Distinguishing Migration From Isolation: A Markov Chain Monte Carlo Approach

Rasmus Nielsen and John Wakeley

Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts 02138

ABSTRACT

A Markov chain Monte Carlo method for estimating the relative effects of migration and isolation on genetic diversity in a pair of populations from DNA sequence data is developed and tested using simulations. The two populations are assumed to be descended from a panmictic ancestral population at some time in the past and may (or may not) after that be connected by migration. The use of a Markov chain Monte Carlo method allows the joint estimation of multiple demographic parameters in either a Bayesian or a likelihood framework. The parameters estimated include the migration rate for each population, the time since the two populations diverged from a common ancestral population, and the relative size of each of the two current populations and of the common ancestral population. The results show that even a single nonrecombining genetic locus can provide substantial power to test the hypothesis of no ongoing migration and/or to test models of symmetric migration between the two populations. The use of the method is illustrated in an application to mitochondrial DNA sequence data from a fish species: the threespine stickleback (*Gasterosteus aculeatus*).

T HE analysis of population subdivision has been a major focus in population genetics and molecular ecology. However, most models of population subdivision make one of two rather extreme assumptions: either (1) the populations have been exchanging migrants at a constant rate for an infinitely long time or (2) the populations are descended from a common ancestral population at some time in the past and have since diverged without gene flow. The first of these is typically referred to as equilibrium migration and the second is often called isolation or historical association. Here we consider a more general model that includes both of these as special cases.

The objective of most biological studies that assume equilibrium migration is to describe the pattern of migration among populations. Usually, some measure of population subdivision, such as F_{ST}, is calculated and used to estimate migration rates. For example, under the symmetric island model (Wright 1931), F_{ST} can easily be transformed into an estimate of the migration rate. Beerli and Felsenstein (1999) recently introduced a modified F_{ST}-based method for asymmetric equilibrium migration. Other direct methods for estimating migration also exist, notably the maximum-likelihood methods developed by Nath and Griffiths (1996), Beerli and Felsenstein (1999), and Bahlo and Griffiths (2000). When isolation without gene flow is assumed, it is of interest to estimate the divergence time between populations. In this case it is possible to transform an estimate of F_{ST} into an estimate of divergence time. In addition, there are other methods of estimating the divergence time between populations, including both maximum likelihood (Nielsen 1998; Nielsen *et al.* 1998) and moment-based approaches (Wakeley and Hey 1997).

However, there are very few methods that assist us in choosing the appropriate model, *e.g.*, to distinguish between isolation and migration as explanations for observed patterns of genetic divergence. Exceptions are the method of Wakeley (1996b), designed for nonrecombining DNA sequence data, and the method of Nielsen and Slatkin (2000), applicable to unlinked restriction fragment length polymorphism (RFLP) or single nucleotide polymorphism (SNP) data. Under somewhat restrictive assumptions about historical population sizes and other aspects of demography, these methods do allow some statistical testing of the demographic models, but with low power.

Despite the fundamental importance of assessing the relative contributions of migration and isolation, there currently exists no framework for obtaining joint estimates of divergence times and migration rates from DNA sequence data. A familiar example, in which the lack of appropriate statistical methods has been felt, is the out-of-Africa controversy in human genetics. Summaries of the data, such as F_{ST}, can be explained either by assuming short divergence times and little migration or long divergence times and strong migration. It is clear, though, that genetic data can contain substantial information regarding the relative contributions of migration and historical association. For example, Wakeley (1996a) showed that the variance of pairwise nucleotide differences is larger under equilibrium migration than

Corresponding author: Rasmus Nielsen, Department of Biometrics, 439 Warren Hall, Cornell University, Ithaca, NY 14853-7801. E-mail: rn28@cornell.edu

under isolation, even when F_{ST} is the same in both models. However, the variance of pairwise differences is a fairly drastic summary of the information in the data. As a single statistic, like F_{ST}, it will not likely be useful in distinguishing among more complicated demographic histories.

The aim of this article is to develop a framework for jointly estimating divergence times and migration rates between two populations from DNA sequence data. We consider a general model, described in the next section, in which the migration rate for each population may be different, and the ancestral population and the two descendent populations may all be of different effective sizes. We make use of both likelihood and Bayesian approaches to ancestral inference. The likelihood function and the posterior distributions are calculated using a Markov chain Monte Carlo (MCMC) method. Our results show that it is possible to obtain reliable joint estimates of migration rates and divergence times from single-locus DNA sequence data. In addition, it is possible to distinguish models of low gene flow and short divergence times from models of high gene flow and long divergence times and to assess asymmetries in the rates of migration between the two populations.

MODEL

The demographic model we consider is a model of divergence between two populations that arose from a single ancestral population at generation t in the past. In the most general version of the model we allow all three populations, the ancestor and both descendants, to have different effective population sizes. The two populations may exchange migrants, and we allow the migration rate to differ between the two. A graphical representation of this demographic model is depicted in Figure 1. There are five parameters: t, N_1, N_2, N_A, m_1, and m_2. N_1 and N_2 are the effective population sizes of the first and the second sampled populations, respectively. N_A is the population size of the ancestral population, m_1 is the proportion of population 1 that is replaced by migrants from population 2 in each generation, and m_2 is the proportion of population 2 that is replaced by migrants from population 1 each generation. We also consider a restricted three-parameter model in which $N_1 = N_2 = N_A$, and $m_1 = m_2$.

We assume selective neutrality and that there is no further population subdivision within each population. Further, we assume that the population sizes, N_1, N_2, and N_A, do not change over time. To model the mutational process we use an infinite sites model without recombination (WATTERSON 1975). However, the method could easily be extended to more general mutation models. We let μ be the rate of neutral infinite sites mutation per sequence per generation.

The coalescent process under this model follows from the now-classical theory of KINGMAN (1982a,b); see also
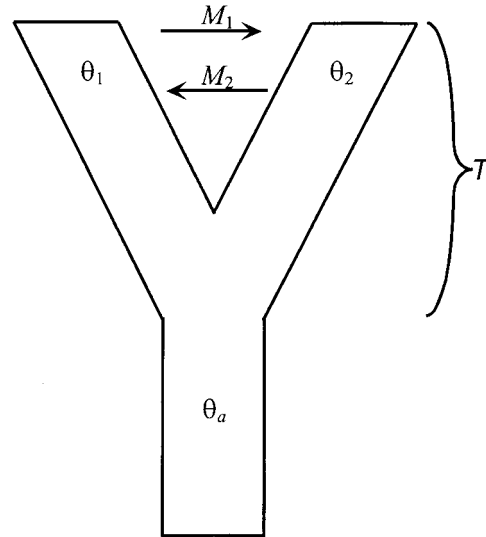

FIGURE 1.—A graphical representation of the model considered in this article.

TAVARÉ (1984) and HUDSON (1991). Here, we scale all parameters by N_1 and consider the behavior of the model as $N_1 \to \infty$ and $4N_1\mu \to \theta$, $N_2/N_1 \to r$, $N_A/N_1 \to a$, $2N_1 m_1 \to M_1$, and $2N_2 m_2 \to M_2$. A coalescence process then arises from a large class of exchangeable population genetic models including Moran models and Wright-Fisher models (WILKINSON-HERBOTS 1998). In this coalescence process, time is measured in units of $2N_1$ generations. Assume that at a given point in time there are n_1 ancestral gene copies in population 1 and n_2 ancestral gene copies in population 2. Looking back in time, coalescence events occur in population 1 and population 2 at rate $n_1(n_1 - 1)/2$ and $n_2(n_2 - 1)/(2r)$, respectively, up until the time of population divergence $T = t/2N_1$. At the same time migration events are occurring at rate $n_1 M_1$ and $n_2 M_2$ and mutations arise independently on each lineage according to a Poisson process with rate $\theta/2$. At time T, the two populations merge (looking backward in time) and a new ancestral sample is created by setting $n_A = n_1 + n_2$. Thereafter, coalescence events occur at rate $n_A(n_A - 1)/(2a)$ until only one gene copy is left in the sample. This is the stochastic process that we use to assign probabilities to different genealogies.

