,\ \mathcal{M}_1,\ \mathcal{M}_2].$$

Takahata and Slatkin (1990) found that Kingman's

results could not be extended to cases with geographical structure. They were unable to obtain the distribution of times to coalescence in the presence of migration. We have avoided this difficulty by having the genealogy G specify not only the coalescences but also the times and places of migration events. With this information in G, its probability density becomes for two populations a product of terms for the intervals between events in the genealogy

$$\text{Prob}(G|\mathcal{P}) = \prod_{i=1}^{2}\left(\left[\frac{2}{\Theta_i}\right]^{v_i}\mathcal{M}_i^{w_i}\right)\prod_{j=1}^{T}p_j, \quad (2)$$

where

$$p_j = \exp\left(-u_j\sum_{i=1}^{2}\left(k_{ji}\mathcal{M}_i + \frac{k_{ji}(k_{ji}-1)}{\Theta_i}\right)\right). \quad (3)$$

Prob($G|\mathcal{P}$) is obtained by computing for each interval the probability that nothing happens during the interval times the probability of the particular event at the bottom (beginning) of the interval. The variable v_i is the total number of coalescences in subpopulation i and w_i is the sum of the numbers of migration events to subpopulation i over all time intervals T. Furthermore, p_j is the probability that no event occurs during time interval j, u_j is the length of time interval j, and k_{ji} is the number of lineages in subpopulation i during time interval j.

If we extend the parameter estimation from a single locus to many unlinked loci, we face the additional problem that, although μ may be constant per locus, it can vary between loci. Neglecting variation of μ, combining $\mathcal{L}$ loci using the likelihood framework is simple:

$$L(\overline{\mathcal{P}}) = \prod_{l=1}^{\mathcal{L}}L_l(\mathcal{P}). \quad (4)$$

On the other hand, if we assume that μ follows a gamma distribution and therefore can vary between loci, we need to include an additional parameter, the shape parameter α of this distribution. The other parameter of this gamma distribution is a scale parameter that can be conveniently chosen to be $\overline{\mu}/\alpha$. By parameterizing in this way, $1/\alpha$ is the squared coefficient of variation of $\overline{\mu}$. We cannot estimate the mean mutation rate $\overline{\mu}$ of this gamma distribution directly. As N_e is the same for all loci, we can use an arbitrary Θ (we have chosen the effective population size of the first population) and scale the other parameters relative to it. We replace μ in the parameters Θ_i and $\mathcal{M}_i$ with a new variable, the mutation rate x. Θ_ξ is the mutation rate x scaled by $4N_e^{(1)}$. We integrate over all possible values of Θ_ξ instead of x. This does not change the shape of our likelihood surface, because $N_e^{(1)}$ in Θ_ξ is constant, so that we have

$$L(\overline{\mathcal{P}},\alpha) = \prod_{l}^{\mathcal{L}}\left(\frac{1}{\Gamma(\alpha)\beta^\alpha}\int_0^\infty e^{-\Theta_\xi/\beta}\Theta_\xi^{\alpha-1}L_l(\mathcal{P}')\,d\Theta_\xi\right), \quad (5)$$

where

$$\beta = \frac{\Theta_1}{\alpha},$$

$$\mathcal{P}' = \left[\Theta_\xi, \; \Theta_2\frac{\Theta_\xi}{\Theta_1}, \; \mathcal{M}_1^{(\xi)}, \; \mathcal{M}_2^{(\xi)}\right],$$

$$\Theta_\xi = 4N_e^{(1)}x,$$

and

$$\mathcal{M}_i^{(\xi)} = \frac{\gamma_i\Theta_1}{\Theta_i\Theta_\xi}.$$

This integral over all mutation rate values is implemented using a discrete gamma approximation with 100 intervals (cf. Yang 1994).

Note that the gamma distribution described here is the distribution of μ across loci, not the more commonly used gamma distribution of rates across sites within a locus. If needed the latter can be approximated by Hidden Markov model methods in the case of DNA sequences.

IMPLEMENTATION

To calculate the likelihood $L(\mathcal{P})$ for the parameters $\mathcal{P}$ using (1) we ought to sum over all possible genealogies, including all topologies with all possible branch lengths and with all possible migration scenarios. This is not possible, as even with two tips there are an infinite number of possible sequences of migration events with different branch length. We have to resort to an approximation of the likelihood function using a Metropolis-Hastings Markov chain Monte Carlo approach (Metropolis et al. 1953; Hammersley and Handscomb 1964; Hastings 1970; Kuhner et al. 1995). The derivation of the importance sampling function is shown in the appendix.

The parameter estimation is simple in principle: we have to find the maximum of the likelihood function, which has five or six dimensions, depending on whether μ is allowed to vary between loci. This is done by a modified damped Newton-Raphson method (Dahlquist and Björk 1974) using explicit equations for the first and second derivatives.

The genealogy sampler is of more interest. A large sample of genealogies $G_1, G_2, \ldots, G_m$ is drawn from a distribution whose density is proportional to $\mathrm{Prob}(D|G)$ $\mathrm{Prob}(G|\mathcal{P}_0)$. This sampling scheme is biased toward genealogies that contribute more to the likelihood with a given parameter set $\mathcal{P}_0$. We find the relative likelihood at other $\mathcal{P}$ values by dividing by the density from which we sampled the genealogies (see also appendix). If m is the number of sampled genealogies, we get

$$\frac{L(\mathcal{P})}{L(\mathcal{P}_0)} \cong \frac{1}{m}\sum_i^m \frac{\mathrm{Prob}(D|G_i)\;\mathrm{Prob}(G_i|\mathcal{P})}{\mathrm{Prob}(D|G_i)\;\mathrm{Prob}(G_i|\mathcal{P}_0)}$$

$$\cong \frac{1}{m}\sum_i^m \frac{\mathrm{Prob}(G_i|\mathcal{P})}{\mathrm{Prob}(G_i|\mathcal{P}_0)}. \tag{6}$$

In the program Coalesce (Kuhner et al. 1995) a region of internal nodes in the genealogy is erased and rebuilt according to the coalescent model. This scheme is not applicable to migration models. We adopt a new scheme that is much more general and that can be used in any extension to the coalescent model. The first genealogy for each locus is constructed with a UPGMA method (as implemented by M. K. Kuhner and J. Yamato in Felsenstein 1993), and the minimal necessary migration events are inserted using a Fitch parsimony algorithm (Fitch 1971). The times for coalescent events or migration events on this first genealogy are calculated with (3) using a uniform random number for p_j and solving for u_j, and parameter values, which are either guessed by the researcher or calculated using an F_{ST}-based method (see appendix).

