Available online at www.sciencedirect.com
SCIENCE DIRECT®
Theoretical Population Biology 68 (2005) 41–53
Theoretical Population Biology
www.elsevier.com/locate/ytpbi

Stepwise mutation likelihood computation by sequential importance sampling in subdivided population models

Maria De Iorio[a,b], Robert C. Griffiths[c,*], Raphael Leblois[d], François Rousset[d]

[a]Department of Mathematics, Imperial College, UK
[b]Department of Epidemiology and Public Health, Imperial College, St Mary's Campus, Norfolk Place, London W2 1PG, UK
[c]Department of Statistics, University of Oxford, 1, South Parks Road, Oxford OX1 3TG, UK
[d]Laboratoire Génétique et Environnement, CNRS-UMR 5554, 34095 Montpellier, France

Received 28 October 2003
Available online 10 May 2005

Abstract

An importance sampling algorithm for computing the likelihood of a sample of genes at loci under a stepwise mutation model in a subdivided population is developed. This allows maximum likelihood estimation of migration rates between subpopulations. The time to the most recent common ancestor of the sample can also be computed. The technique is illustrated by an analysis of a data set of Australian red fox populations.

MSC: primary 60G40; 93E25; 92D25

Keywords: Coalescent process; Importance sampling; Migration; Stepwise mutation

1. Introduction

Recently there has been much research into computation methods of ancestral inference from samples of genes conditional on their observed type configuration using importance sampling (IS), MCMC and Bayesian techniques. In the simplest stepwise mutation model inference involves finding a maximum likelihood estimate of θ, the mutation rate. Wilson and Balding (1998) use a Bayesian MCMC scheme implemented in `Micsat` for microsatellite data, Beerli and Felsenstein (1999) use an MCMC scheme implemented in `Migrate` to estimate migration rates from data which could have a stepwise mutation mechanism. Nielsen (1997) uses an IS algorithm based on an algorithm in Griffiths and Tavaré (1994). Stephens and Donnelly (2000) develop a sequential IS technique, improving the algorithm of Griffiths and Tavaré (1994) to find the likelihood of samples of genes under a general mutation model which includes the stepwise model. Stephens (2001) gives an algorithm based on IS to simulate genealogies of selected alleles in a population of variable size. The paper also describes the use of IS methods in population genetics in a more general framework. Chen et al. (2005) improve sequential IS by running multiple processes and resampling at sequential steps in the algorithm. A different approach based on F-statistics and analysis of variance analogues is reviewed in Excoffier (2001) and Rousset (2001). De Iorio and Griffiths (2004a,b) develop a technique to construct sequential IS proposal distributions on coalescent histories in population genetic models based on the diffusion process generator that describes the distribution of population gene frequencies. The technique extends proposal distributions of Stephens and Donnelly (2000) to a wider class of models, with a focus on subdivided population models of the island model type. Apart from likelihood calculations, ancestral inference questions involving time, such as the time to the most recent common

*Corresponding author. Fax: +44 1865 272595.
E-mail address: griff@stats.ox.ac.uk (R.C. Griffiths).

0040-5809/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.tpb.2005.02.001

ancestor (TMRCA) of the genes in a data set, can be answered by including the time between events in the underlying coalescent process.

The ancestry of a sample of n genes is described by a coalescent tree (Kingman, 1982), where pairs of ancestral lines coalesce at unit rate forming a tree back in time to the ancestor of the sample genes. Mutations occur at rate $\theta/2$ along the edges of the coalescent tree according to a Poisson process.

An accessible introduction to the coalescent process is Nordborg (2001), with subdivided coalescent studies by Notohara (1990) and Herbots (1997). Two papers on ancestral inference are Griffiths (2001) and Stephens (2001). This current paper derives a detailed algorithm from the techniques in De Iorio and Griffiths (2004b) for likelihood calculations under a stepwise mutation model in a subdivided population.

In the stepwise mutation model the allele type space is the set of integers $\{\ldots -2,-1,0,1,\ldots\}$. Transitions of allele type when a mutation occurs are made according to a random walk from state j to a state $j+Z$, where Z is an integer-valued random variable. In the simplest case studied by Ohta and Kimura (1973), Moran (1975), Moran (1976), $Z=\pm 1$ with probability $1/2$. There is a distribution of the configuration of types in a sample with positions measured relative to the most recent common ancestor of the sample, which could be taken without loss of generality to be 0. The distribution is shift invariant, only depending on the relative positions of the types on a line.

In this paper an IS algorithm for computing the likelihood of a sample of genes at loci under a stepwise mutation model in a subdivided population is developed. This allows maximum likelihood estimation of migration rates between subpopulations. The algorithm and program code used were thoroughly checked. Intermediate calculations also agree with exact analytic results for probabilities of identity of type under a finite island model. See Nagylaki (1983) for the finite island model, and Rousset (1996) for an adaption to different mutation models.

The precision of migration and mutation rate maximum likelihood estimates is checked by an analysis of multilocus simulated data from a two population model.

To illustrate the algorithm with a real data set, an analysis is performed on data from two Australian red fox (*Vulpes vulpes*) populations.

2. Coalescent histories and importance sampling

Let E be the set of possible types of a gene. Denote the sample configuration of the numbers of different types as $\boldsymbol{n}=(n_j, j\in E)$, and $p(\boldsymbol{n})$ the probability of obtaining a sample $\boldsymbol{n}$. $\boldsymbol{e}_j$ will denote the jth unit vector. For $j\in E$, let $\pi(j\mid \boldsymbol{n})$ be the probability that an additional gene chosen from the population is of type j, given that we have an observed configuration $\boldsymbol{n}$. These conditional distributions are important in the sampling distribution and coalescent history process. If $j_1,j_2,\ldots,j_n$ are the types of genes sequentially sampled such that there are $n_j, j\in E$ genes of type j in the sample, then

$$p(\boldsymbol{n})=\frac{n!}{\prod_{j\in E} n_j!}\prod_{l=1}^{n}\pi(j_l\mid \boldsymbol{e}_{j_1}+\cdots+\boldsymbol{e}_{j_{l-1}}).$$

The distribution $p(\boldsymbol{n})$ is invariant under sequential sampling order.

A coalescent history $\{H_k,\ k=0,-1,\ldots,-m\}$ is defined as the set of ancestral configurations at the embedded events in the Markov process where coalescence, mutation or other events take place. H_0 denotes the current state, and H_{-m} the state when a singleton ancestor is reached. The Markov nature of the process implies that

$$p(H_k)=\sum_{\{H_{k-1}\}} p(H_k\mid H_{k-1})p(H_{k-1}). \tag{2.1}$$

$p(H_k)$ and $\{p(H_{k-1})\}$ are unknown, whereas the probabilities $p(H_k\mid H_{k-1})$ are easily derived from the distribution of the coalescent tree. A coalescent history is illustrated in Fig. 1. In (2.1) history probabilities are evaluated in the direction from the ancestor type to the sample data configuration. However a reverse history process from the sample data to the ancestor is required to efficiently evaluate the sample likelihood. A sequential importance sampling representation is based on an approximation $\hat{p}(H_{k-1}\mid H_k)$ to the unknown reverse probabilities $p(H_{k-1}\mid H_k)$. In one step

$$\begin{aligned} p(H_k)&=\sum_{\{H_{k-1}\}}\frac{p(H_k\mid H_{k-1})}{\hat{p}(H_{k-1}\mid H_k)}p(H_{k-1})\hat{p}(H_{k-1}\mid H_k)\\ &=\mathbb{E}_{\hat{p}}\sum_{\{H_{k-1}\}}\left[\frac{p(H_k\mid H_{k-1})}{\hat{p}(H_{k-1}\mid H_k)}\mid H_k\right]. \end{aligned} \tag{2.2}$$