STATISTICS

We are interested in making inferences about the parameters $\Theta = \{\theta, M_1, M_2, r, a, T\}$ using the observed sequence data X. The first approach we use is to calculate (or approximate) the likelihood function for the parameters $L(\Theta \mid X) \propto \Pr(X \mid \Theta)$ to obtain point estimates and confidence intervals for the parameters and to perform statistical tests.

Bayesian inference: In addition to a likelihood framework, we also explore the utility of a Bayesian approach

to estimating parameters and making probabilistic statements regarding Θ. In this, we assume the existence of some prior distribution for Θ, $f(\Theta)$, and make inferences regarding Θ on the basis of the posterior distribution $f(\Theta \mid X)$. The posterior distribution is obtained from the prior distribution using

$$f(\Theta \mid X) = \Pr(X \mid \Theta) f(\Theta) / \Pr(X). \qquad (1)$$

We assume a uniform prior for M_1, M_2, T, and θ and assume that $\log(r)$ and $\log(a)$ are uniformly distributed on an interval symmetric around 0. Assuming uniform priors implies that we assign equal probability mass to all possible values of the parameter; *i.e.*, we assume that all possible values of the parameter are equally likely before observing the data. To ensure that the posterior distributions are proper, *i.e.*, that they are real probability distributions, we must constrain the parameter space. Unless otherwise stated, the parameters were constrained as follows: $M_i \in [0, 10]$, $i = 1, 2$; $T \in [0, 10]$; $\log(r) \in [-10, 10]$; $\log(a) \in [-10, 10]$. These ranges should encompass most biologically important values. In no cases discussed in this article did the maximum-likelihood estimate of a parameter equal the upper boundary of the assumed parameter range. However, if very large parameter estimates are found, the support of the priors can be expanded appropriately.

Integrated likelihood: Note that when a uniform prior distribution is used, the likelihood function $L(\Theta \mid X)$ is given by the posterior distribution. In other cases, the likelihood function can be deduced by comparing the posterior distribution with the prior distribution. In most cases we concentrate on the interpretation of the posterior distribution in a likelihood framework. We summarize the information for a single parameter in terms of the integrated likelihood for the parameter. The integrated likelihood function for a parameter is obtained by integrating over the prior distribution of all other parameters. Although integrated likelihood is a basically Bayesian idea, most of the properties of a regular likelihood function also apply to the integrated likelihood function. The review by Berger *et al.* (1999) provides a discussion of the merits and utility of integrated likelihood.

Hypotheses testing: To perform hypothesis tests and model selection we use log-likelihood-ratio tests. The standard theory stipulates that approximately minus two times the log-likelihood ratio of two nested models is asymptotically χ_j^2 distributed, where j is the difference in the number of parameters between the two models. However, this result is in general not applicable to the current problems. The first reason for this is that the parameters of interest, in several cases, are at the boundary of the parameter space, for instance, if we wish to test the hypothesis $M_1 = 0$. When the parameter is at the boundary of the parameter space, this violates the regularity conditions under which the χ^2 approximation has been derived. Under the assumption that there is

one parameter of interest, that this parameter is on the boundary of the parameter space, and that no other parameters are on the boundary, the log-likelihood-ratio test statistic should instead be asymptotically distributed as a random variable that takes the value 0 with probability 0.5 and takes on a value from a χ_1^2 distribution with probability 0.5 (Chernoff 1954).

However, there is another important reason not to use the large sample approximations in this case: the data from individual nucleotide sites are all correlated because of shared common ancestry. The likelihood function cannot be written as the product of the likelihood in multiple independent data points. In this sense, there is only one independent data point and it is not appropriate to appeal to the standard asymptotic results.

To perform valid likelihood-ratio tests we therefore explore the use of parametric bootstrapping. Parametric bootstrapping proceeds by approximating the null distribution of the likelihood-ratio test statistic by the distribution of test statistics obtained from samples simulated under the null hypothesis. The simulations are performed by assuming that the true value of the nuisance parameters equals the value estimated under the null hypothesis. In the present case, such simulations can be performed using standard coalescence simulation methods. Unfortunately, in many applications it may be too time consuming to perform parametric bootstrapping. Therefore, we also explore another approach for model selection, which is based on the Akaike information criterion (AIC; *e.g.*, Akaike 1985). According to the AIC, the best model is the model that minimizes $-2 \times [\log(L) - d_i]$, where d_i is the number of parameters of model i.

MARKOV CHAIN MONTE CARLO

The simulation method we use to approximate the likelihood function/posterior distribution of the parameters is similar to the methods applied by Wilson and Balding (1998) and Nielsen (2000a), which are related to the methods by Kuhner *et al.* (1995) and Beerli and Felsenstein (1999). These use the Metropolis-Hastings method (Metropolis *et al.* 1953; Hastings 1970) to integrate over genealogies stochastically. We wish to approximate the posterior distribution of the parameters $f(\Theta \mid X) = cf(X \mid \Theta) f(\Theta)$, where c is a constant proportionality factor. It is not possible in general to calculate $f(X \mid \Theta)$, except by conditioning on the underlying gene genealogy, represented by G. The posterior distribution can then be written as

$$f(\Theta \mid X) = cf(\Theta) \int_{G \in \Gamma} f(X \mid \Theta, G) f(G) \, dG, \qquad (2)$$

and the problem of estimating the posterior distribution/likelihood function is reduced to the problem of solving the integral in (2), where Γ refers to the set of all possible gene genealogies.

Under the infinite sites model $f(X \mid \Theta, G)$ can easily be

calculated by mapping mutations on the gene genealogy (G). The number of mutations on an edge is Poisson distributed with rate $\theta t/2$, where t is the length of the edge. The total data likelihood conditional on the gene genealogy can, therefore, be evaluated as the product of multiple independent Poisson random variables.

Metropolis-Hastings MCMC: To evaluate the integral in (2) it is necessary to use Monte Carlo methods. As in WILSON and BALDING (1998) and NIELSEN (2000a), it is possible to define a Markov chain with state space given by Γ and with stationary distribution of genealogies proportional to $f(X \mid \Theta, G)f(G)$. By sampling from this chain at stationarity it is possible to approximate the posterior distribution of Θ. To assure that the chain has the desired stationary distribution, we use the Metropolis-Hastings method (METROPOLIS *et al.* 1953; HASTINGS 1970). At each step in the chain, updates from the current state (Θ_i, G_i) to a new state (Θ^*, G^*) are proposed with probability $q[(\Theta_i, G_i) \rightarrow (\Theta^*, G^*)]$. With probability

$$\min\left\{1, \frac{f(X \mid \Theta^*, G^*)f(G^*, \Theta^*)q[(\Theta^*, G^*)\rightarrow(\Theta_i, G_i)]}{f(X \mid \Theta_i, G_i)f(G_i, \Theta_i)q[\Theta_i, G_i)\rightarrow(\Theta^*, G^*)]}\right\} \quad (3)$$

the new proposed state is accepted; *i.e.*, (Θ_{i+1}, G_{i+1}) = (Θ^*, G^*), otherwise (Θ_{i+1}, G_{i+1}) = (Θ_i, G_i). If the chain is constructed such that it is aperiodic and irreducible, it will under quite general conditions have $f(\Theta \mid X)$ as its stationary distribution (*e.g.*, RIPLEY 1987). The Markov chain is constructed by suggesting updates to the gene genealogy and to the parameters one by one. The chain therefore consists of a mixture of the six type of updates for Θ and updates of the genealogy, including the times of migration. The algorithmic details of these six update types are described in the APPENDIX.