The genealogies are sampled as follows (cf. Figure 2): A coalescent node or tip z on this genealogy is chosen at random, the lineage below it is dissolved, and the node is used as the starting point to simulate the ancestry using a migration-coalescent process through a series of time intervals (Figure 2). The other lineages do not change and are taken as given and form the partial genealogy G_p. For the further description of the re-arrangement process we use "up" and "top" for being or moving closer to the tips of the genealogy and "down" and "bottom" for being or moving toward the root of the genealogy. Starting at the chosen node z (see Figure 2), we draw a new time interval by solving for u_j in

$$p_j' = \exp\left(-u_j\left[\mathcal{M}_i + \frac{2k_{ji}'}{\Theta_i}\right]\right) \tag{7}$$

after substituting a uniform deviate for p_j', where k_{ji}' is the number of lineages of population i in the partial genealogy G_p during the time interval j. If this new time is further back in time than the bottom of the time interval j on the partial genealogy G_p, the active lineage advances down to that time and the simulation process starts again. When the newly drawn time is in the active time interval, an event, either a migration or coalescence, is drawn according to their relative probabilities of occurrence, e.g.,

$$\mathrm{Prob}(\text{Migration event in population } i) = \frac{\mathcal{M}_i}{\mathcal{M}_i + 2k_{ji}'/\Theta_i}. \tag{8}$$

If a migration event occurs the process moves down to the time when the migration event happened. Below a migration event the active lineage is in the other population and the migration-coalescent process continues (see also Figure 2C). After a coalescent event occurs the process stops. The transformation from one genealogy into another allows change from any genealogy over several steps into any other and therefore allows the process to search the whole genealogy space. On the other hand, the resimulation of a single line guarantees that newly found genealogy is similar to the old genealogy.
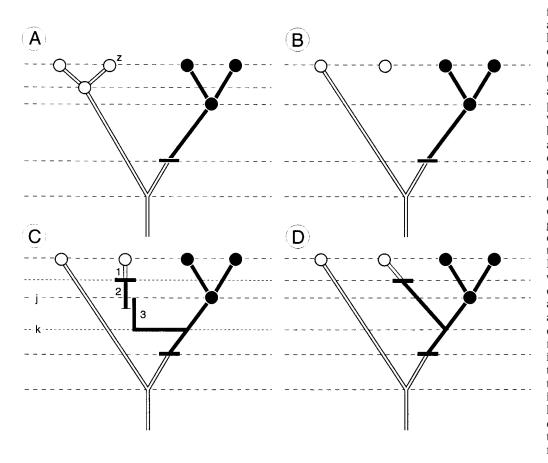
Figure 2.—Transition from an old genealogy G_o to a new genealogy G_n. Dotted lines are the times of coalescences or migrations. (A) Genealogy G_o with migration. The black bar marks a migration from the white population to the black or vice versa; z is the node to be picked. (B) Partial genealogy G_p after drawing the coalescent node z at random and dissolving the branch to the next coalescent below. (C) Simulation of the coalescent with migration. One possible outcome with three consecutive steps is shown: (1) using Equation 7 a new time interval is drawn and a migration event from white to black is also drawn (Equation 8) and the lineage is extended down to that event; (2) a new time interval is drawn: it extends too far back, so the lineage advances down to the time j; (3) a new time interval is drawn with a coalescent event at its bottom end. The process stops at time k. (D) the final configuration G_n.

In cases where all the lineages have not coalesced by the time of the root of the partial genealogy, simulation on the active lineage and the bottom lineage of G_p continues until the lineages coalesce (Figure 3). When simulating on two lineages we draw new time intervals using (3) instead of (7). There is one exception to this rule at the bottom of the genealogy: if by chance the dissolved lineage is the bottommost lineage of one of the populations on G_o, we keep simulating with a single active lineage below the root of G_p until the former root of G_o is reached and then start to simulate on both of the two remaining lines. The lineage between the root

of G_p and the old root on G_o on the partial genealogy has a known history, so there is no need to simulate this history again.

A change to the newly found genealogy G_n is accepted with probability

$$r = \min(1, r_M r_H), \qquad (9)$$

where r_M is the Metropolis term

$$r_M = \frac{\text{Prob}(D|G_n)\ \text{Prob}(G_n|\mathcal{P})}{\text{Prob}(D|G_o)\ \text{Prob}(G_o|\mathcal{P})} \qquad (10)$$
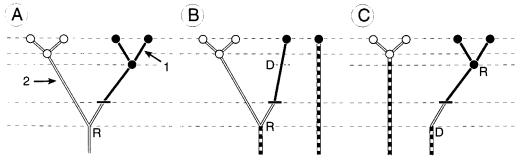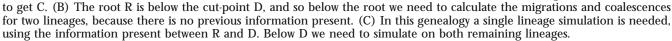
and r_H is the Hastings term

Figure 3.—Simulation on one or two lineages at the bottom of a genealogy. Striped lineages are active. On these lineages the migration-coalescent process is used to find new time intervals and events. (A) old genealogy G_o, with 1 indicating the branch to be cut to get genealogy B, and 2 indicating the branch to be cut to get C. (B) The root R is below the cut-point D, and so below the root we need to calculate the migrations and coalescences for two lineages, because there is no previous information present. (C) In this genealogy a single lineage simulation is needed, using the information present between R and D. Below D we need to simulate on both remaining lineages.