The full sequential IS representation from continuing (2.2) over states $H_0,H_{-1},\ldots,H_{-m}$ is

$$\begin{aligned} p(H_0)=\mathbb{E}_{\hat{p}}\Bigg[&\frac{p(H_0\mid H_{-1})}{\hat{p}(H_{-1}\mid H_0)}\cdots\frac{p(H_{-m+1}\mid H_{-m})}{\hat{p}(H_{-m}\mid H_{-m+1})}\\ &\times p(H_{-m})\Bigg], \end{aligned} \tag{2.3}$$

where $\mathbb{E}_{\hat{p}}$ is taken over histories $H_{-1},\ldots,H_{-m}$ with $\hat{p}(\cdot\mid\cdot)$ being the reverse chain transition probabilities. Probabilities of a history sample path $\mathscr{H}$ are evaluated in forward and reverse directions in the numerator and denominator of (2.3). The likelihood of the data can be evaluated by repeated simulation of sample histories in a reverse direction from the current sample configuration H_0 to H_{-m} under $\hat{p}(\cdot)$ with transition probabilities $\hat{p}(H_{k-1}\mid H_k)$, then averaging the sequential IS weights

$$\frac{p(H_0\mid H_{-1})}{\hat{p}(H_{-1}\mid H_0)}\cdots\frac{p(H_{-m+1}\mid H_{-m})}{\hat{p}(H_{-m}\mid H_{-m+1})}p(H_{-m})$$

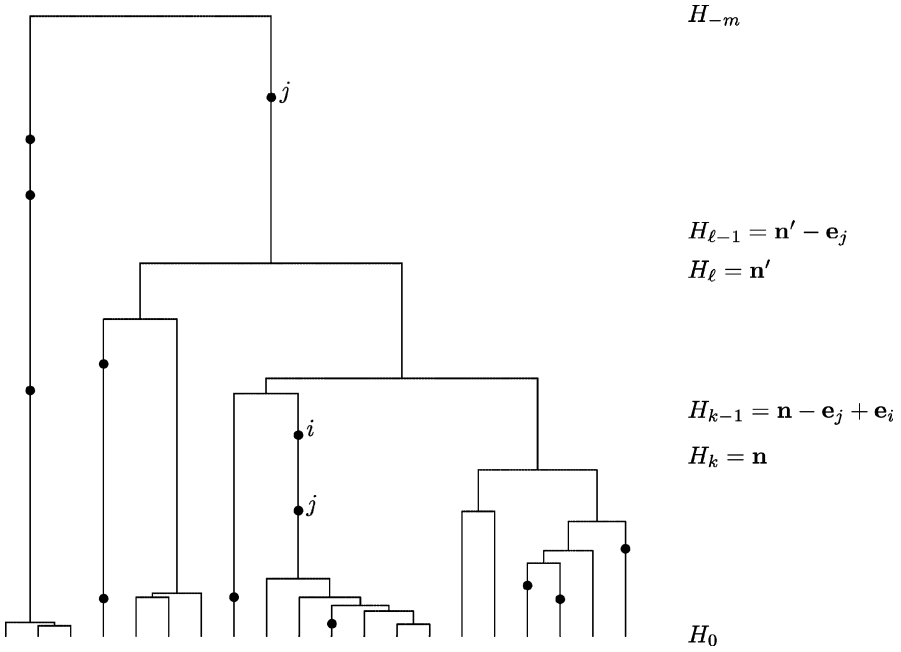

Fig. 1. Coalescent tree.

obtained on each run to obtain an estimate of the likelihood.

In a panmictic model if a historical configuration is $H_k = \boldsymbol{n}$ then either $H_{k-1} = \boldsymbol{n} - \boldsymbol{e}_j$ for some $j \in E$ corresponding to coalescence of two type j genes, or $H_{k-1} = \boldsymbol{n} + \boldsymbol{e}_i - \boldsymbol{e}_j$ for some $i,j \in E$ corresponding to mutation forward in time from i to j chosen with transition probability matrix P. (See Fig. 1 for an illustration.) In detail (2.1) becomes

$$p(\boldsymbol{n}) = \frac{\theta}{n+\theta-1} \sum_{i,j\in E, n_j>0} \frac{n_i + 1 - \delta_{ij}}{n} P_{ij} p(\boldsymbol{n} - \boldsymbol{e}_j + \boldsymbol{e}_i) + \frac{n-1}{n+\theta-1} \sum_{j\in E, n_j>0} \frac{n_j - 1}{n-1} p(\boldsymbol{n} - \boldsymbol{e}_j). \tag{2.4}$$

In (2.4) $\delta_{ij} = 1$ if $i = j$ or $\delta_{ij} = 0$ otherwise. Griffiths and Tavaré (1994) derive (2.4) from a coalescent argument and also by considering a sample from the population frequencies in a diffusion process model.

It is possible to express the reverse chain probabilities by using Bayes' rule as

$$P(H_{k-1} \mid H_k) = \begin{cases} \dfrac{n_j - 1}{n+\theta-1} \cdot \dfrac{n_j}{n} \cdot \dfrac{1}{\pi(j \mid \boldsymbol{n} - \boldsymbol{e}_j)} & \text{if } H_{k-1} = \boldsymbol{n} - \boldsymbol{e}_j, \\ \dfrac{\theta}{n+\theta-1} \cdot \dfrac{n_j}{n} \cdot \dfrac{\pi(i \mid \boldsymbol{n} + \boldsymbol{e}_i - \boldsymbol{e}_j)}{\pi(j \mid \boldsymbol{n} - \boldsymbol{e}_j)} & \text{if } H_{k-1} = \boldsymbol{n} + \boldsymbol{e}_i - \boldsymbol{e}_j. \end{cases} \tag{2.5}$$

IS distributions are found by substituting an approximation $\widehat{\pi}$ to obtain $\widehat{P}(H_{k-1} \mid H_k)$. $p(\boldsymbol{n})$ can then be calculated by simulation. Stephens and Donnelly (2000) construct an IS proposal distribution on coalescent histories approximating $\pi(\cdot \mid \boldsymbol{n})$, by $\widehat{\pi}(\cdot \mid \boldsymbol{n})$, the stationary distribution in a Markov chain with transition probability matrix

$$\frac{\theta P + \boldsymbol{n}}{n+\theta}. \tag{2.6}$$

That is, for $j \in E$,

$$\widehat{\pi}(j \mid \boldsymbol{n}) = \sum_{i\in E} \widehat{\pi}(i \mid \boldsymbol{n}) \frac{\theta P_{ij} + n_j}{n+\theta}. \tag{2.7}$$

An explicit solution to (2.7) is that

$$\widehat{\pi}(j \mid \boldsymbol{n}) = \sum_{i\in E} \frac{n_i}{n} \sum_{k=0}^{\infty} \rho^k (1-\rho) P_{ij}^{(k)} = \sum_{i\in E} \frac{n_i}{n} Q_{ij}, \tag{2.8}$$

where $\rho = \theta/(n+\theta)$ and the transition matrix $Q = (1-\rho)(I - \rho P)^{-1}$. De Iorio and Griffiths (2004a,b) provide three ways of justifying the approximation $\widehat{\pi}$; from an approximation to the generator of the diffusion process describing the distribution of the population gene frequencies; from the recursive equations for the sampling distribution; and from a coalescent argument.