Smoothing function: To obtain smooth likelihood surfaces from a single run of the Markov chain, the method of NIELSEN (2000b) was used. In brief, the likelihood function is estimated by simulating a Markov chain with stationary distribution of G and Θ_0 proportional to $\Pr(X \mid G)P_{\Theta_0}(G)P(\Theta_0)$, and estimates of the likelihood function for a particular value of Θ are then obtained as

$$L(X \mid \Theta) \approx c \frac{\sum_{i=1}^n r_i(\Theta, \Theta_0^{(i)})w_i(\Theta, \Theta_0^{(i)}, G_i)}{n\int r(\Theta, \Theta_0)\,dP(\Theta_0)}, \quad w_i(\Theta, \Theta_0, G)$$

$$= \frac{P(\Theta)}{P(\Theta_0)}, \quad (4)$$

where c is a constant of proportionality and $r(\Theta, \Theta_0)$ is a weighting function. $r(\Theta, \Theta_0)$ is used to down-weight values of $w(\Theta, \Theta_0, G)$ for which $|\Theta - \Theta_0|$ is large. The weighting function $r(\Theta, \Theta_0)$ may take any functional form and is here defined as $r(\Theta, \Theta_0) = e^{-(|\Theta_0-\Theta|)^2/c}$, with $c = 0.01$. This function was chosen because it was found in preliminary investigations to lead to fast convergence. This method can be thought of as a weighted importance sampling scheme where the importance sampling function is partially determined by the data.
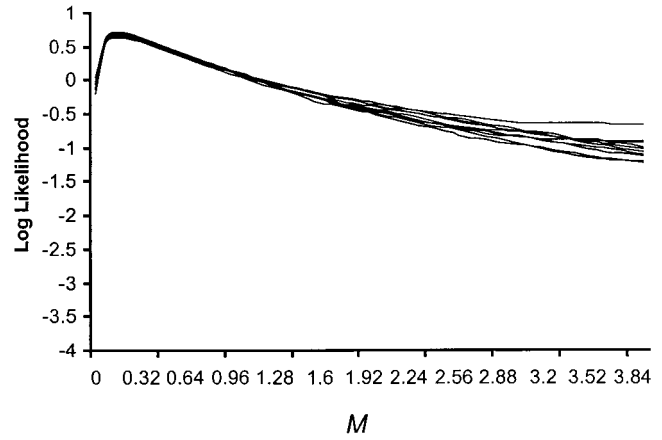

FIGURE 2.—Ten independent replicates of the likelihood surfaces for M obtained for the data set from ORTI *et al.* (1994) containing 23 sequences and 35 segregating sites.

CONVERGENCE

In Markov chain Monte Carlo methods it is typically easy to show that the Markov chain will converge and that a consistent estimate of the likelihood function can be obtained. However, it is usually much more difficult to show how many simulated steps of the chain are needed to ensure convergence of the ergodic averages. If too few steps are simulated, the estimate of the likelihood function will be biased and will depend on the initial values of the parameters. The first issue we explore is, therefore, the degree to which assumption of stationarity and convergence to the ergodic averages are satisfied in the present case for realistic simulation times. We consider a simple version of the model with 3 parameters: θ, M, and T; that is, with $a = 1$, $r = 1$, and $M_1 = M_2$. To evaluate the method we use the data set previously published by ORTI *et al.* (1994) containing 23 sequences and 35 segregating sites. The integrated likelihood function of M obtained using 10 replicate runs of the Markov chain is plotted in Figure 2. The surfaces were obtained using 1,000,000 steps in the chain and a "burn-in" time of 100,000 steps. Only samples of the chain taken after the burn-in time are used for estimating the posterior distribution/likelihood surface. The starting values of the parameters and the genealogy differed between runs. The high degree of similarity of these distributions suggests that stationarity and convergence to the ergodic averages has been achieved. Thus, 1,000,000 steps in the chain appear sufficient for data sets of this size to provide reliable likelihood surfaces/posterior distributions.

PROPERTIES OF THE ESTIMATORS

We also explored the statistical properties of the point estimators of M and T under the three-parameter model. One hundred data sets were simulated under this model for each of two parameter settings: ($M = 0$, $T = 2$, $\theta = 10$) and ($M = 1$, $T = \infty$, $\theta = 10$). In

these simulations it was assumed that 30 individuals were sampled from each population. A burn-in time of 500,000 steps was chosen and the chains were run for an additional 5,000,000 steps to evaluate the likelihood surfaces/posterior distributions. Some sample posterior distributions of T and M are shown in Figure 3. The mode of the integrated likelihood function of each parameter is used as an estimator of the parameter. In the 100 simulations the mean estimate of M was 0.0055 for the case ($M = 0$, $T = 2$, $\theta = 10$) and 1.02 for the case ($M = 1$, $T = \infty$, $\theta = 10$). The estimate of M will of course always be slightly upwardly biased when the true parameter is at the boundary of the parameter space ($M = 0$). The distribution of the estimates of M is shown in Figure 4. Note that the variance of the estimates obtained in the model ($M = 1$, $T = \infty$, $\theta = 10$) is quite high. Nonetheless, it is encouraging that reasonable estimates can be obtained in such cases by using only a single locus without recombination, even when taking the possibility of recent common ancestry of the populations into account.

The estimator of T does not appear to have similarly desirable properties, at least not in the case of $T = \infty$. There are two reasons for this. First, the Monte Carlo variance for the parameter T seems to be quite large in many cases (compare Figure 3c to Figure 3d). Since the likelihood surface often is very flat for this parameter, the estimates may not be reliable. The second reason is that the integrated likelihood surface often has multiple peaks, *e.g.*, Figure 3c. This is in fact a real property of the likelihood function, rather than an artifact of the Monte Carlo variance. The multimodality can easily be explained by considering the structure of the underlying gene genealogies and is a consequence of having only a limited number of migration events occurring in the ancestry of the sample. Consider the hypothetical case where the times of migration events in the genealogy are known and fixed. Assuming low migration, the likelihood will then always be higher when T is slightly smaller than the age of the migration event than if T is slightly larger. The likelihood surface for T may therefore increase as T approaches the time of a migration event and decrease at the time right after a migration event. Obviously, the times of migration events are not known and fixed in real data, but may be quite well determined if there are sufficient nucleotide data, causing the integrated likelihood surface for T to have multiple modes.

In contrast, for the case of ($M = 0$, $T = 2$, $\theta = 10$), the estimator seems to perform quite well with an average estimate of $\hat{T} = 2.16$ and multimodal likelihood surfaces appear to be rare.