$$r_H = \frac{\text{Prob}(G_o|G_n)}{\text{Prob}(G_n|G_o)}. \quad (11)$$

This acceptance probability r is based on the work of Hastings (1970). In our specific implementation it can be simplified. The scheme for changing the genealogy creates a new genealogy from an old one by first randomly removing a branch with all its possible migration events, which creates a partial genealogy G_p. The probability of a particular change is only governed by the number of coalescent nodes in the genealogy, and this number is constant so that the probabilities $\text{Prob}(G_p|G_n)$ and $\text{Prob}(G_p|G_o)$ are equal. For a specific triplet of original, partial, and new genealogies G_o, G_p, and G_n

$$\text{Prob}(G_n|G_o) = \text{Prob}(G_p|G_o)\,\text{Prob}(G_n|G_p) \quad (12)$$

holds. The resimulation process (7)–(8) is driven only by the parameters $\mathcal{P}$ and so $\text{Prob}(G|G_p)$ is proportional to $\text{Prob}(G|\mathcal{P})$ so that

$$r_H = \frac{\text{Prob}(G_p|G_n)\,\text{Prob}(G_o|G_p)}{\text{Prob}(G_p|G_o)\,\text{Prob}(G_n|G_p)}$$

$$= \frac{\text{Prob}(G_o|\mathcal{P})}{\text{Prob}(G_n|\mathcal{P})}. \quad (13)$$

This results in cancellation of the $\text{Prob}(G|\mathcal{P})$ terms in r, so that we have

$$r = \min\left(1, \frac{\text{Prob}(D|G_n)}{\text{Prob}(D|G_o)}\right). \quad (14)$$

If the genealogy is accepted, it is the starting point for a new cycle; otherwise the previous one continues to be used. After having sampled a large number of genealogies we estimate the parameters by maximizing (6). In theory a single long Markov chain would be enough to find the genealogies that contribute most to the likelihood, but if we start from a very bad $\mathcal{P}_0$ we can need an excessively long time to find this region. If the likelihood of a new genealogy is only slightly better than that of the old one, the search is moving nearly randomly and can spend a long time in regions that do not contribute much to the final likelihood. The sampler will move more quickly if the likelihood surface is steep. If we start with bad parameters there is some chance that the likelihood surface is locally very flat. To overcome this, we restart a new Markov chain with the improved $\mathcal{P}$ and the last genealogy of the most recent chain. Our implementation uses a number of short chains in which the parameter estimates are based only on a few genealogies to find regions with good parameter values. It then uses an arbitrary number of long chains to get good estimates of the parameters. All but the last chain are discarded and not used for the final result. Kuhner *et al.* (1998) have not found a decrease in performance compared to the scheme of taking several chains into account for the final result. If we run this genealogy sampling process long enough and with enough chains, the sampler will find its way to the region

with the biggest contribution to the likelihood and therefore approximate the maximum-likelihood estimate of our parameters $\mathcal{P}$. The main problem with importance sampling schemes is that little is known about how we can be sure that we have sampled enough genealogies from the region that contributes most to our likelihood function. It is common practice in our group to run 10 "short" chains and then 2 "long" chains. Using random values for the first parameter set, we have found that for a given data set the results converge to a value that is not dependent on these initial parameter values when the total numbers of sampled genealogies exceed 30,000 genealogies (data not shown). These are 10 short chains with 1000 genealogies and 2 long chains with 10,000 genealogies sampled. It seems to be sufficient to deliver acceptable estimates for simulated data and averages of those, although we would recommend choosing a better start parameter than a random guess and running the program at least twice as long. The number of genealogies needed for an accurate estimate is certainly dependent on many factors, such as starting parameters, first genealogy, number of sampled individ-

TABLE 1

Simulation of 100 single-locus datasets with 25 sampled individuals for each population and 500 or 1000 bp, respectively, for different known parameter combinations

Truth		Means		SD		
Θ	γ	Θ	γ	Θ	γ	N
		500 bp				
0.001	0.1	0.0011	0.86	0.0001	0.42	100
	1.0	0.0010	2.91	0.0001	0.72	100
	10.0	0.0013	15.84	0.0002	2.99	81
0.01	0.1	0.0099	0.19	0.0011	0.03	100
	1.0	0.0104	1.30	0.0011	0.18	100
	10.0	0.0131	14.61	0.0016	2.10	100
0.1	0.1	0.0976	0.17	0.0101	0.03	100
	1.0	0.1008	0.92	0.0105	0.12	100
	10.0	0.1067	8.06	0.0113	0.93	100
		1000 bp				
0.001	0.1	0.0008	0.29	0.0001	0.12	100
	1.0	0.0010	2.98	0.0001	0.78	100
	10.0	0.0017	21.41	0.0004	4.88	92
0.01	0.1	0.0102	0.13	0.0011	0.02	100
	1.0	0.0103	1.03	0.0011	0.13	100
	10.0	0.0112	10.40	0.0012	1.32	100
0.1	0.1	0.0985	0.18	0.0102	0.03	100
	1.0	0.1013	0.94	0.0106	0.12	100
	10.0	0.1042	7.83	0.0108	0.87	100

SD, standard deviation. $\Theta = 4N_e\mu$, where N_e is the effective population size and μ is the mutation rate. $\gamma = 4N_e m$, where m is the migration rate. In two cases the numbers of replicates N are 81 and 92, respectively, instead of 100: the program failed to find the maximum of the likelihood surface and these cases were excluded. The values shown are for population 1; those for population 2 are similar.