A stochastic interpretation of (2.8) is to choose a type $i \in E$ gene with probability n_i/n, then obtain a type $j \in E$ gene from a Geometric $(\theta/(n+\theta))$ number of mutations according to the transition matrix P. In a stepwise mutation model where $P_{ij} = 1/2$ if $|i-j| = 1$, or zero otherwise, it is shown

Table 1
IS proposal distribution and importance weights for a coalescent model with migration

H_{k-1}	Proposal distribution	Importance weight
$\boldsymbol{n}-\boldsymbol{e}_{\alpha j}$	$\frac{n_{\alpha j}(n_{\alpha j}-1)q_\alpha^{-1}}{\widehat{\pi}(j \mid \alpha, \boldsymbol{n}-\boldsymbol{e}_{\alpha j})D(\boldsymbol{n})}$	$\frac{n_\alpha}{n_{\alpha j}}\widehat{\pi}(j \mid \alpha, \boldsymbol{n}-\boldsymbol{e}_{\alpha j})$
$\boldsymbol{n}-\boldsymbol{e}_{\alpha j}+\boldsymbol{e}_{\alpha i}$	$\frac{n_{\alpha j}\theta P_{ij}\widehat{\pi}(i \mid \alpha, \boldsymbol{n}-\boldsymbol{e}_{\alpha j})}{\widehat{\pi}(j \mid \alpha, \boldsymbol{n}-\boldsymbol{e}_{\alpha j})D(\boldsymbol{n})}$	$\frac{(n_{\alpha i}+1-\delta_{ij})}{n_{\alpha j}}\frac{\widehat{\pi}(j \mid \alpha, \boldsymbol{n}-\boldsymbol{e}_{\alpha j})}{\widehat{\pi}(i \mid \alpha, \boldsymbol{n}-\boldsymbol{e}_{\alpha j})}$
$\boldsymbol{n}-\boldsymbol{e}_{\alpha j}+\boldsymbol{e}_{\beta j}$	$\frac{n_{\alpha j}m_{\alpha\beta}\widehat{\pi}(j \mid \beta, \boldsymbol{n}-\boldsymbol{e}_{\alpha j})}{\widehat{\pi}(j \mid \alpha, \boldsymbol{n}-\boldsymbol{e}_{\alpha j})D(\boldsymbol{n})}$	$\frac{(n_{\beta j}+1)}{n_{\alpha j}}\frac{n_\alpha}{(n_\beta+1)}\frac{\widehat{\pi}(j \mid \alpha, \boldsymbol{n}-\boldsymbol{e}_{\alpha j})}{\widehat{\pi}(j \mid \beta, \boldsymbol{n}-\boldsymbol{e}_{\alpha j})}$

in this paper that

$$Q_{ij}=\frac{1-\rho}{\sqrt{1-\rho^2}}\cdot\left[\frac{\rho}{1+\sqrt{1-\rho^2}}\right]^{|j-i|}. \tag{2.9}$$

In a subdivided population model with g subpopulations let S be the subpopulation type space. A gene's type is then indexed by $S\times E$, the subpopulation it is in, and its allele type. Possible transitions back in time to a sample of genes at a prior history event to a gene type (α,j) are: coalescence of a pair of genes of type (α,j); mutation forward in time from type (α,i) to (α,j) with rate $\theta/2$ and transition probability P_{ij}; and migration back in time of a type j gene from subpopulation α to β at rate $m_{\alpha\beta}/2$. Let $m_\alpha=\sum_{\beta\neq\alpha}m_{\alpha\beta}$, and denote $(q_\alpha,\alpha\in S)$ as relative subpopulation sizes. A Wright–Fisher model in discrete time gives rise to this model as subpopulation sizes tend to infinity. Let $(N_\alpha)_{\alpha\in S}$ be the subpopulation sizes, $N=\sum_{\alpha\in S}N_\alpha$, $q_\alpha=N_\alpha/N$, $\alpha\in S$, and $v_{\alpha\beta},\alpha,\beta\in S$ be the probability that the parent of an offspring in subpopulation α is from subpopulation β in the previous generation. The backward migration rates are defined as $m_{\alpha\beta}=2Nv_{\alpha\beta}$, $\alpha,\beta\in\Gamma$, $\alpha\neq\beta$ with the overall rate $m_\alpha=\sum_{\beta\neq\alpha}m_{\alpha\beta}$. If $(\widetilde{m}_{\beta\alpha})$ are the forward migration rates then $\widetilde{m}_{\beta\alpha}=N_\alpha m_{\alpha\beta}/N_\beta$. The model considered here is the usual coalescent time scaled model where time is measured in units of N generations, and $N\to\infty$ while migration and mutation rates are kept constant. A careful treatment of the limit is in Herbots (1997). The analogue of (2.4) for a subdivided population model, derived in De Iorio and Griffiths (2004b), is

$$\begin{aligned}\left(\sum_{\alpha=1}^{g}n_\alpha(n_\alpha-1)q_\alpha^{-1}+\sum_{\alpha=1}^{g}n_\alpha m_\alpha+n\theta\right)p(\boldsymbol{n})&\\ =\sum_{\alpha=1}^{g}\sum_{j\in E}n_\alpha(n_{\alpha j}-1)q_\alpha^{-1}p(\boldsymbol{n}-\boldsymbol{e}_{\alpha j})&\\ +\theta\sum_{\alpha=1}^{g}\sum_{i,j\in E}(n_{\alpha i}+1-\delta_{ij})P_{ij}p(\boldsymbol{n}+\boldsymbol{e}_{\alpha i}-\boldsymbol{e}_{\alpha j})&\\ +\sum_{\alpha=1}^{g}\sum_{j\in E}\sum_{\beta\neq\alpha}m_{\alpha\beta}\frac{n_\alpha}{n_\beta+1}(n_{\beta j}+1)&\\ \times p(\boldsymbol{n}-\boldsymbol{e}_{\alpha j}+\boldsymbol{e}_{\beta j}).&\end{aligned} \tag{2.10}$$

Bahlo and Griffiths (2001) obtain a general solution for $p(\boldsymbol{n})$ when the sample size is $n=2$, though the form of solution is not simple. For $(\alpha,j)\in S\times E$, let $\pi(j\mid\alpha,\boldsymbol{n})$ be the probability that an additional gene taken from subpopulation α is of type j, given that we have an observed configuration $\boldsymbol{n}=(n_{\alpha j})$. The scaling is such that $\sum_{j\in E}\pi(j\mid\alpha,\boldsymbol{n})=1$. The reverse chain probabilities $p(H_{k-1}\mid H_k)$ can be expressed in terms of $\pi(\cdot\mid\alpha,\boldsymbol{n})$. The proposal distribution $\widehat{p}(H_{k-1}\mid H_k)$ and one-step IS weights, $p(H_k\mid H_{k-1})/\widehat{p}(H_{k-1}\mid H_k)$ based on approximate distributions $\widehat{\pi}(\cdot\mid\alpha,\boldsymbol{n})$, derived in De Iorio and Griffiths (2004b), are shown in Table 1. The probability distributions $\widehat{\pi}(\cdot\mid\alpha,\boldsymbol{n})$ are defined by a system of equations

$$\begin{aligned}(n_\alpha q_\alpha^{-1}+m_\alpha+\theta)\widehat{\pi}(j\mid\alpha,\boldsymbol{n})=n_{\alpha j}q_\alpha^{-1}+\theta\sum_{i\in E}P_{ij}\widehat{\pi}(i\mid\alpha,\boldsymbol{n})&\\ +\sum_{\beta\neq\alpha}m_{\alpha\beta}\widehat{\pi}(j\mid\beta,\boldsymbol{n}).&\end{aligned} \tag{2.11}$$

The overall event rate in subpopulation α is $d_\alpha/2$, where $d_\alpha=n_\alpha(n_\alpha-1)q_\alpha^{-1}+n_\alpha m_\alpha+n_\alpha\theta$, and the total event rate is $D(\boldsymbol{n})/2$ where $D(\boldsymbol{n})=\sum_{\alpha=1}^{g}d_\alpha$. In practice it is easiest to choose a gene of type (α,j) to change with probability

$$\frac{n_{\alpha j}((n_\alpha-1)q_\alpha^{-1}+m_\alpha+\theta)}{D(\boldsymbol{n})}, \tag{2.12}$$

then select an associated event from the conditional proposal distribution found by dividing the proposal distribution in Table 1 by (2.12).