A traditional Bayesian estimator based on the posterior expectation of the parameter was not used because it in many cases was very sensitive to the choice of upper bound for the parameter. This was expected for T, since the integrated likelihood function in some cases was an increasing function of T. However, even for many of the
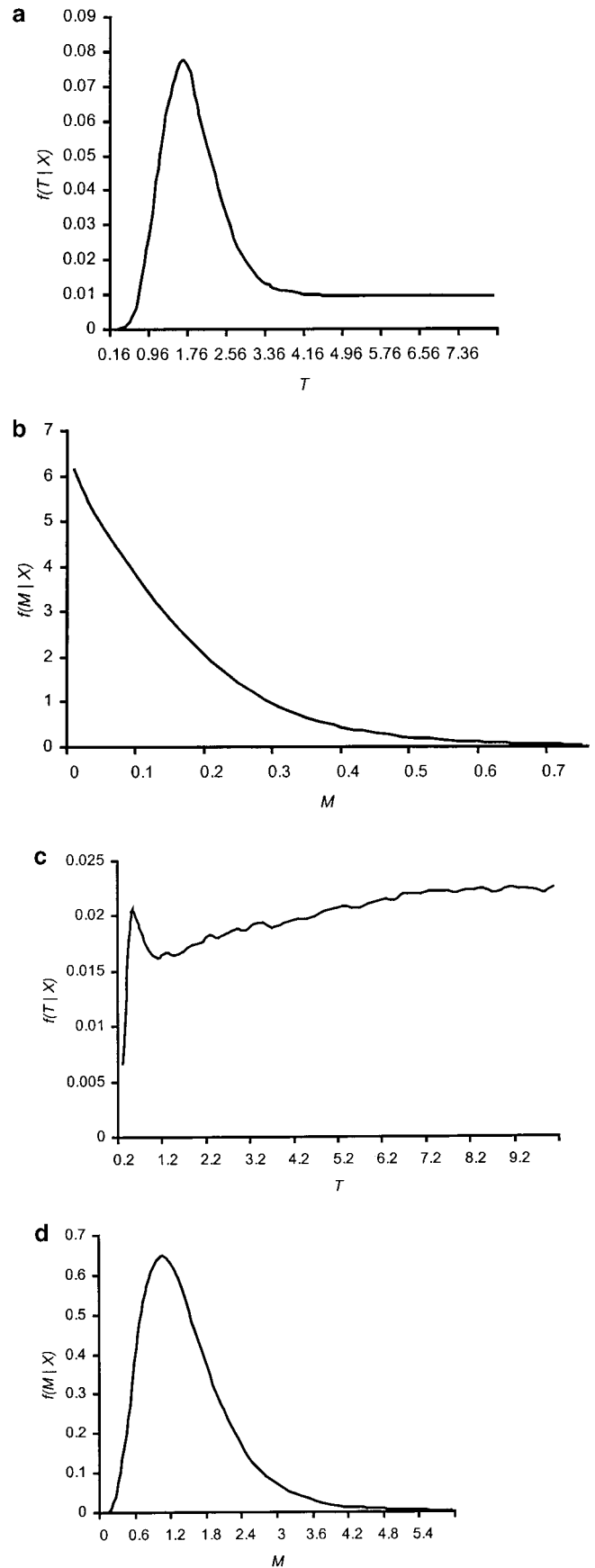

FIGURE 3.—Sample posterior distributions of M and T estimated from data simulated assuming $M = 0$, $T = 2$, and $\theta = 10$ (a and b) and $M = 1$, $T = \infty$, and $\theta = 10$ (c and d).
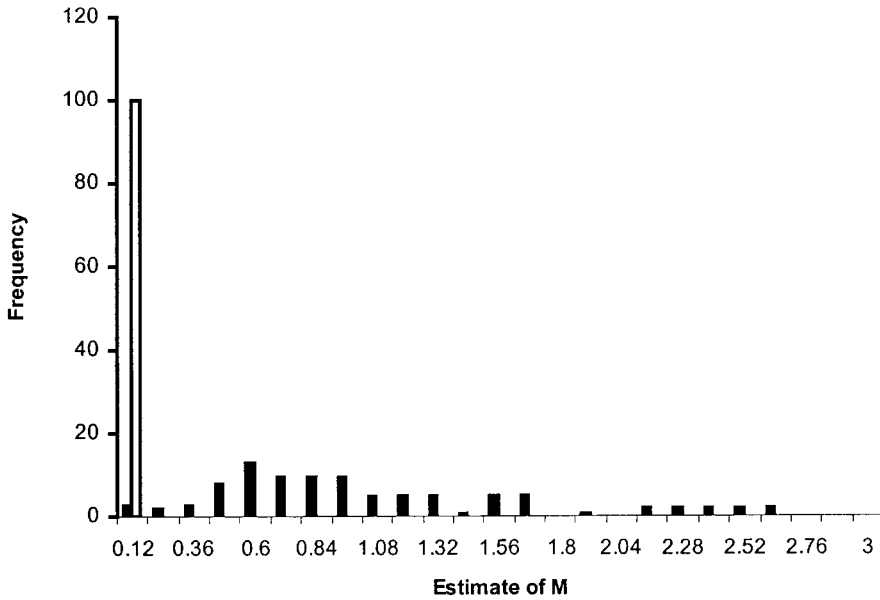
Figure 4.—The distribution of estimates of M assuming $M = 0$, $T = 2$, and $\theta = 10$ (open bar) and $M = 1$, $T = \infty$, and $\theta = 10$ (solid bars) in 100 replicate simulations.

other parameters, the posterior expectation was quite sensitive to the choice of prior. Considering data from multiple loci could reduce this problem.

DETECTING MIGRATION

The next issue we focus on is the possibility of detecting migration when it occurs. Can we reject a model of isolation without gene flow in favor of a model that includes migration? To address this issue we examine the log-likelihood ratio $\log[\max\{L(M = 0 \mid X)\}/\max\{L(M \mid X)\}]$. A plot of the likelihood ratio in the 100 simulations for $(M = 0, T = 2, \theta = 10)$ (open bars) and $(M = 1, T = \infty, \theta = 10)$ (solid bars) is shown in Figure 5. Note first that, as expected, minus two times the log-likelihood ratio is far from being χ_1^2 distributed when the null hypothesis is true (open bars). The deviation is larger than expected for a test in which the

parameter is on the boundary of the parameter space (Chernoff 1954), emphasizing that the usual asymptotics should not be applied for a population sample of a single locus. In the present case, as was also found in Nielsen and Palsboll (1999), application of the standard χ^2 approximation results in a conservative test.

However, note the difference in the distribution of the likelihood ratio in the case of $(M = 0, T = 2, \theta = 10)$ and $(M = 1, T = \infty, \theta = 10)$. This suggests that the method indeed has a lot of power to distinguish between the two models. In fact, using parametric bootstrapping, the hypothesis of $M = 0$ would be rejected in 100 of 100 cases when $M = 1$, $T = \infty$, and $\theta = 10$. Using the AIC, a model with migration would be accepted in 98 out of 100 cases when $M = 1$ and in 0 of 100 cases when $M = 0$. In this case, the AIC provides a very powerful method of distinguishing between models with and without migration.
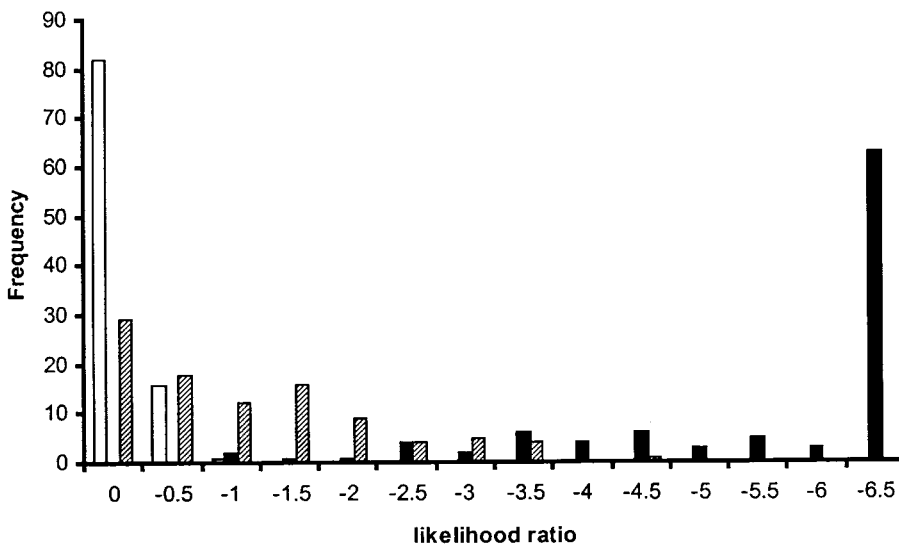

Figure 5.—The distribution of the log-likelihood ratio of the hypothesis $M = 0$ vs. $M \geq 0.0$. One hundred replicate simulations were performed for each of the parameters settings $M = 0$ and $T = 2$ (open bars), $M = 1$ and $T = \infty$ (solid bars), and $M = 0.25$ and $T = 1$ (hatched bars). In all cases $\theta = 10$.
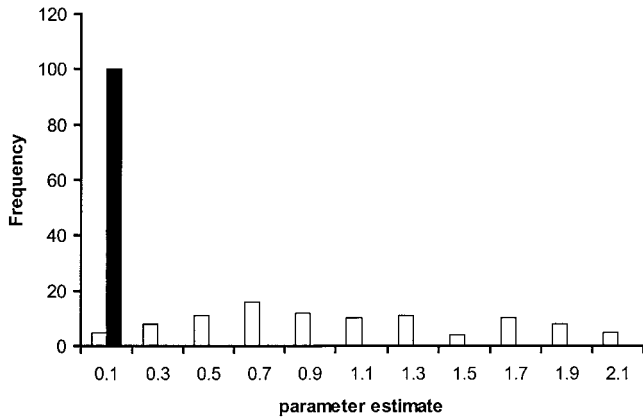
FIGURE 6.—The distribution of estimates of M_1 (open bars) and M_2 (solid bar) in 100 data sets simulated assuming $\theta = 10.0$, $M_1 = 1.0$, $M_2 = 0.0$, $r = 1.0$, $a = 1.0$, and $T = \infty$.