TABLE 2

Influence of number of sites and number of loci on parameter estimates

	Loci							
	1 [100]		2 [50]		5 [20]		10 [10]	
bp	Θ	γ	Θ	γ	Θ	γ	Θ	γ
	Means							
500	0.0102	1.13	0.0088	1.06	0.0097	0.99	0.0108	0.97
1,000	0.0098	1.05	0.0104	1.10	0.0104	1.16	0.0100	0.77
2,000	0.0107	0.99	0.0098	1.02	0.0104	1.00	0.0100	0.83
5,000	0.0103	0.96	0.0103	0.90	0.0104	0.87	0.0107	0.89
10,000	0.0102	0.88	0.0103	0.89	0.0100	0.82	0.0109	0.83
	Standard deviations							
500	0.0011	0.17	0.013	0.19	0.0023	0.25	0.0036	0.34
1,000	0.0010	0.13	0.0015	0.20	0.0024	0.28	0.0034	0.26
2,000	0.0011	0.13	0.0014	0.18	0.0024	0.25	0.0034	0.29
5,000	0.0011	0.12	0.0015	0.15	0.0024	0.21	0.0036	0.31
10,000	0.0011	0.11	0.0015	0.14	0.0023	0.19	0.0037	0.29

A total of 25 individuals were sampled for each population. For each mean a total of 100 loci were examined. The numbers of replicates are in brackets. The true values are $\Theta = 0.01$ and $\gamma = 1.0$. The values shown are for population 1; those for population 2 are similar.

uals, variability in the data set, and number of migration events.

SIMULATION RESULTS

We have simulated genealogies with known true parameters $\mathcal{P}_T$ using a coalescent with migration and evolved data according to our data models along the branches starting from the root of the genealogy (Hudson 1983). The data that resulted at the tips were used as the input data for our method. The mutation rate was held constant for all simulations except one, in which we used a gamma-distributed mutation rate with shape parameter $\alpha = 1.0$. For all other simulations we used 10 short chains sampling 1000 genealogies, keeping 100 of these genealogies for the parameter estimation and 3 long chains sampling 10,000 genealogies, using 1000 of them. The last chain delivered the parameters shown in the tables, and all other chains were discarded.

For the comparison with an F_{ST}-based estimator (see appendix) we used a symmetrical parameter setup for the simulations, $\Theta_1 = \Theta_2$ and $\gamma_1 = \gamma_2$. However, the parameters that were estimated were not constrained to be symmetric.

Our simulations generally recover the true Θ_T very well (Table 1). With 1000 bp the means are better than with 500 bp. The means for γ are not close to the true γ_T. With low Θ_T values γ is grossly overestimated. The upward bias in γ is huge with 500-bp data and shrinks when longer sequences are used (Tables 1 and 2).

Sequences in two populations can be similar to each

other due to migration or due to low variation caused by a small Θ. The simulated datasets with low Θ sometimes show by chance no variation at all in short sequences. In these cases one would expect the estimate of Θ to be 0.0 and the estimate of $\mathcal{M}$ to be indeterminate. The likelihood surface for such data is very flat and some of the runs (Table 1) failed to find a maximum. These cases were omitted from the tables. If the migration parameter γ is high, it is consistently underestimated with high Θ. This is caused by our allowing only a limited number of migration events per genealogy in the Markov chain Monte Carlo runs, which were 1000 migration events per genealogy, so that solutions with higher numbers of migration events were not proposed. Our method delivers similar estimates for Θ and similar or better estimates for γ than an F_{ST}-based estimator (Table 3).

The number of cases in which the F_{ST}-based estimator (Table 3) is usable is often dramatically lower than that with the ML estimator (Table 1; see also appendix). For example, with $\Theta_T = 0.001$ and $\gamma_T = 10$ the F_{ST}-based estimator could be used in only 45 out of 100 runs. This behavior seems not to improve with longer sequences. Increasing the number of sites and the number of loci helps to reduce the variance of the ML estimates (Table 2 and Figure 4). Estimates of γ from datasets with very long sequences are biased and are smaller than γ_T. This is the result of low acceptance rates during a run, when the starting genealogy is so well defined by the data that in our simulation runs the chains were too short and too few different migration scenarios were tried. The estimation of nonsymmetrical true parameters works

TABLE 3

F_{ST}-based estimator with three parameters ($\Theta \equiv \Theta_1 \equiv \Theta_2, \mathcal{M}_1, \mathcal{M}_2$, see also appendix)

Truth		Means		SD		
Θ	γ	Θ	γ	Θ	γ	N
		500 bp				
0.001	0.1	0.0009	0.70	0.0002	0.32	68
	1.0	0.0011	7.39	0.0002	4.60	63
	10.0	0.0011	19.45	0.0002	10.04	45
0.01	0.1	0.0087	129.40	0.0012	130.10	80
	1.0	0.0091	4.11	0.0012	1.04	77
	10.0	0.0118	14.40	0.0018	4.73	53
0.1	0.1	0.0720	0.70	0.0090	0.39	75
	1.0	0.0901	3.34	0.0107	0.63	84
	10.0	0.0868	15.96	0.0122	4.43	59
		1000 bp				
0.001	0.1	0.0008	0.63	0.0001	0.23	74
	1.0	0.0012	8.07	0.0002	3.83	70
	10.0	0.0012	7.94	0.0002	2.26	55
0.01	0.1	0.0101	0.69	0.0018	0.25	76
	1.0	0.0094	2.51	0.0012	0.58	78
	10.0	0.0091	602.17	0.0014	588.26	59
0.1	0.1	0.0726	0.71	0.0900	0.28	77
	1.0	0.0869	13.38	0.0102	6.65	85
	10.0	0.0940	43.11	0.0140	25.07	54

Means and standard deviations of 100 single-locus datasets with 25 individuals were sampled for each population and 500 bp or 1000 bp, respectively. In all cases in which the homozygosity within a population is not less than the homozygosity between populations, the F_{ST} estimator fails. Those replicates have been discarded. N shows the number of replicates remaining. The values shown are for population 1; those for population 2 are similar.

quite well (Table 4). We encounter similar biases as with symmetrical parameters: with low Θ the γ estimates are biased upward and with high Θ they are biased downward.