Here we describe briefly how a coalescent approximation gives rise to $\widehat{\pi}$. Let $B_{\alpha j}$ be the event that a gene from subpopulation α of type j is involved in the first event back in the coalescent history of the process and $\boldsymbol{Y}$ a random vector describing the configuration of types. Then

$$\begin{aligned}p(B_{\alpha j}\cap\{\boldsymbol{Y}=\boldsymbol{n}\})&=p(\boldsymbol{n})\mathbb{P}(B_{\alpha j}\mid\boldsymbol{Y}=\boldsymbol{n})\\ &=\frac{n_\alpha(n_\alpha-1)q_\alpha^{-1}}{D(\boldsymbol{n})}\sum_{j\in E}\frac{n_{\alpha j}-1}{n_\alpha-1}p(\boldsymbol{n}-\boldsymbol{e}_{\alpha j})\\ &\quad+\frac{n_\alpha\theta}{D(\boldsymbol{n})}\sum_{i,j\in E}\frac{n_{\alpha i}+1-\delta_{ij}}{n_\alpha}\\ &\quad\times P_{ij}p(\boldsymbol{n}-\boldsymbol{e}_{\alpha j}+\boldsymbol{e}_{\alpha i})\end{aligned}$$

$$+\frac{n_\alpha m_\alpha}{D(\boldsymbol{n})}\sum_{j\in E}\sum_{\beta\neq\alpha}\frac{m_{\alpha\beta}}{m_\alpha}\cdot\frac{n_{\beta j}+1}{n_\beta+1}$$
$$\times p(\boldsymbol{n}-\boldsymbol{e}_{\alpha j}+\boldsymbol{e}_{\beta j}). \quad (2.13)$$

Exchangeability in the order of sampled genes implies that

$$\pi(j\mid\alpha,\boldsymbol{n}-\boldsymbol{e}_{\alpha j})p(\boldsymbol{n}-\boldsymbol{e}_{\alpha j})=\frac{n_{\alpha j}}{n_\alpha}p(\boldsymbol{n}),$$
$$\pi(i\mid\alpha,\boldsymbol{n}-\boldsymbol{e}_{\alpha j})p(\boldsymbol{n}-\boldsymbol{e}_{\alpha j})=\frac{n_{\alpha i}+1-\delta_{ij}}{n_\alpha}p(\boldsymbol{n}-\boldsymbol{e}_{\alpha j}+\boldsymbol{e}_{\alpha i}),$$
$$\pi(j\mid\beta,\boldsymbol{n}-\boldsymbol{e}_{\alpha j})p(\boldsymbol{n}-\boldsymbol{e}_{\alpha j})=\frac{n_{\beta j}+1}{n_\beta+1}p(\boldsymbol{n}-\boldsymbol{e}_{\alpha j}+\boldsymbol{e}_{\beta j}). \quad (2.14)$$

Substituting from (2.14) into (2.13)

$$\pi(j\mid\alpha,\boldsymbol{n}-\boldsymbol{e}_j)\mathbb{P}(B_{\alpha j}\mid\boldsymbol{Y}=\boldsymbol{n})D(\boldsymbol{n})/n_\alpha$$
$$=(n_{\alpha j}-1)q_\alpha^{-1}+\theta\sum_{i\in E}P_{ij}\pi(i\mid\alpha,\boldsymbol{n}-\boldsymbol{e}_j)$$
$$+\sum_{\beta\neq\alpha}m_{\alpha\beta}\pi(j\mid\beta,\boldsymbol{n}-\boldsymbol{e}_j). \quad (2.15)$$

The system (2.15) is exact, rather than approximate. To obtain the approximate system (2.11) assume that

$$\mathbb{P}(B_{\alpha j}\mid\boldsymbol{Y}=\boldsymbol{n})=\frac{n_\alpha(n_\alpha-1)q_\alpha^{-1}+n_\alpha m_\alpha+n_\alpha\theta}{D(\boldsymbol{n})}\cdot\frac{n_{\alpha j}}{n_\alpha},$$

the probability of the last history event being in subpopulation α, times the approximate probability $n_{\alpha j}/n_\alpha$. Then substituting in (2.15) and setting $\boldsymbol{n}-\boldsymbol{e}_j\to\boldsymbol{n}$ yields (2.11).

A stochastic interpretation of the distribution $\widehat{\pi}$ is shown in De Iorio and Griffiths (2004b) to be the following. Let $M^\circ=(m_{\alpha\beta}/m_\alpha)$ be a transition probability matrix with diagonal elements zero constructed from the migration rate matrix M, so that the rows of M° each add to 1. Denote, for $\alpha\in S$,

$$\phi_\alpha=\frac{m_\alpha}{n_\alpha q_\alpha^{-1}+m_\alpha},\quad \rho_\alpha=\frac{\theta}{n_\alpha q_\alpha^{-1}+m_\alpha+\theta},$$

and the transition probability matrix $P_\alpha=(1-\rho_\alpha)(I-\rho_\alpha P)^{-1}$. A mechanism for choosing a gene of type $j\in E$ from the distribution $\widehat{\pi}(j\mid\alpha,\boldsymbol{n})$ is the following. Choose a sequence of subpopulations $\alpha_0,\alpha_1,\ldots,\alpha_\tau$, for the migration path of a gene, starting with $\alpha_0=\alpha$ and stopping at step τ in subpopulation α_τ, with probability

$$\phi_{\alpha_0}\phi_{\alpha_1}\cdots\phi_{\alpha_{\tau-1}}(1-\phi_{\alpha_\tau})\cdot m^\circ_{\alpha_0\alpha_1}m^\circ_{\alpha_1\alpha_2}\cdots m^\circ_{\alpha_{\tau-1}\alpha_\tau}.$$

ϕ_α can be interpreted as the probability of moving from subpopulation α to another subpopulation, while $1-\phi_\alpha$ is the probability of stopping in subpopulation α. Next choose a type at random from subpopulation α_τ, such that the probability of choosing a gene of type i is $n_{\alpha_\tau i}/n_{\alpha_\tau}$. Mutate back along the migration path to α_0, so that a sample path probability of a sequence of mutations which start with type $i_{\alpha_\tau}=i$ and end with a type $i_0=j$ gene is

$$\frac{n_{\alpha_\tau i_{\alpha_\tau}}}{n_{\alpha_\tau}}P_{\alpha_\tau;i_\tau i_{\tau-1}}\cdots P_{\alpha_1;i_2 i_1}P_{\alpha_0;i_1 i_0}.$$

An interpretation of P_α is that there are a geometrically distributed number of mutations with parameter ρ_α, the probability of a mutation, and a transition matrix P for type changes, in each subpopulation $\alpha\in\{\alpha_\tau,\ldots,\alpha_0\}$ visited in the migration path. The stochastic structure described above can be seen by rewriting (2.11) as

$$\widehat{\pi}(j\mid\alpha,\boldsymbol{n})=(1-\rho_\alpha)(1-\phi_\alpha)\frac{n_{\alpha j}}{n_\alpha}+(1-\rho_\alpha)\phi_\alpha$$
$$\times\sum_{\beta\neq\alpha}m^\circ_{\alpha\beta}\widehat{\pi}(j\mid\beta,\boldsymbol{n})$$
$$+\rho_\alpha\sum_{i\in E}P_{ij}\widehat{\pi}(i\mid\alpha,\boldsymbol{n}). \quad (2.16)$$

In the simple stepwise mutation model $P_\alpha=Q$, in (2.8) with parameter $\rho=\rho_\alpha$.

3. Stepwise mutation model

The population genetics model considered in this paper is a subdivided population with stepwise mutations on the line. The gene type space is then $E=\{\ldots,-2,-1,0,1,2,\ldots\}$ with a mutation transition matrix of the form

$$P_{ij}=u_{j-i},\quad i,j=0,\pm1,\pm2,\ldots\ .$$

A sample of $(n_\alpha;\alpha\in S)$ genes is taken from the subpopulations. $\boldsymbol{n}=(n_{\alpha j};(\alpha,j)\in S\times E)$ is the collection of the number of genes in subpopulation α of type j. Backwards migration rates from subpopulation α to β are denoted by $(m_{\alpha\beta};\alpha,\beta\in S)$. $p(\boldsymbol{n})$ will denote the probability of a configuration $\boldsymbol{n}$ under the model.