FIGURE 7.—The joint integrated likelihood surface for T and M estimated from the data by ORTI *et al.* (1994). Darker values indicate higher likelihood.

To investigate the power to distinguish models when there is just a small amount of migration, simulations were performed assuming $M = 0.25$, $T = 1$, and $\theta = 10$. Note (Figure 5, hatched bars) that the likelihood ratios are somewhat intermediate in this case. Using the AIC, the model without migration would be rejected in 39 out of 100 cases and accepted in 61 of 100 cases. Using parametric bootstrapping the model of no migration would have been rejected in 65 out of 100 cases. This suggests that likelihood-ratio tests with reasonable power can in fact be established. When computationally possible, it is recommended to obtain critical values for hypothesis testing using the parametric bootstrap. For the present parameter settings, 100 simulations for a single data set of this size would take ∼1 week on a medium fast workstation.

DETECTING UNEQUAL MIGRATION

Another possible application of the method is to detect unequal, or asymmetric, migration, *i.e.*, a higher rate of migration from one population to the other than in the opposite direction. To investigate the properties of the method in this regard, an equilibrium migration model with parameters $\theta = 10.0$, $M_1 = 1.0$, $M_2 = 0.0$, $r = 1.0$, $a = 1.0$, and $T = \infty$ was simulated. One hundred simulations were performed, each using 10,000,000 steps in the Markov chain to estimate the posterior distribution/likelihood surface. Estimates of M_1 and M_2 were obtained using the mode of the posterior distribution of these parameters and are shown in Figure 6. Estimates of M_2 were all in the range 0.0–0.2 but only five estimates of M_1 fell in this range.

Using the AIC, a model with $M_1 = 0$ would be rejected in 67 of 100 cases and $M_2 = 0$ would be rejected in 0 out of 100 cases. Parametric bootstrapping to test models was not performed in this case. Performing such a test would take a few weeks on a medium fast workstation. In
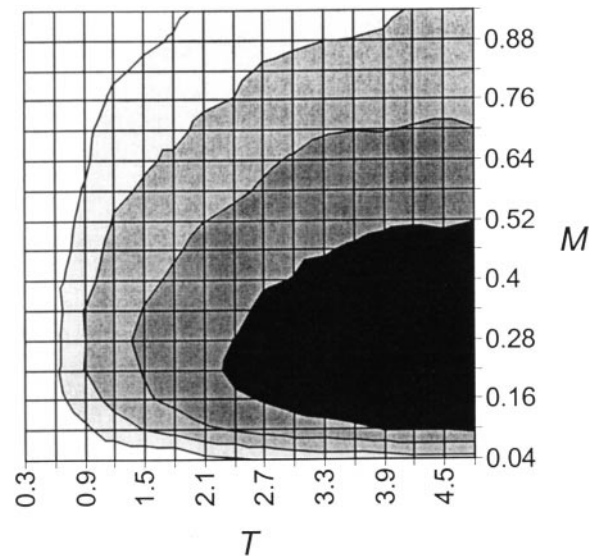
any case, it is obvious from these simulations that it is in fact possible to detect unequal migration and that a single nonrecombining locus might be sufficient in most cases for this purpose as long as θ is sufficiently large. If necessary, multiple independent loci could be used to further increase the power.

APPLICATIONS TO THE DATA BY ORTI *et al.* (1994)

A previous method for distinguishing migration from isolation on the basis of variance of pairwise differences was established by WAKELEY (1996b). That method was applied to a data set by ORTI *et al.* (1994) of mtDNA sequences from the western and eastern Pacific stickleback populations, and the hypothesis of pure isolation ($M = 0$) was rejected with a P value of 0.013. To compare the methods and to provide an illustration of the new method on the basis of real data, we reanalyzed the data set by ORTI *et al.* (1994) using our new method. The integrated likelihood function for T and M is shown in Figure 7. Thus, our analysis agrees well with that of WAKELEY (1996b). The data seem to be most compatible with a model of large divergence times with rates of migration of ∼$M = 0.3$. A model of moderate migration and very long divergence times is more compatible with the data than a model of short divergence times and low migration rates.

Further, using our method we can also extract information regarding the direction of gene flow between the populations. The integrated likelihood surfaces for M_1 and M_2 are shown in Figure 8. The data appear to be compatible with a model of unequal gene flow between the populations. A strictly decreasing likelihood surface for M_2 indicates low levels of gene flow from
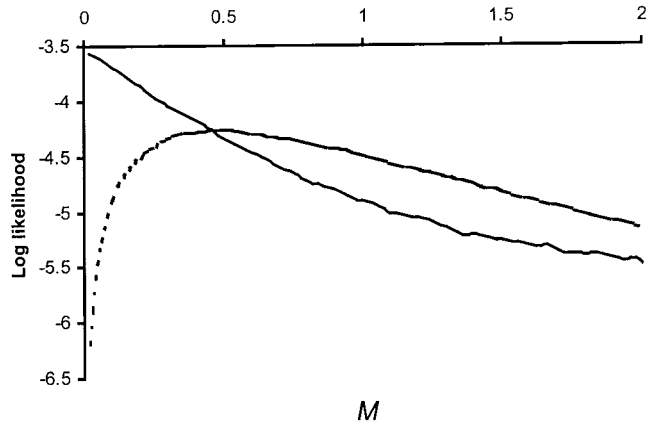
FIGURE 8.—The integrated likelihood surfaces for M_1 (dots) and M_2 (solid lines) estimated from the data by ORTI *et al.* 1994.

FIGURE 10.—The integrated likelihood surface for a estimated from the data by ORTI *et al.* (1994).

the eastern Pacific to the western Pacific populations. In contrast, the estimate of M_1 based on the integrated likelihood function is ∼0.5 and the likelihood ratio of the hypothesis $M_1 = 0$ is >2.0. There seems to be ongoing gene flow from the western to the eastern Pacific populations but little gene flow in the opposite direction.