The assumption that mutation rate varies according to a gamma distribution adds more noise to the estimation. All biases are similar to the ones shown in the tables. The estimation of the shape parameter $1/\alpha$ is certainly not very precise with only a few loci (Figure 5). One of the runs shown in Figure 5 obviously contains almost no information about $1/\alpha$ and its log-likelihood curve is nearly flat, so that almost any value of $1/\alpha$ can be accepted for this dataset. The increase of the log-likelihood values with very small $1/\alpha$ and far from the maximum is due to imprecisions in the calculation using the discrete gamma approximation. In evaluations using a more exact but much slower quadrature scheme (Sikorski *et al.* 1984) the peaks of the log-likelihood curves are at similar values, but log-likelihood for very low $1/\alpha$ values are monotonically decreasing and not,

as with the discrete gamma approximation, slightly increasing.

DISCUSSION

The estimation of the migration parameter γ seems to be rather imprecise, and this is true for both migrate and F_{ST} (*e.g.*, Tables 1 and 3). The standard deviations for γ are large for single-locus estimates. More sites per locus give better estimates but the improvement of the variance is small, and given the obstacles of sequencing long stretches of DNA for several individuals one might better invest in the investigation of an additional unlinked locus. Each new locus has its own genealogy that is not correlated with that of the other loci and so increases the power of estimating parameters more, as we can see in Table 2.

There is considerable bias in the estimates of γ when datasets either have no or little genetic variation or have very high variation. Datasets with almost no variation inflate $\mathcal{M}$ and produce an upward bias. The likelihood surfaces become very flat because of the lack of variation, and almost any migration parameter value is possible. These problems can be overcome by adding more variation, which for low true population size Θ_T means either adding more base pairs or adding more individuals. With high Θ_T and high true migration parameter γ_T, the estimates are biased downward because of our need to truncate the number of migration events in the reconstructed genealogies at some high arbitrary number. If we did not bias against high numbers of migration events, the program could add more and more migration events and would eventually crash the computer by running out of memory or would run too long.

In single runs we can also see a "fatal attraction" to 0.0 if the true parameters are very close to 0.0. The genealogy sampler is using the current parameter values to propose coalescences and migration events. If one or more of these parameters are very close to 0.0, then it can be seen by inspecting (2) that as a result our procedure will either most often propose an immediate coalescent event if one of the Θ is close to 0.0 or rarely propose an immigration event with very low γ. If a chain never proposes any instances of an event, its rate of occurrence will be estimated to 0.0, and this situation may persist indefinitely, a fatal attraction. Even if the parameters are prevented from becoming exactly zero this means that it may take large amounts of simulation to escape these values.

Because long sequences define the genealogy very well, our Markov chain Monte Carlo sampler will often reject newly proposed genealogies and only a few different genealogies are sampled for the parameter estimation. Therefore, the program does not readily explore the whole migration-genealogy space and tends to stick to the good starting genealogy. This generates a bias in

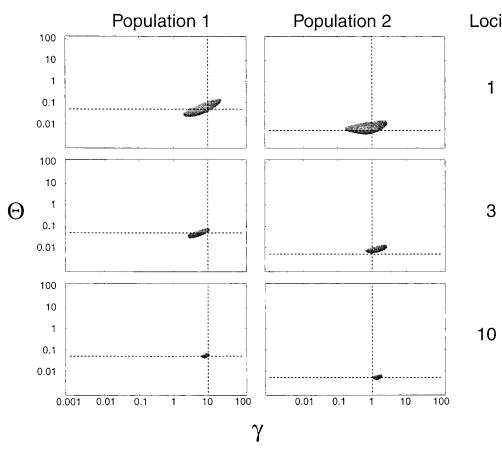## Population 1   Population 2   Loci

Figure 4.—Example of estimation of population parameters for two populations with 1, 3, and 10 loci. The graphs are cross-sections through the parameter space passing through the peak of the likelihood surface. The axes are on a logarithmic scale. The dashed lines give the true values of the parameter values: $\Theta_1 = 0.05$, $\Theta_2 = 0.005$, $\gamma_1 = 10.0$, $\gamma_2 = 1.0$. The gray area is an approximate 95% confidence set based on a likelihood-ratio test criterion defined as the range of parameter values with log-likelihood values equal to or higher than the maximum -0.5×9.4877 ($\chi^2_{\text{d.f.} = 4; \alpha = 0.05}$).

$\mathcal{M}$ because we have started with a genealogy that has a minimal number of migration events on it (Table 2). We may be able to overcome this mixing problem by adding a "heating" scheme to our importance sampling. Currently the only way to overcome this is to run the chains much longer than the defaults (for practical guidance see Beerli 1997).

Summing over loci assuming that the mutation rate is constant between loci delivers much better estimates of the parameters than single-locus estimation. This model is rather unnatural for real data, especially microsatellite data or electrophoretic data. Summing over loci using a gamma-distributed mutation rate and estimating the shape parameter $1/\alpha$ increases the variation of the

TABLE 4

Simulation with unequal known parameters of 100 two-locus datasets with 25 sampled individuals for each population and 500 bp per locus

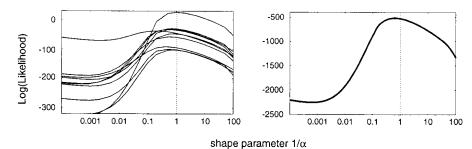|	Population 1		Population 2	
	Θ	γ	Θ	γ
Truth	0.0500	10.00	0.0050	1.00
Means	0.0476	8.35	0.0048	1.21
SD	0.0052	1.09	0.0005	0.15

SD, standard deviation.

parameters but allows more realistic estimation of the parameters with several loci.

Conclusion: Our method based on coalescents with migration delivers similar or better estimates than the F_{ST} method based on the expectations derived by Maynard Smith (1970), which for certain data also produces fine results. Qualitative comparison with other methods than the F_{ST} measure, for example Rannala and Hartigan (1996), or the cladistic approach of Slatkin and Maddison (1989; see also Hudson *et al.* 1992, their Table 1) shows that all these methods have a tendency to overestimate γ when Θ is not very high and all methods have quite large variances. The maximum-likelihood estimators have smaller variances than the other approaches. But the biggest drawback of the F_{ST} methods compared to ML methods seems to be their inability to estimate population size and migration rate for data where the homozygosity between populations happens to be equal to or less than the homozygosity within a population (see also Hudson *et al.* 1992).