Fourier transform methods will be used in the following sections to solve recursive equations obtained. The Fourier transform of an absolutely convergent series $\{a_j;j=0,\pm1,\pm2\ldots\}$ will be denoted by

$$a^*(\xi)=\sum_{j=-\infty}^{\infty}e^{\imath\xi j}a_j,$$

where $\imath=\sqrt{-1}$.

3.1. A sample of two genes

It is possible to obtain a formula for $p(\boldsymbol{n})$ when $n=2$ by using Fourier transform methods. We provide details of the simplest case with two subpopulations labelled α and β. Let $p(\alpha,\beta;\Delta)$ be the probability that two genes chosen from subpopulations α and β are separated by a signed distance $\Delta=0,\pm1,\pm2,\ldots$. $p(\alpha,\alpha;\Delta)$ is interpreted as the probability that the signed distance of the second gene from the first gene chosen from subpopulation α is Δ. If $\boldsymbol{n}_{\alpha\beta}$ is a sample of two genes from

subpopulations α and β with respective positions i and j such that $i-j=\Delta$, then for $\alpha\neq\beta$, $p(\alpha,\beta;d)=p(\boldsymbol{n}_{\alpha\beta})$, while for $\alpha=\beta$, $p(\alpha,\alpha;\Delta)=p(\boldsymbol{n}_{\alpha\alpha})/(2-\delta_{\Delta,0})$, where $\delta_{\Delta,0}=1$ if $\Delta=0$ or $\delta_{\Delta,0}=0$ if $\Delta\neq 0$. The scaling is such that $\sum_{\Delta=-\infty}^{\infty}p(\alpha,\beta;d)=1$ for α and β equal or unequal. Without loss of generality choosing $j=0$, substituting in (2.10), (or a derivation from first principles for a sample of two genes) and using translation invariance gives the following equations for $\alpha\neq\beta$,

$$\begin{aligned}&(m_\alpha+m_\beta+2\theta)p(\alpha,\beta;\Delta)\\&\quad=2\theta\sum_{k=-\infty}^{\infty}p(\alpha,\beta;\Delta-k)u_k\\&\qquad+m_\alpha p(\beta,\beta;\Delta)+m_\beta p(\alpha,\alpha;\Delta),\\&(q_\alpha^{-1}+m_\alpha+\theta)p(\alpha,\alpha;\Delta)\\&\quad=q_\alpha^{-1}\delta_{\Delta,0}+\theta\sum_{k=-\infty}^{\infty}p(\alpha,\alpha;\Delta-k)u_k\\&\qquad+m_\alpha p(\alpha,\beta;\Delta).\end{aligned}\tag{3.1}$$

Denote $p^*_{\alpha\beta}(\xi)$ as the Fourier transform of $p(\alpha,\beta;d)$, and $\theta^*=\theta(1-u^*(\xi))$. Then from (3.1), omitting the argument ξ for ease of notation,

$$\begin{aligned}&(m_\alpha+m_\beta+2\theta^*)p^*_{\alpha\beta}=m_\alpha p^*_{\beta\beta}+m_\beta p^*_{\alpha\alpha},\\&(q_\alpha^{-1}+m_\alpha+\theta^*)p^*_{\alpha\alpha}=q_\alpha^{-1}+m_\alpha p^*_{\alpha\beta},\\&(q_\beta^{-1}+m_\beta+\theta^*)p^*_{\beta\beta}=q_\beta^{-1}+m_\beta p^*_{\alpha\beta}.\end{aligned}$$

The solution of (3.2) is

$$p^*_{\alpha\beta}=\frac{A}{B},\tag{3.2}$$

where

$$\begin{aligned}A&=q_\beta^{-1}m_\alpha(q_\alpha^{-1}+m_\alpha+\theta^*)+q_\alpha^{-1}m_\beta(q_\beta^{-1}+m_\beta+\theta^*),\\B&=(m_\alpha+m_\beta+2\theta^*)(q_\alpha^{-1}+m_\alpha+\theta^*)(q_\beta^{-1}+m_\beta+\theta^*)\\&\quad-m_\alpha m_\beta(q_\alpha^{-1}+m_\alpha+q_\beta^{-1}+m_\beta+2\theta^*).\end{aligned}$$

$p^*_{\alpha,\alpha}$ and $p^*_{\beta,\beta}$ can then be found from the second equation of (3.2). Probabilities are found by inversion of the corresponding transform, with

$$p(\alpha,\beta;d)=\frac{1}{2\pi}\int_{-\pi}^{\pi}e^{-\imath d\xi}p^*_{\alpha\beta}(\xi)\,d\xi.\tag{3.3}$$

The general case of more than two subpopulations is similar, with a system of equations to solve for $\alpha\neq\beta$, $\alpha,\beta=1,\ldots,g$ of

$$\begin{aligned}&(m_\alpha+m_\beta+2\theta^*)p^*_{\alpha\beta}=\sum_{\gamma\neq\alpha}m_{\alpha,\gamma}p^*_{\gamma\beta}+\sum_{\gamma\neq\beta}m_{\beta,\gamma}p^*_{\gamma\alpha},\\&(q_\alpha^{-1}+m_\alpha+\theta^*)p^*_{\alpha\alpha}=q_\alpha^{-1}+\sum_{\gamma\neq\alpha}m_{\alpha\gamma}p^*_{\alpha\gamma}.\end{aligned}\tag{3.4}$$

In a symmetric one-step mutation model where $u_{-1}=u_{+1}=1/2$, $u_j=0$ if $j\neq\pm1$, $u^*(\xi)=\cos(\xi)$, and $\theta^*=\theta(1-\cos(\xi))$.

3.2. Importance sampling distribution π

3.2.1. Single population

The system of equations (2.7) in a stepwise mutation model becomes

$$\widehat{\pi}(j\mid\boldsymbol{n})=\frac{n_j}{n+\theta}+\frac{\theta}{n+\theta}\sum_{i=-\infty}^{\infty}u_{j-i}\widehat{\pi}(i\mid\boldsymbol{n}).\tag{3.5}$$

The Fourier transform equation corresponding to (3.5) is

$$\widehat{\pi}^*(\xi\mid\boldsymbol{n})=\frac{n^*(\xi)}{n+\theta}+\frac{\theta}{n+\theta}u^*(\xi)\widehat{\pi}^*(\xi\mid\boldsymbol{n}).\tag{3.6}$$

Thus

$$\widehat{\pi}^*(\xi\mid\boldsymbol{n})=\frac{n^*(\xi)}{n+\theta(1-u^*(\xi))}.\tag{3.7}$$

Inverting the transform

$$\widehat{\pi}(k\mid\boldsymbol{n})=\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{e^{-k\xi\imath}\cdot n^*(\xi)}{n+\theta(1-u^*(\xi))}\,d\xi.\tag{3.8}$$

In a symmetric one-step mutation model there is an explicit solution

$$\begin{aligned}\widehat{\pi}(k\mid\boldsymbol{n})&=\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{e^{-k\xi\imath}\cdot n^*(\xi)}{n+\theta(1-\cos(\xi))}\,d\xi\\&=\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{\sum_{j=-\infty}^{\infty}n_j\cos((j-k)\xi)}{n+\theta(1-\cos(\xi))}\,d\xi\\&=\sum_{j=-\infty}^{\infty}\frac{n_j}{n+\theta}c_{j-k}(\rho)\\&=\frac{n_k}{n+\theta}c_0(\rho)+\sum_{\ell=1}^{\infty}\frac{n_{k+\ell}+n_{k-\ell}}{n+\theta}c_\ell(\rho),\end{aligned}\tag{3.9}$$

where $\rho=\theta/(n+\theta)$, and

$$\begin{aligned}c_\ell(\rho)&=\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{\cos(\ell\xi)}{1-\rho\cos(\xi)}\,d\xi\\&=\frac{1}{\sqrt{1-\rho^2}}\cdot\left[\frac{\rho}{1+\sqrt{1-\rho^2}}\right]^{|\ell|}.\end{aligned}\tag{3.10}$$

In the derivation of (3.9) if $j\neq k$,

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{\sum_{j=-\infty}^{\infty}n_j\,\sin((j-k)\xi)}{n+\theta(1-\cos(\xi))}\,d\xi=0$$

because the integrand is an odd function of ξ about zero.