Another question we are able to address is the difference in effective population size between the two populations. If a model without gene flow were applied to these stickleback populations, the migration from western to eastern Pacific could inflate the estimates of effective population size of the recipient population. For example, based on the average number of pairwise differences within populations, the estimate of the effective population size of the eastern Pacific population is 4.3 times the size of the western Pacific population. When migration is taken into account, the estimate based on the mode of the integrated likelihood function is that
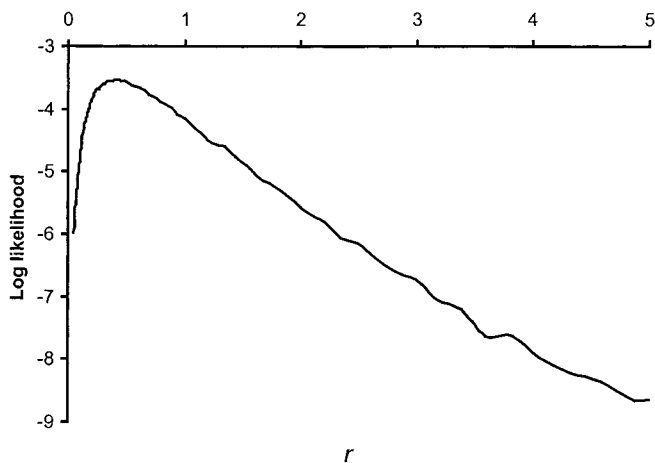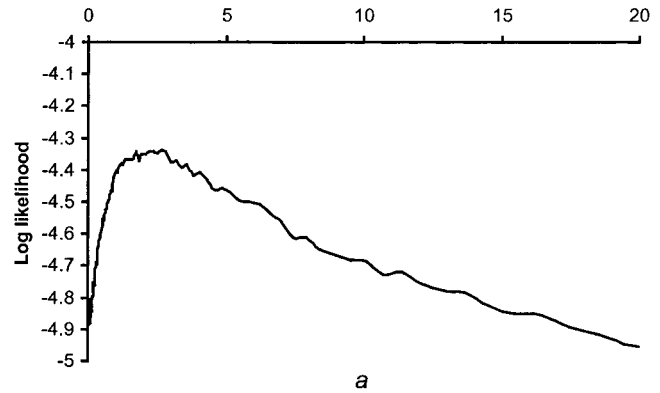
the eastern Pacific population is 3.0 times larger than the size of the western Pacific population (Figure 9).

We can also consider another parameter of interest: the size of the ancestral population relative to the current size of one of the populations. The integrated likelihood for this parameter (a) is shown in Figure 10. Note that the likelihood surface is very flat. The reason is that the data are compatible with a model of long divergence times, in which case there is only little information regarding a preserved. The MLE of a is ∼1.7.

DISCUSSION

We demonstrated that estimates of the parameters θ, M_1, M_2, r, a, and T can be obtained within either a Bayesian or likelihood framework using an MCMC method. In most cases, reliable parameter estimates can be obtained in a few hours on a desktop computer. We also found substantial power to test competing demographic hypotheses using a likelihood-ratio test. For example, using parametric bootstrapping, a model of no migration is rejected at the 5% level in 100/100 cases when the true model is equilibrium migration with $M = 1.0$ and $\theta = 10.0$. This is a dramatic improvement over the methods of WAKELEY (1996b) and NIELSEN and SLATKIN (2000). For example, WAKELEY's (1996b) method would have ∼30% power under similar circumstances.

The simulations presented here support the use of the AIC (*e.g.*, AKAIKE 1985) as a means of identifying plausible demographic models in ecological genetic studies. It is possible to address the classic problem of distinguishing between short divergence times with low gene flow and long divergence times with moderate gene flow as explanations of population divergence. In addition, this method can be used to assess the importance of asymmetries in migration rates between populations. A well-known case where these tools could be of use is in studies of the emergence of modern humans. On the one hand, the out-of-Africa hypothesis (*e.g.*,

FIGURE 9.—The integrated likelihood surface for r estimated from the data by ORTI *et al.* (1994).

STRINGER and ANDREWS 1988; VIGILANT *et al.* 1991) assumes low migration and recent divergence among human populations, and on the other, the "multiregional" model (*e.g.*, BROOKS and WOOD 1990; WOLPOFF 1996) dictates long divergence times between human populations. With appropriate modifications, the method presented here could be used to resolve this and other disputes.

Our method has several important limitations. Most importantly, it assumes an infinite sites model of mutation. This restriction may be prohibitive in many applications. However, the method can be extended to the case of the finite sites model, although this may significantly increase the required computational time. In general, computational time is a limitation in these methods. Most reasonably sized data sets can be analyzed using the present methods, but there may be data sets where the computational time is prohibitive, especially when applied to finite sites models. Also, since the method is based on MCMC, it is very important to assess the convergence in each particular study. We recommend running multiple chains using different starting values for the parameters. If all chains give similar results, they provide good evidence that convergence of the ergodic averages was achieved.

For the purpose of hypothesis testing, we recommend using parametric bootstrapping. Use of the classical χ^2 approximation, possibly with the correction by CHERNOFF (1954), would lead to conservative tests with reduced power in the examples we investigated here. Tests based on parametric bootstrapping are computationally intensive and a heuristic application of the apparently conservative χ^2 criterion or of the AIC criterion may therefore in some cases be preferable.

We did not report the results based on Bayesian tests. From a practical standpoint, such tests were not desirable because the parameter estimates were quite sensitive to the choice of prior. Also, using improper uniform prior distributions on the interval $[0, \infty)$ results in improper posterior distributions for most parameters.

Computer programs implementing the methods are available on request from Rasmus Nielsen (rn28@cornell.edu).

LITERATURE CITED

AKAIKE, H., 1985 Prediction and entropy, pp. 1–24 in *A Celebration of Statistics*, edited by A. C. ATKINSON and S. E. FIENBERG. Springer-Verlag, Berlin.

BAHLO, M., and R. C. GRIFFITHS, 2000 Inference from gene trees in a subdivided population. Theor. Popul. Biol. **57:** 79–95.

BEERLI, P., and J. FELSENSTEIN, 1999 Maximum-likelihood estimation of migration rates and effective population numbers in two populations using a coalescent approach. Genetics **152:** 763–773.

BERGER, J. O., B. LISEO and R. L. WOLPERT, 1999 Integrated likelihood methods for eliminating nuisance parameters. Stat. Sci. **14:** 1–28.

BROOKS, A. S., and B. WOOD, 1990 Paleoanthropology: the Chinese side of the story. Nature **344:** 288–289.

CHERNOFF, H., 1954 On the distribution of the likelihood ratio. Ann. Math. Stat. **25:** 573–578.

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

HUDSON, R. R., 1991 Gene genealogies and the coalescent process, pp. 1–44 in *Oxford Surveys of Evolutionary Biology*, edited by D. FUTUYMA and J. ANTONOVICS. Oxford University Press, Oxford.

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

KINGMAN, J. F. C., 1982b On the genealogy of large populations. J. Appl. Prob. **19:** 27–43.

KUHNER, M. K., J. YAMATO and F. FELSENSTEIN, 1995 Estimating effective population size and mutation rate from sequence data using Metropolis-Hastings sampling. Genetics **140:** 1421–1430.

METROPOLIS, N., A. W. ROSENBLUTH, M. N. ROSENBLUTH, A. H. TELLER and E. TELLER, 1953 Equations of state calculations by fast computing machines. J. Chem. Phys. **21:** 1087–1091.

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

NIELSEN, R., 1998 Maximum likelihood estimation of population divergence times and population phylogenies under the infinite sites model. Theor. Popul. Biol. **53:** 143–151.

NIELSEN, R., 2000a Estimation of population parameters and recombination rates using single nucleotide polymorphisms. Genetics **154:** 931–942.

NIELSEN, R., 2000b Improved MCMC methods for some population genetic models. Technical report no. M-1556. Department of Biometrics, Cornell University, Ithaca, NY.

NIELSEN, R., and P. J. PALSBOLL, 1999 Tests of microsatellite evolution: multi-step mutations and constraints on allele size. Mol. Phylogenet. Evol. **11:** 477–484.

NIELSEN, R., and M. SLATKIN, 2000 Analysis of population subdivision using di-allelic models. Evolution **54:** 44–50.

NIELSEN, R., J. L. MOUNTAIN, J. P. HUELSENBECK and M. SLATKIN, 1998 Maximum-likelihood estimation of population divergence times and population phylogeny in models without mutation. Evolution **52:** 669–677.

ORTI, G., M. A. BELL, T. E. REIMCHEN and A. MEYER, 1994 Global survey of mitochondrial DNA sequences in the threespine stickleback: evidence for recent migrations. Evolution **48:** 608–622.

RIPLEY, B., 1987 *Stochastic Simulation.* Wiley, New York.

STRINGER C. B., and P. ANDREWS, 1988 Genetic and fossil evidence for the origin of modern humans. Science **239:** 1263–1268.