Direct comparison of the different methods is currently difficult because each one estimates a different number of parameters or uses a different migration model. We expand our model to accept more populations and different migration schemes. This should then facilitate direct comparison. We believe that taking the history of mutation into account and summing over all possible genealogies gains the most information from

Figure 5.—Log-likelihood curves of shape parameter $1/\alpha$. The true $1/\alpha$ is 1.0. The curves were evaluated with Θ and γ parameters from the maximum-likelihood estimate. Left, 10 runs with 10 loci each. Right, sum of the log-likelihood curves over 10 runs with 10 loci each.

the data. We believe that our approach and the approaches of Nath and Griffiths (1996) and Bahlo and Griffiths (http://www.maths.monash.edu.au/~mbahlo/mpg/gtree.html/) will produce better estimates than other approaches.

The methods described here are available in the program Migrate over the World Wide Web at http://evolution.genetics.washington.edu/lamarc.html/. We distribute it as C source code and also as binaries for several platforms including PowerMacintoshes and Windows 95/98 for Intel-compatible processors.

LITERATURE CITED

Beerli, P., 1997 Migrate documentation version 0.3. Distributed over the Internet: http://evolution.genetics.washington.edu/lamarc.html/.

Chib, S., and E. Greenberg, 1995 Understanding the Metropolis-Hastings algorithm. Am. Stat. **49:** 327–335.

Dahlquist, G., and A. Björk, 1974 *Numerical Methods.* Prentice-Hall, New York.

Felsenstein, J., 1973 Maximum likelihood estimation of evolutionary trees from continuous characters. Am. J. Hum. Genet. **25:** 471–492.

Felsenstein, J., 1988 Phylogenies from molecular sequences: inference and reliability. Annu. Rev. Genet. **22:** 521–565.

Felsenstein, J., 1992a Estimating effective population size from sample sequences: a bootstrap Monte Carlo integration method. Genet. Res. **60:** 209–220.

Felsenstein, J., 1992b Estimating effective population size from samples of sequences: inefficiency of pairwise and segregating sites as compared to phylogenetic estimates. Genet. Res. **59:** 139–147.

Felsenstein, J., 1993 PHYLIP: phylogenetic inference package version 3.5c. Distributed over the Internet: http://evolution.genetics.washington.edu/phylip.html/.

Felsenstein, J., and G. A. Churchill, 1996 A hidden Markov model approach to variation among sites in rate of evolution. Mol. Biol. Evol. **13:** 93–104.

Fitch, W. M., 1971 Toward defining the course of evolution: minimum change for a specified tree topology. Syst. Zool. **20:** 406–416.

Griffiths, R., and S. Tavaré, 1994 Sampling theory for neutral alleles in a varying environment. Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci. **344:** 403–410.

Hammersley, J., and D. Handscomb, 1964 *Monte Carlo Methods.* Methuen and Co., London.

Hastings, W., 1970 Monte Carlo sampling methods using Markov chains and their applications. Biometrika **57:** 97–109.

Hudson, R., 1983 Properties of a natural allele model with intragenic recombination. Theor. Popul. Biol. **23:** 183–201.

Hudson, R., M. Slatkin and W. Maddison, 1992 Estimation of levels of gene flow from DNA sequence data. Genetics **132:** 583–589.

Hudson, R. R., 1990 Gene genealogies and the coalescent process, pp. 1–44 in *Oxford Surveys in Evolutionary Biology,* Vol. 7, edited by R. Dawkins and M. Ridley. Oxford University Press, Oxford, U.K.

Kass, R. E., B. P. Carlin, A. Gelman and R. M. Neal, 1998 Markov chain Monte Carlo in practice: a roundtable discussion. Am. Stat. **52:** 93–100.

Kingman, J., 1982a The coalescent. Stoch. Proc. Appl. **13:** 235–248.

Kingman, J., 1982b On the genealogy of large populations, pp. 27–43 in *Essays in Statistical Science,* edited by J. Gani and E. Hannan. Applied Probability Trust, London.

Kishino, H., and M. Hasegawa, 1989 Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in hominoidea. J. Mol. Evol. **29:** 170–179.

Kuhner, M., J. Yamato and J. Felsenstein, 1995 Estimating effective population size and mutation rate from sequence data using Metropolis-Hastings sampling. Genetics **140:** 1421–1430.

Kuhner, M., J. Yamato and J. Felsenstein, 1998 Maximum likelihood estimation of population growth rates based on the coalescent. Genetics **149:** 429–434.

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

Maynard Smith, J., 1970 Population size, polymorphism, and the rate of non-darwinian evolution. Am. Nat. **104:** 231–237.

Metropolis, N., A. Rosenbluth, M. Rosenbluth, A. Teller and E. Teller, 1953 Equation of state calculations by fast computing machines. J. Chem. Phys. **21:** 1087–1092.

Nath, H., and R. Griffiths, 1993 The coalescent in two colonies with symmetric migration. J. Math. Biol. **31:** 841–851.

Nath, H., and R. Griffiths, 1996 Estimation in an island model using simulation. Theor. Popul. Biol. **50:** 227–253.

Nei, M., and M. W. Feldman, 1972 Identity of genes by descent within and between populations under mutation and migration pressures. Theor. Popul. Biol. **3:** 460–465.

Notohara, M., 1990 The coalescent and the genealogical process in geographically structured population. J. Math. Biol. **29:** 59–75.

Rannala, B., and J. Hartigan, 1996 Estimating gene flow in island populations. Genet. Res. **67:** 147–158.

Sikorski, K., F. Stenger and J. Schwing, 1984 Algorithm 614: A FORTRAN subroutine for numerical integration in H_p. ACM Transact. Math. Soft. **10:** 152–160.

Slatkin, M., 1985 Rare alleles as indicators of gene flow. Evolution **39:** 53–65.