3.2.2. Subdivided populations

Taking the Fourier transform of (2.11) produces a system of equations for $\alpha\in S$ of

$$\begin{aligned}&(n_\alpha^{-1}q_\alpha^{-1}+m_\alpha+\theta(1-u^*(\xi)))\widehat{\pi}^*(\xi\mid\alpha,\boldsymbol{n})\\&\quad=n^*_\alpha q_\alpha^{-1}+\sum_{\beta\neq\alpha}m_{\alpha\beta}\widehat{\pi}^*(\xi\mid\beta,\boldsymbol{n}).\end{aligned}\tag{3.11}$$

Let $A(\xi)=\mathrm{Diag}\big(n_\alpha q_\alpha^{-1}+m_\alpha+\theta(1-u^*(\xi))\big)-M$, $b(\xi)=(n^*_\alpha q_\alpha^{-1})$ and $\widehat{\pi}^*(\xi\mid\boldsymbol{n})=(\widehat{\pi}(\xi\mid\alpha,\boldsymbol{n}))$. The system (3.11)

can be written in the matrix form

$$\widehat{\pi}(\xi \mid \boldsymbol{n}) = A(\xi)^{-1} b(\xi)$$

with a solution of

$$\widehat{\pi}(j \mid \alpha, \boldsymbol{n}) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-j\xi_l} A(\xi)^{-1} b(\xi)\, d\xi. \tag{3.12}$$

An approximate solution to these equations can be found by simple numerical integration on equally spaced points in $[-\pi, \pi]$.

3.2.3. Two subpopulations

The general stochastic representation in Section 2 gives an expression in a model with two subpopulations labelled α and β of

$$\widehat{\pi}(j \mid \alpha, \boldsymbol{n}) = \sum_{k=-\infty}^{\infty} \sum_{v_\alpha=1}^{\infty} \sum_{v_\beta=0}^{\infty} \left[\frac{n_{\alpha k}}{n_\alpha}(1-\phi_\alpha) + \frac{n_{\beta k}}{n_\beta}(1-\phi_\beta) \right] \times \phi_\alpha^{v_\alpha - 1} \phi_\beta^{v_\beta} [P_\alpha^{v_\alpha} P_\beta^{v_\beta}]_{kj}. \tag{3.13}$$

The probability of v_α, v_β visits to α, β starting at α and ending at α is $\phi_\alpha^{v_\alpha - 1}(1-\phi_\alpha)\phi_\beta^{v_\beta}$, and similarly $\phi_\alpha^{v_\alpha - 1}\phi_\beta^{v_\beta}(1-\phi_\beta)$, for ending in β. The transition matrices P_α, P_β commute, so the term containing them in (3.13) is well defined. Denote the mean quantities

$$\widehat{\mu}(\alpha, \boldsymbol{n}) = \sum_{j=-\infty}^{\infty} j\widehat{\pi}(j \mid \alpha, \boldsymbol{n}), \quad \bar{n}_\alpha = \sum_{j=-\infty}^{\infty} j \frac{n_{\alpha j}}{n_\alpha},$$

$$\widehat{\mu}(\beta, \boldsymbol{n}) = \sum_{j=-\infty}^{\infty} j\widehat{\pi}(j \mid \beta, \boldsymbol{n}), \quad \bar{n}_\beta = \sum_{j=-\infty}^{\infty} j \frac{n_{\beta j}}{n_\beta}.$$

If the mean mutation distance $\sum_{j=-\infty}^{\infty} jP_{ij}$ is always zero from any position i then by considering the first step in the migration walk

$$\begin{aligned} \widehat{\mu}(\alpha, \boldsymbol{n}) &= \bar{n}_\alpha(1-\phi_\alpha) + \phi_\alpha \widehat{\mu}(\beta, \boldsymbol{n}), \\ \widehat{\mu}(\beta, \boldsymbol{n}) &= \bar{n}_\beta(1-\phi_\beta) + \phi_\beta \widehat{\mu}(\alpha, \boldsymbol{n}). \end{aligned} \tag{3.14}$$

The solution of (3.14) is

$$\begin{aligned} \widehat{\mu}(\alpha, \boldsymbol{n}) &= \frac{\bar{n}_\alpha(1-\phi_\alpha) + \phi_\alpha(1-\phi_\beta)\bar{n}_\beta}{1-\phi_\alpha\phi_\beta}, \\ \widehat{\mu}(\beta, \boldsymbol{n}) &= \frac{\bar{n}_\beta(1-\phi_\beta) + \phi_\beta(1-\phi_\alpha)\bar{n}_\alpha}{1-\phi_\alpha\phi_\beta}. \end{aligned} \tag{3.15}$$

The coefficients of $\bar{n}_\alpha, \bar{n}_\beta$ in (3.15) are the probabilities that the parent gene of the gene chosen comes from subpopulations α, β.

The representation (3.13) is transparent to understand, however it is easier to find a detailed solution directly from (3.12). We assume a symmetric one-step mutation model. Then the determinant of $A(\xi)$, $|A(\xi)| = (n_\alpha q_\alpha^{-1} + m_\alpha + \phi)(n_\beta q_\beta^{-1} + m_\beta + \phi) - m_\alpha m_\beta$, where $\phi = \theta(1-\cos(\xi))$. $|A(\xi)| = (\phi - \lambda_1) \times (\phi - \lambda_2)$, where λ_1, λ_2 are the roots of $\phi^2 + \phi(x_\alpha + x_\beta) + x_\alpha x_\beta - m_\alpha m_\beta = 0$, with $x_\alpha = n_\alpha q_\alpha^{-1} + m_\alpha$, and similarly for x_β. Both roots are real, less than zero, and unequal. Consider the expression

$$\begin{aligned} |A(\xi)|^{-1} &= \frac{1}{\lambda_1 - \lambda_2} \cdot \left[\frac{1}{\phi - \lambda_1} - \frac{1}{\phi - \lambda_2} \right] \\ &= \frac{1}{\lambda_1 - \lambda_2} \left[\frac{1}{-\lambda_1 + \theta} \cdot \frac{1}{1 - \rho_1 \cos(\xi)} \right. \\ &\quad \left. - \frac{1}{-\lambda_2 + \theta} \cdot \frac{1}{1 - \rho_2 \cos(\xi)} \right], \end{aligned} \tag{3.16}$$

where $\rho_i = \theta/(-\lambda_i + \theta), i = 1, 2$. From the matrix equation for the inversion of the transform and the form of the inverse of $A(\xi)$,

$$\begin{aligned} \widehat{\pi}(j \mid \alpha, \boldsymbol{n}) = \int_{-\pi}^{\pi} e^{-\iota\xi j} |A(\xi)|^{-1} \Big[& (n_\beta q_\beta^{-1} + m_\beta \\ & + \theta(1-\cos(\xi)) q_\alpha^{-1} n_\alpha^* + m_\alpha q_\beta^{-1} n_\beta^* \Big] d\xi. \end{aligned} \tag{3.17}$$

From Eqs. (3.17) and (3.16)

$$\widehat{\pi}(j \mid \alpha, \boldsymbol{n}) = \frac{1}{\lambda_1 - \lambda_2} \cdot \sum_{k=-\infty}^{\infty} \left[\frac{a_1(k,j)}{-\lambda_1 + \theta} - \frac{a_2(k,j)}{-\lambda_2 + \theta} \right], \tag{3.18}$$

where for $i = 1, 2$

$$\begin{aligned} a_i(k,j) = q_\alpha^{-1} n_{\alpha k} \Big[& (n_\beta q_\beta^{-1} + m_\beta + \theta) c_{k-j}(\rho_i) \\ & - \frac{\theta}{2} (c_{k-j+1}(\rho_i) + c_{k-j-1}(\rho_i)) \Big] \\ & + q_\beta^{-1} n_{\beta k} m_\alpha c_{k-j}(\rho_i). \end{aligned} \tag{3.19}$$

The summation in (3.18) is finite because only a finite number of $\{n_{\alpha k}, n_{\beta k}\}$ are non-zero.