TAVARÉ, S., 1984 Lines of descent and genealogical processes, and their application in population genetics models. Theor. Popul. Biol. **26:** 119–164.

VIGILANT, L., M. STONEKING, H. HARPENDING, K. HAWKES and A. C. WILSON, 1991 African populations and the evolution of human mitochondrial DNA. Science **253:** 1503–1507.

WAKELEY, J., 1996a The variance of pairwise nucleotide differences in two populations with migration. Theor. Popul. Biol. **49:** 39–57.

WAKELEY, J., 1996b Distinguishing migration from isolation using the variance of pairwise differences. Theor. Popul. Biol. **49:** 369–386.

WAKELEY, J., and J. HEY, 1997 Estimating ancestral population parameters. Genetics **145:** 847–855.

WATTERSON, G. A., 1975 On the number of segregating sites in genetical models without recombination. Theor. Popul. Biol. **7:** 256–276.

WILKINSON-HERBOTS, H. M., 1998 Genealogy and subpopulation differentiation under various models of population structure. J. Math. Biol. **37:** 535–585.

WILSON, I. J., and D. J. BALDING, 1998 Genealogical inference from microsatellite data. Genetics **150:** 499–510.

WOLPOFF, M. H., 1996 Interpretations of multiregional evolution. Science **274:** 704–707.

WRIGHT, S., 1931 Evolution in Mendelian populations. Genetics **16:** 97–159.

APPENDIX

In the following we describe the details of the different types of updates of the Markov chain. In this description, we consider the times of migration as a part of the description of a genealogy (G). G therefore consists of a topology, a vector of coalescence times, and a vector of migration times for each edge in G. The algorithm is similar to the proposal algorithm from Beerli and Felsenstein (1999).

Updates of G: Updates of the gene genealogy are performed by simulating the ancestry of a lineage from the coalescence prior conditional on the remaining part of the genealogy. The simulations can be described by the following algorithm:

1. Choose a random edge in the tree uniformly among all edges.
2. Detach the edge from the tree at the point in which the edge coalesces with some other "sister-edge" in the genealogy.
3. Start simulations of the new path of the edge at the time of the origin of the edge, *i.e.*, at the time where two sister-edges coalesced into the focal edge or at time 0 if the edge existed at this point in time.
4. Stop the simulations when the edge coalesces with a new sister-edge.

The simulations of the path of the edge are performed relatively easily using the usual coalescence simulation methods. For example, if the edge at a given time exists in population 1 and there are n_1 ancestral gene copies existing in population 1, the edge migrates into population 2 at rate M_1 and coalesces with each lineage in population 1 at rate 1, when time is measured in units of $2N_1$. It is therefore possible to simulate the ancestry of the lineage conditional on the remaining gene genealogy simply by simulating a series of exponential random variables. However, such simulations are somewhat more complicated than usual coalescence simulations because at each step it is necessary to keep track of the identity of all edges existing in the population at a particular time. This means that the gene genealogy is not only represented in terms of the genealogical structure itself but also in terms of the times of migrations of edges between the populations. Also, some minor technical difficulties must be overcome when the new history of the edge changes the time of the root (the time to the most recent common ancestor of the sample).

Since the path of the gene genealogy is simulated according to the prior distribution, the acceptance probability for this type of update is simply min{1, $f(X \mid \Theta_i, G^*)/f(X \mid \Theta_i, G_i)$}. The conditional likelihood $f(X \mid \Theta, G)$ can be calculated for any Θ and G under the infinite sites model by mapping mutations on the genealogy (there is only one unique way of doing this under the infinite sites model).

Although it may be possible to establish more efficient types of update under the infinite sites model, this type of update was chosen because it can also be applied when the mutational model is more complex.

Updates of M: M_1 and M_2 are defined either on the interval $[0, \infty)$ or on the interval $[0, M_{max}]$. The following description assumes that $M_i \in [0, M_{max}]$. The update algorithm for the case of $M_i \in [0, \infty)$ follows easily from this description. Let the current value of M_j be M_{ji}; a proposal value of M_j (M_j^*) is then drawn uniformly from the interval $[M_{ji} - \Delta_M, M_{ji} + \Delta_M]$, where Δ_M is chosen such that $\Delta_M \leq M_{max}$. If $M_j^* < 0$, set $M_j^* = -M_j^*$, and if $M_j^* > M_{max}$, set $M_j^* = 2M_{ji} - M_j^*$. Using this reflection of M_j^* around 0 and M_{max} it is guaranteed that $q[(\Theta^*, G_i) \rightarrow (\Theta_i, G_i)] = q[(\Theta_i, G_i) \rightarrow (\Theta^*, G_i)]$, where $\Theta^* = \{M_j^*\} \cup \Theta_i \backslash \{M_{ji}\}$. Also, since the structure of the genealogy does not change, the acceptance probability does not depend on the data and is just given by min{1, $f(G_i \mid \Theta^*)/f(G_i \mid \Theta_i)$}, which can be calculated quite fast. Let τ_k be the time between the $k - 1$th and the kth coalescence or migration event and define $\psi_k = \Sigma_{i=1}^k \tau_k$. If $\psi_k > T$ and $\psi_{k-1} < T$, define $\tau_k = T - \psi_{k-1}$. Also denote the number of active ancestral lineages in population j in the time interval between the $k - 1$th and the kth coalescence or migration event by n_{jk}. Then the ratio $f(G_i \mid \Theta^*)/f(G_i \mid \Theta_i)$ is given by

$$\prod_{k: \Psi_{k-1} < T} h_M(\tau_k, n_{1k}, n_{1k}),$$

$$h_M(\tau_k, n_{1k}, n_{1k}) = \begin{cases} \dfrac{M_j^*}{M_{ij}} e^{\tau_j(M_{ij} - M_j^*) n_{jk}} \\ \quad \text{if interval } k \text{ ends in a migration event} \\ \quad \quad \text{from population } j \text{ to the other population} \\ e^{\tau_j(M_{ij} - M_j^*) n_{jk}} \\ \quad \text{else.} \end{cases}$$

$$(A1)$$

Updates of r: Updates of r are chosen similarly to updates of M_j. It is assumed that $\log(r)$ is distributed uniformly on the interval $[\log(1/r_{max}), \log(r_{max})]$. A new value of r (r^*) is proposed by choosing $\log(r^*)$ uniformly in the interval $[r_i - \Delta_r, r_i + \Delta_r]$, where Δ_r is chosen such that $\Delta_r \leq r_{max}$. If $r^* < 1/r_{max}$, set $\log(r^*) = 2 \log(1/r_{max}) - \log(r^*)$ and if $r^* < r_{max}$, set $\log(r^*) = 2 \log(r_{max}) - \log(r^*)$. The acceptance ratio of this type of update is again just given by the Metropolis factor

$$\dfrac{f(G_i \mid \Theta^*)}{f(G_i \mid \Theta_i)} = \prod_{k: \Psi_{k-1} < T} h_r(\tau_k, n_{2k}),$$

$$h_r(\tau_k, n_{2k}) = \begin{cases} \dfrac{r_i}{r^*} e^{\tau_k \left(M_2 n_{2k} + \binom{n_{2k}}{2} \right)(r_i^{-1} - r^{*-1})} \\ \quad \text{if interval } k \text{ ends with a migration event} \\ \quad \quad \text{from or coalescent event in population 2} \\ e^{\tau_k \left(M_2 n_{2k} + \binom{n_{2k}}{2} \right)(r_i^{-1} - r^{*-1})} \\ \quad \text{else,} \end{cases}$$

$$(A2)$$

where now $\Theta^* = \{r^*\} \cup \Theta_i \backslash \{r_i\}$.