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

Slatkin, M., and R. Hudson, 1991 Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. Genetics **129:** 555–562.

Slatkin, M., and W. Maddison, 1989 A cladistic measure of gene flow inferred from the phylogenies of alleles. Genetics **123:** 603–613.

Swofford, D., G. Olsen, P. Waddell and D. Hillis, 1996 Phylogenetic inference, pp. 407–514 in *Molecular Systematics,* edited by D. Hillis, C. Moritz and B. Mable. Sinauer Associates, Sunderland, MA.

Takahata, N., 1988 The coalescent in two partially isolated diffusion populations. Genet. Res. **52:** 213–222.

Takahata, N., and M. Slatkin, 1990   Genealogy of neutral genes in two partially isolated populations. Theor. Popul. Biol. **38:** 331–350.

Tufto, J., S. Engen and K. Hindar, 1996   Inferring patterns of migration from gene frequencies under equilibrium conditions. Genetics **144:** 1911–1921.

Wakeley, J., 1998   Segregating sites in Wright's island model. Theor. Popul. Biol. **53:** 166–174.

Yang, Z., 1994   Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximative methods. J. Mol. Evol. **39:** 306–314.

Communicating editor: S. Tavaré

APPENDIX

Derivation of the importance sampling function: We want to calculate Equation 1 but would need to sum over all possible genealogies. This function can be transformed into an importance sampling function by assuming that $L(\mathcal{P})$ is an expectation and we sample from a distribution whose density is g instead of the correct density, f,

$$L(\mathcal{P}) = \text{Prob}(D|\mathcal{P}) = \sum_G \text{Prob}(G|\mathcal{P})\,\text{Prob}(D|G) \quad \text{(A1)}$$

$$= \mathbb{E}_f(h) = \sum_G hf \quad \text{(A2)}$$

$$= \sum_G h\frac{f}{g}g = \mathbb{E}_g\left(h\frac{f}{g}\right). \quad \text{(A3)}$$

Suppose that

$$g = \frac{\text{Prob}(G|\mathcal{P}_0)\,\text{Prob}(D|G)}{\sum_G \text{Prob}(G|\mathcal{P}_0)\,\text{Prob}(D|G)}, \quad \text{(A4)}$$

$$h = \text{Prob}(D|G), \quad \text{(A5)}$$

and

$$f = \frac{\text{Prob}(G|\mathcal{P})}{\sum_G \text{Prob}(G|\mathcal{P})} = \text{Prob}(G|\mathcal{P}), \quad \text{(A6)}$$

because

$$\sum_G \text{Prob}(G|\mathcal{P}) = 1; \quad \text{(A7)}$$

then we have

$$\mathbb{E}_f(h) = \sum_G \frac{\text{Prob}(G|\mathcal{P})}{\sum_G \text{Prob}(G|\mathcal{P})}\,\text{Prob}(D|G)$$

$$= \sum_G \text{Prob}(G|\mathcal{P})\,\text{Prob}(D|G) \quad \text{(A8)}$$

$$= \mathbb{E}_g\left(\frac{f}{g}h\right)$$

$$= \mathbb{E}_g\bigg[(\text{Prob}(G|\mathcal{P})$$

$$\times\,\text{Prob}(D|G)\Big/\left(\frac{\text{Prob}(G|\mathcal{P}_0)\,\text{Prob}(D|G)}{\sum_G \text{Prob}(G|\mathcal{P}_0)\,\text{Prob}(D|G)}\right)\bigg], \quad \text{(A9)}$$

so that

$$L(\mathcal{P}) = L(\mathcal{P}_0)\,\mathbb{E}_g\left(\frac{\text{Prob}(G|\mathcal{P})\,\text{Prob}(D|G)}{\text{Prob}(G|\mathcal{P}_0)\,\text{Prob}(D|G)}\right), \quad \text{(A10)}$$

where

$$L(\mathcal{P}_0) = \sum_G \text{Prob}(G|\mathcal{P}_0)\,\text{Prob}(D|G).$$

The expectation can be estimated by its average over the simulation

$$\frac{L(\mathcal{P})}{L(\mathcal{P}_0)} \cong \frac{1}{m}\sum_i^m \frac{\text{Prob}(G_i|\mathcal{P})}{\text{Prob}(G_i|\mathcal{P}_0)}. \quad \text{(A11)}$$

In Markov chain Monte Carlo approaches, the goal is to sample from the posterior and concentrate the sampling on regions with higher probabilities (Hammersley and Handscomb 1964; Chib and Greenberg 1995; Kass *et al.* 1998). In our scheme we are approximating the target function $\text{Prob}(G|\mathcal{P})\,\text{Prob}(D|G)$ with $\text{Prob}(G|\mathcal{P}_0)\,\text{Prob}(D|G)$; this may be a rather crude approximation when $\mathcal{P}_0$ is very different from $\mathcal{P}$, but after several updating chains the target and sampling distribution are very similar. When Θ and Θ_0 are very similar, our approach is nearly optimal if it is run long enough to avoid problems of autocorrelation. Sampling from the prior alone, $\text{Prob}(G|\mathcal{P})$ for example, is very inefficient as was shown by J.F. in 1988 (J. Felsenstein, unpublished data).

Data models: Our migration estimation divides naturally into two parts: calculation of $\text{Prob}(G|\mathcal{P})$ and $\text{Prob}(D|G)$. This makes it easy to implement models for different kinds of data: any data model influences only the latter calculation, which is the genealogy likelihood calculation.

For sequences we are using the model of change originated by one of us (J.F.) as implemented in PHYLIP 3.2 in 1984, described in Kishino and Hasegawa (1989) and described as F84 by Swofford *et al.* (1996). It is a variant of the Kimura two-parameter model, which allows for different transition and transversion ratios and variable base frequencies. In `migrate` the model is the same as in PHYLIP 3.5 and includes also rate variation among sites (Felsenstein and Churchill 1996), but the inclusion of rate variation between sites was not tested.