3.2.4. Computer implementation of the IS algorithm

An implementation of the IS algorithm for two subpopulations (labelled 1 and 2) was based on the formula (3.18) for $\widehat{\pi}$. Using the example of locus DB4 in Table 2 with $\theta = 2.0$, $m_{12} = 5$, and $m_{21} = 3$, and relative subpopulation sizes of (0.25, 0.75) five duplicate computations of the likelihood of the sample and TMRCA were made, each with one million runs and different starting seeds. The accuracy is quite good, with likelihood values $4.2, 4.4, 4.5, 4.2, 4.3$ times 10^{-18} and mean TMRCA values in coalescent units of $3.219, 3.218, 3.209, 3.217, 3.217$. The time taken for a million runs is approximately 25 min on a 2.4 Ghz Pentium 4 computer.

3.2.5. Simulation study

A simulation study was undertaken using the IS algorithm for maximum likelihood estimation of mutation and migration rates. The model has two subpopulations with symmetric migration and a one-step mutation model. Ten data sets of 5 independent loci and 10 data sets of 20 independent loci were used to check the accuracy of estimates, and to see the effect of the number of loci on parameter estimation. Because of

computational constraints the study was limited. Likelihood estimation was done on computers of the Centre Informatique National de l'Enseignement Supérieur (CINES, France), occupying roughly 10 000 h on 500 MHz processors. Simulated data sets were generated using a discrete generation coalescent process in a two population model with equal sizes $N_1 = N_2 = 1000$ genes, mutation rate $\theta = 4.0$, and symmetric migration rates $\gamma = m_{12} = m_{21} = 4.0$. Leblois et al. (2003) has details about the discrete generation simulation technique. Likelihoods were computed for each locus using the IS algorithm described in this paper. Likelihoods for each locus were then multiplied to find the overall likelihood for the multilocus data set at each parameter point. Finally, maximum likelihood estimates were computed from the likelihood surface obtained from the likelihood at the different parameter points by kriging as described in the Appendix A. A computation based on 5000 runs at 50 values of the parameter vector yielded less accurate estimates of likelihood than a computation based on 500,000 runs at the same 50 values, but the maximum likelihood estimates were identical (details not shown). Thus, accurate estimation of the likelihood for each point is not essential. Further computations were based on 5000 runs at 5000 points. For the two cases of 5 and 20 loci, the relative mean bias and relative mean square error (MSE) were calculated for estimates of the migration rate and mutation rate. Since migration was symmetric in the model, the two migration rate estimates were pooled.

The relative bias (bias/expectation) and relative MSE (MSE/squared expectation) are presented in Table 2. They show good performance of the algorithm, in particular in estimating θ. As expected, the precision increases with the number of loci for estimation of both mutation and migration parameters. In both cases of 5 and 20 loci the bias and the MSE are very small (below 10%). On the other hand, our results show a slight overestimation of the migration parameter γ. Nevertheless, the MSE is not very high indicating that estimates are close to parameter values in the model.

Part of the variance of the estimator is due to the variance of the maximum likelihood estimator, and part of it is due to the imprecision in locating the maximum of the likelihood function through sampling points and smoothing the surface. Efficiency of the latter procedure is demonstrated if it contributes only a small part of the total variance. In this case, the correlation between independent applications of the procedure to the same data set should be high. Thus, we independently applied this procedure (hypercube sampling, likelihood evaluation, and kriging) twice to the ten 5-loci data sets. The correlation for pairs of estimates was >0.975 for all parameters, demonstrating the efficiency of the procedure.

The discrete generation simulation model differed slightly from the continuous time coalescent model under which estimation was based. The limited number of intervals used in the likelihood grid, or the simulation model, may account for the slight bias of migration estimates.

For comparison `Migrate` (Beerli and Felsenstein, 2001) was used with default settings to analyze the same simulated data sets. Using the default settings, `Migrate` takes about the same computer time as our algorithm to analyze the data. In the study `Migrate` estimates were not precise with a low number of loci. Precision in estimating γ, but not θ, increased with the number of loci (Table 3). Note that estimation with `Migrate` assumed two different parameters θ_1 and θ_2, instead of a single θ parameter for both populations as assumed in the previous estimations.

3.2.6. *Microsatellite application*

Microsatellite loci are highly variable and the presence of back mutations cannot be ignored. Therefore inference from such data can be challenging and likelihood evaluation extremely computer intensive. The algorithm described in this paper constitutes a significant improvement over previous methods. Nielsen (1997) developed an algorithm based on the pioneering work of Griffiths and Tavaré (1994) to obtain maximum likelihood estimates of the mutation parameter θ at microsatellite loci. The method is computationally inefficient, even for a single locus, and computational time might be large, as many runs through the Markov chain do not contribute anything to the likelihood value. In fact, only a few simulated genealogies will contribute significantly to the likelihood evaluation, while most of

Table 2
Bias and precision of estimates from simulated data sets using the algorithm described in this paper

		5 loci	20 loci
θ	Bias	0.16	0.03
θ	MSE	0.09	0.01
γ	Bias	0.19	−0.05
γ	MSE	0.55	0.22

Table 3
Bias and precision of estimates from simulated data sets using `Migrate`

		5 loci	20 loci
θ	Bias	0.23	−0.60
θ	MSE	0.30	0.40
γ	Bias	1.2	0.25
γ	MSE	3.1	0.46

the computational effort will be spent on genealogies with very small probability, i.e. not consistent with the observed data. This problem is especially evident in runs in which too many mutation events occur. Nielsen (1997) proposed truncating such runs according to a rule based on the expected number of mutations. The IS proposal described in this paper implicitly solves this problem, concentrating most of the computational effort on trees with high probability, given the observed data. Of course the extension to subdivided populations of the island model type here is new. Stephens and Donnelly (2000) make a comparison on example data sets of their technique with their implementation of the Griffiths–Tavaré technique, and a comparison with `Batwing`, an implementation of a Bayesian technique of Wilson and Balding (1998). The Stephens–Donnelly technique is the most efficient of the three. Their microsatellite state space is truncated, so their distributions $\widehat{\pi}$ are computed from a system of linear equations rather than using (3.9). Chen and Liu (2000) show that in a data example the Griffiths–Tavaré technique combined with path resampling is as efficient as the Stephens–Donnelly technique and produces the same likelihood curves (see also Chen et al. (2005)). Path resampling could be used generally in other IS schemes such as in this paper.