Updates of a: Updates of a are generated in a manner identical to the method used for generating updates to

r; *i.e.*, a new value of a (a^*) is chosen in an interval of size $2\Delta_a$ around the current value of a (a_i). The acceptance ratio is again just given by the Metropolis factor,

$$\frac{f(G_i \mid \Theta^*)}{f(G_i \mid \Theta_i)} = \prod_{k:\Psi_k > T} \frac{a_i}{a^*}\exp\left(\tau_k\binom{n_k}{2}(a_i^{-1} - a^{*-1})\right), \quad \text{(A3)}$$

where n_k is the number of edges in the genealogy in the kth interval, $\Theta^* = \{a^*\} \cup \Theta_i\backslash\{a_i\}$, and τ_k now is defined so that $\tau_k = \psi_k - T$ if $\psi_k > T$ and $\psi_{k-1} < T$.

Updates of θ: This type of update is similar to the updates previously described. A new value of θ (θ*) is drawn from the interval $[\theta_i - \Delta_\theta, T_i + \Delta_\theta]$, where Δ_θ is chosen such that $\Delta_\theta \leq \theta_{max}$. The acceptance probability is

$$\frac{f(G_i \mid \Theta^*)}{f(G_i \mid \Theta_i)} = \left(\frac{\theta^*}{\theta_i}\right)^S e^{\beta_i(\theta_i - \theta^*)}, \quad \text{(A4)}$$

where $\Theta^* = \{\theta^*\} \cup \Theta_i\backslash\{\theta_i\}$, S is the total number of mutations in the tree, and β_i is the total tree length of genealogy G_i.

Updates of *T*: In the following discussion, time is increasing backward in time and the time of sampling is $t = 0$. Updates to T are chosen using the method described for the other parameters, *i.e.*, uniformly in an interval $[T_i - \Delta_T, T_i + \Delta_T]$. However, some care must be taken to preserve the reversibility conditions of the chain. If we denote the time at which edge j arises by a coalescence event by t_j^{start} and the time it ends in a coalescence event with another edge by t_j^{end}, then it will occur for some j that $t_j^{\text{start}} < T^*$ and $t_j^{\text{end}} < T_i$ if $T^* > T_i$ and T_i is larger than the time to the most recent common ancestor (TMRCA); *i.e.*, edge j exists in the time interval $[T_i, T^*]$.

A solution to this problem is simply to simulate the path of migrations of the edges in the time interval $[T_i, T^*]$. An update to T is always a joint update of T and G. A method for performing such simulations is given by the following algorithm:

1. Set $j = 1$.
2. If edge j exists in time interval $[T_i, T^*]$, simulate migrations on the edge according to a two-state continuous time Markov chain with infinitesimal generator

$$q = \begin{bmatrix} -M_1 & M_1 \\ M_2 & -M_2 \end{bmatrix}$$

starting at the current state of the edge at time $\min\{T_i, t_j^{\text{start}}\}$ and stopping at time $\max\{T^*, t_j^{\text{end}}\}$.

3. If edge j exists in time interval $[T_i, T^*]$, $t_j^{\text{end}} < T^*$, j has an index larger than its sister edge (edge j is a right sister edge) and edge j does not end up in the same population as its sister edge: Repeat (2).
4. If j is less than the total number of edges in the genealogy, set $j = j + 1$ and repeat.

This algorithm does not simulate migration paths ac-

cording to the prior distribution, but does simulate them according to an approximation of this distribution. A Hastings term ensures that the Markov chain has the desired stationary distribution. Also, note that this algorithm will be very slow when M_1 and/or M_2 and $\kappa_j = t_j^{\text{end}} - \min\{T_i, t_j^{\text{start}}\}$ is small and an odd number of migration events must occur on a right edge for it to end (coalesce) in the same population as its sister edge. The algorithm can be speeded up in such cases by letting the first migration event be simulated conditional on at least one migration event occurring. For example, if edge j exists in population 1 at time $\min\{T_i, t_j^{\text{start}}\}$ but must be in population 2 at time t_j^{end}, then the probability density of the time to the first migration event (t_M), conditional on at least one migration event occurring, is given by

$$f(t_M \mid \text{at least one migration event}) = \frac{M_1 e^{-M_1(t_M - \min\{T_i, t_j^{\text{start}}\})}}{1 - e^{-M_1 K_j}},$$

$$t_j^{\text{end}} > t_M \geq \min\{T_i, t_j^{\text{start}}\}. \quad \text{(A5)}$$

Using the inverse transformation method, the time to the first migration event can be simulated by choosing

$$t'_M = \min\{T_i, t_j^{\text{start}}\} - (\log(1.0 - U(1.0 - \exp[-M_{1\kappa j}])))/M_1, \quad \text{(A6)}$$

where U is a uniform $[0, 1]$ random number. It is thereby possible to simulate migration paths relatively fast even when M_1 and κ are small and an odd number of migration events must occur on a right edge for it to end (coalesce) in the same population as its sister edge.

If $T^* < T_i$ the migration events occurring in the interval $[T^*, T_i]$ are simply erased. The acceptance ratio for this type of update is

$$\frac{f(G^* \mid \Theta^*)q[(\Theta^*, G^*)\to(\Theta_i, G_i)]}{f(G_i \mid \Theta_i)q[(\Theta_i, G_i)\to(\Theta^*, G^*)]} = g_1(\Theta^*, G^*, \Theta_i, G_i)$$

$$\times \sum_{j \in B}\sum\sum_{k:T^*<t_{jk}\leq T_i} g_2(M_1, M_{2i}, t_{jk}, t_{j(k-1)})g_3(M_1, M_2, \kappa_j), \quad \text{(A7)}$$

where

$$g_1(\Theta^*, G^*, \Theta_i, G_i) = \sum_{k:T^*<\psi_k\leq T_i} \exp\left[\tau_k\left(\binom{n_{1i} + n_{2i}}{2}a_i^{-1} - n_1 M_1 - n_2 M_2\right.\right.$$

$$\left.\left. - \binom{n_{1i}}{2} - \binom{n_{2i}}{2}r_i^{-1}\right)\right]$$

$$\times \begin{cases} a_i & \text{if interval } k \text{ ends in a coalescence} \\ & \quad \text{event in population 1} \\ a_i/r_i & \text{if interval } k \text{ ends in a coalescence} \\ & \quad \text{event in population 2,} \\ 1 & \text{else} \end{cases}$$

$$g_2(M_1, M_{2i}, t_{jk}, t_{j(k-1)}) = \begin{cases} e^{-M_1(t_{jk}-t_{j(k-1)})} \\ \quad \text{if migration is from population 1 to population 2} \\ e^{-M_2(t_{jk}-t_{j(k-1)})} \\ \quad \text{if migration is from population 2 to population 1} \end{cases}$$

$$
g_3(M_1, M_2, \kappa_j) = \begin{cases} (M_2 + M_1 e^{-(M_2+M_1)\kappa_j})/(M_1 + M_2) \\[4pt] \quad \text{if odd number of migrations and} \\ \quad\quad \text{first migration is from population 1} \\[4pt] (M_1 + M_2 e^{-(M_2+M_1)\kappa_j})/(M_1 + M_2) \\[4pt] \quad \text{if odd number of migrations and} \\ \quad\quad \text{first migration is from population 2} \\[4pt] 1 - (M_1 + M_2 e^{-(M_2+M_1)\kappa_j})/(M_1 + M_2) \\[4pt] \quad \text{if even number of migrations and} \\ \quad\quad \text{first migration is from population 1} \\[4pt] 1 - (M_2 + M_1 e^{-(M_2+M_1)\kappa_j})/(M_1 + M_2) \\[4pt] \quad \text{if even number of migrations and} \\ \quad\quad \text{first migration is from population 2} \end{cases}
$$

and ψ_k is defined as in the case of updates of M_1 and M_2, B is the set of all edges existing in the interval $[T^*, T_i]$, and t_{jk} is the time of the kth migration on the jth edge, $t_{j0} =. \min\{T_i, t_j^{\text{start}}\}$. If $T^* > T_i$, the acceptance ratio can be calculated similarly as the inverse term. This completes the description of the MCMC algorithm.