For microsatellite data, we have implemented a one-step mutation model in which the probability of making a net change of i steps in time interval u is

$$\text{Prob}(i|u) = \sum_{k=0}^{\infty} \frac{e^{-u}(u/2)^{i+2k}}{(i+k)!\,k!}. \quad \text{(A12)}$$

In addition to the stepwise mutation model a Brownian motion approximation is available; this fast approximation to the exact model is described elsewhere.

Our "infinite" allele model is approximated with a $(k+1)$-allele model. The observed alleles are $\mathcal{A} = (a_1, a_2, \ldots, a_k)$. All unobserved alleles are pooled into a_{k+1}.

$$\text{Prob}(a_i|a_i, u) = e^{-u}f_{a_i},$$

$$\text{Prob}(a_i|a_j, u) = (1 - e^{-u})f_{a_j}$$

$$f_{a_z} = \begin{cases} \dfrac{1}{k+1} & \text{if } a_z \in \mathcal{A}, \\[2ex] 1 - \dfrac{k}{k+1} & \text{if } a_z \notin \mathcal{A}. \end{cases} \quad (A13)$$

F_{ST} calculations: We use a calculation based on $F_W^{(i)}$, the homozygosity within the population i and F_B, the homozygosity between populations. For sequences, F_W and F_B can be calculated using the heterozygosity calculations described in Slatkin and Hudson (1991). The original recurrence equations F_W and F_B were developed by Maynard Smith (1970) and Maruyama (1970). We use an equation based on Nei and Feldman (1972). The exact equations are simplified by removing quadratic terms like μ^2, m^2, and μm and divisions by number of individuals in a population (*e.g.*, m/N). For two populations in equilibrium we get the equation system

$$F_W^{(1)} = \frac{1}{2N_1} + \left(1 - 2\mu - 2m_1 - \frac{1}{2N_1}\right)F_W^{(1)} + 2m_1F_B$$

$$F_W^{(2)} = \frac{1}{2N_2} + \left(1 - 2\mu - 2m_2 - \frac{1}{2N_2}\right)F_W^{(2)} + 2m_2F_B$$

$$F_B = F_B(1 - 2\mu - m_1 - m_2) + m_1 F_W^{(1)} + 2m_2F_W^{(2)},$$
$$(A14)$$

where μ is the mutation rate, m_i is the immigration rate into population *i*, and N_i is the subpopulation size. To fit these calculations into our framework we replace $4N\mu$ with Θ and m/μ with $\mathcal{M}$. We cannot solve for the four parameters we would need for an accurate comparison with our likelihood method. We can solve the equation system (A14) by assuming either that the population sizes are the same in both subpopulations (SN) or that the migration rates are symmetric (SM). With the SN approach solving for $\Theta \equiv \Theta_1 \equiv \Theta_2$, $\mathcal{M}_1$, and $\mathcal{M}_2$ we get

$$\Theta = \frac{2 - F_W^{(1)} - F_W^{(2)}}{2F_B + F_W^{(1)} + F_W^{(2)}},$$

$$\mathcal{M}_1 = \frac{2F_BF_W^{(1)} + F_W^{(1)} - F_W^{(2)} - 2F_B}{(F_W^{(1)} - F_B)(F_W^{(1)} + F_W^{(2)} - 2)},$$

$$\mathcal{M}_2 = \frac{2F_BF_W^{(2)} + F_W^{(2)} - F_W^{(1)} - 2F_B}{(F_W^{(2)} - F_B)(F_W^{(1)} + F_W^{(2)} - 2)}. \quad (A15)$$

TABLE 5

F_{ST}-based parameter estimation

	Population 1		Population 2		
	Θ	γ	Θ	γ	N
	Means				
Truth	0.0500	10.00	0.0050	1.00	—
SD	0.0138	13.80	0.0138	2.08	25 ●
	0.0116	3.56	0.0116	0.93	100 ○
SM	0.1824	72.04	0.0057	1.25	64 ●
	0.1167	46.10	0.0048	1.39	100 ○
	Standard deviations				
SN	0.0029	6.11	0.0029	0.55	25 ●
	0.0012	1.49	0.0012	0.16	100 ○
SM	0.1219	59.01	0.0009	0.21	64 ●
	0.0776	37.55	0.0006	0.19	100 ○

Simulation with unequal known parameters of 100 two-locus datasets with 25 sampled individuals in each population and 500 bp per locus. SN, the population sizes are the same in both subpopulations and migration rates are variable; SM, the migration rates are symmetrical and the population sizes are variable. Simulation runs with negative migration rates were discarded from the analysis in lines with ● signs; in lines with ○, negative values were replaced by 0.0.

The SM approach with Θ_1, Θ_2, and $\mathcal{M} \equiv \mathcal{M}_1 \equiv \mathcal{M}_2$ results in

$$\Theta_1 = \frac{(1 - F_W^{(1)})(F_W^{(1)} + F_W^{(2)} - 2F_B)}{(F_W^{(1)})^2 + F_W^{(1)}F_W^{(2)} - 2F_B^2}$$

$$\Theta_2 = \frac{(1 - F_W^{(2)})(F_W^{(1)} + F_W^{(2)} - 2F_B)}{(F_W^{(2)})^2 + F_W^{(1)}F_W^{(2)} - 2F_B^2}$$

$$\mathcal{M} = \frac{2F_B}{F_W^{(1)} + F_W^{(2)} - 2F_B}. \quad (A16)$$

Comparing (A15) and (A16) one can recognize that the estimation of $\mathcal{M}_i$ in (A15) breaks down when $F_B \geq F_W^{(i)}$. In (A16) the $\mathcal{M}$ can be reasonably estimated only if $2F_B \leq F_W^{(1)} + F_W^{(2)}$. For the comparison with our maximum-likelihood estimator (Table 3) we used SN because of our emphasis on estimation of migration rates. There are differences between the two F_{ST} methods even when the true parameters are symmetrical (data not shown), but we do not have the impression that one F_{ST} method works better than the other over the range we are using for the comparison with our `migrate` method (Table 5).