3.2.7. Red Fox data example

Lade et al. (1996) collected data on seven microsatellite loci from Australian populations of the red fox *Vulpes vulpes*. As an example the two subpopulations from Phillip Island (PI) and the adjacent mainland at San Remo (SR) separated by a bridge are considered here. The data, coded so the minimum position is 0 and two base pairs are taken as a one unit mutation step, are shown in Table 4. A stepwise mutation model with single steps was used to model the data. This implies that two alleles that differ by one mutation step are more closely related than alleles that differ by many mutation steps. The stepwise model is a possible model for microsatellites when interest is in the relatedness between individuals and in population substructuring. Evolution at the different loci was assumed to be independent with the same mutation rate θ and two migration rates between the populations. Likelihoods for each locus were computed using the IS algorithm described in this paper, then multiplied to find the overall likelihood for the loci. Maximum likelihood estimates of the mutation and migration rates were found by fitting a likelihood surface using the kriging method described in the Appendix A with 1000 design points. Raw likelihood points were based on runs of 10,000 replicates. Parameter estimates and likelihoods are shown in Table 5 for three different population size ratios PI:SR. The ratio with maximum likelihood is $(0.25, 0.75)$. There is a large difference between the likelihoods for sizes $(0.25, 0.75)$ and $(0.75, 0.25)$; the former is consistent with the higher diversity in SR, and the latter is unlikely. A maximum likelihood estimate is $\widehat{\theta} = 2$ with an approximate backward migration rate estimate from PI to SR of $\widehat{m}_{PS} = 5$ and in the other direction $\widehat{m}_{SP} = 3$. Forward migration rates are $\widetilde{m}_{SP} = 1.6$ and $\widetilde{m}_{PS} = 9$. A likelihood surface for the migration rates when $\theta = 2.0$ is shown in Fig. 2. Likelihood units are 10^{-89}. Raw likelihood points were based on runs of one million replicates. Pairwise distributions of the number of steps between two genes both from PI, both

Table 4
Microsatellite allele frequencies from the red fox populations of PI and SR

Locus	PI	SR
DB1	$n = 46$	$n = 42$
0		3
1		5
2	32	2
4		10
5		1
7	14	11
9		10
DB3	$n = 46$	$n = 46$
0	38	32
1	8	11
5		3
DB4	$n = 46$	$n = 44$
0	8	20
9	38	14
12		4
14		5
15		1
DB6	$n = 46$	$n = 42$
0	1	2
1	27	27
2	18	13
OB	$n = 46$	$n = 44$
0	28	18
6	8	16
8	10	10
VD10	$n = 42$	$n = 42$
0	5	5
2		13
4	12	12
6	25	12
C213	$n = 46$	$n = 44$
0	4	15
2	42	24
3		5

from SR, or one each from PI and SR are computed from (3.3) and shown in Table 6. The square root of average pairwise difference squares within and between populations is shown in Table 7. There is quite a variation between loci and variation from the theoretical values.

As an illustration the mean and standard deviation of the TMRCA at each locus, shown in Table 8, was computed with the estimated parameters conditional on the stepwise allele configuration observed. If the mutation rate per locus per generation is 10^{-3}, then the effective population size is $N = \theta/(2 \times 10^{-3}) = 1000$, with effective population sizes on PI of 250 and SR of 750. Effective population sizes approximate true population sizes over time by their harmonic means. Supposing a generation time for foxes of five years (Lade et al., 1996), then coalescent time units are in units of 5000 years. The TMRCA of the loci are well before the introduction of foxes into Australia around 1870 (Lade et al., 1996), which is what is expected. Assuming the model holds over the full ancestry may have the effect of inflating

Table 5
Likelihood estimates

(q_1, q_2)	Likelihood $\times 10^{-89}$	$\widehat{\theta}$	$\widehat{m}_{PS}$	$\widehat{m}_{SP}$
(0.5, 0.5)	43×10^3	1.96	2.07	3.92
(0.25, 0.75)	62×10^5	2.00	5.07	3.27
(0.75, 0.25)	64	1.70	1.79	14.6

Table 6
Distance distribution between two genes

$\lvert\Delta\rvert$	PI	SR	PI SR
0	0.565	0.447	0.388
1	0.276	0.333	0.362
2	0.096	0.132	0.150
3	0.038	0.053	0.060
4	0.015	0.021	0.024
5	0.006	0.008	0.010
6	0.002	0.003	0.004
7	0.001	0.001	0.002
8	0.000	0.001	0.000

Table 7
Square root of average square pairwise differences

Locus	PI	SR	PI SR
DB1	3.29	4.33	4.18
DB3	0.54	1.78	1.36
DB4	4.88	8.04	6.76
DB6	0.75	0.77	0.76
OB	5.04	4.88	5.05
VD10	2.78	2.87	3.05
C213	0.81	1.54	1.28
Theory	1.23	1.45	1.54

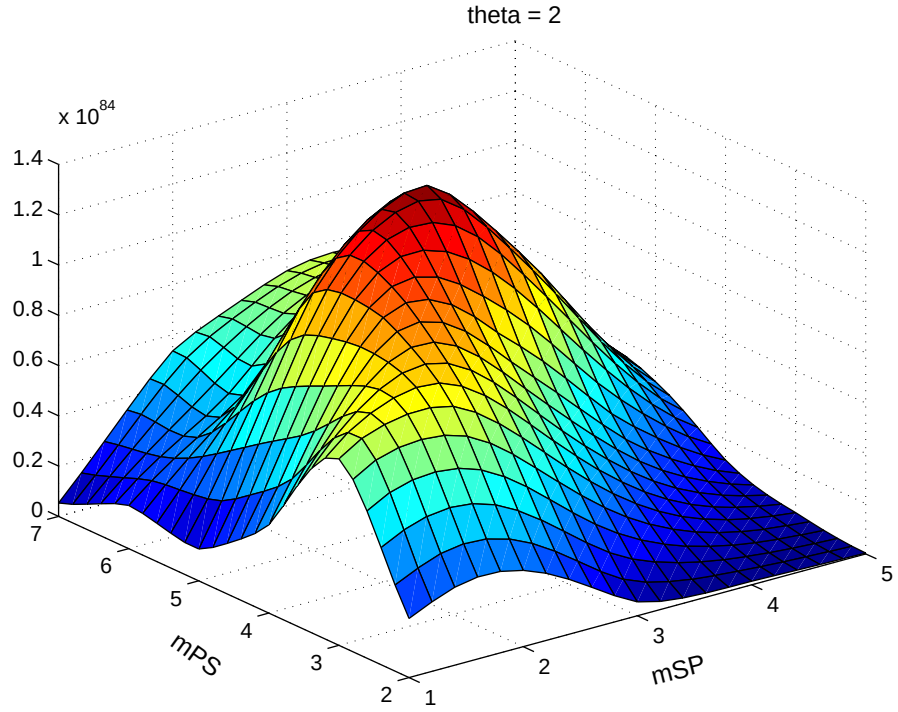

Fig. 2. Likelihood surface for migration rates.

Table 8
TMRCA in years

Locus	mean	sd
DB1	9100	3700
DB3	8200	3600
DB4	16000	5100
DB6	4600	2500
OB	11200	4200
VD10	7700	3400
C213	5500	2800

migration rates, by explaining allele frequency differences by migration, rather than by drift.

Acknowledgements

MDI, RCG were supported by BBSRC Bioinformatics grant 43/BIO14435; RL, FR were supported by grants AIP Biodiversité INRA; CNRS, Programme *Impact des Biotechnologies dans les agro-cosystmes* to I. Olivieri and A. Mignot; and a ACI Jeune chercheur grant supported FR. The author list is alphabetical.

Appendix A

A.1. Likelihood surface

The IS algorithm proposed in this paper allows computation of the likelihood with respect to parameters of interest. We have a major interest in estimating migration and mutation parameters and therefore in estimating the likelihood surface. There are two aspects of computing the likelihood surface. The first is pointwise computation of likelihoods; the second is to find a way to interpolate the surface locally about the points already computed taking into account that the surface is a random realization of the true surface. Potentially there are a number of different methods for parameter estimation: IS with respect to a proposal distribution containing the parameter distribution (Griffiths and Tavaré, 1994; Stephens and Donnelly, 2000); bridge sampling (Fearnhead and Donnelly, 2001); Markov chain Monte Carlo schemes (Beerli and Felsenstein, 1999); and Bayesian methods (Wilson and Balding, 1998). Beaumont et al. (2002) propose a method for approximate Bayesian statistical inference based on summary statistics. Their approach replaces the full data with suitable summary statistics and approximates the posterior density of the parameters of interest using kernel density estimation techniques. The use of summary statistics allows for increased computational efficiency, although it does not make use of all the information in the data, as typically in many settings sufficient statistics are not available.

Here we have adopted a different approach based on a local linear predictor for interpolated points on the computed points which takes into account random variation of the surface. The likelihood of the full sample of chromosomes is evaluated at points in a design set and then the likelihood surface is estimated using kriging methods. Let $y(\boldsymbol{\theta})$ denote the likelihood function of a sample of n genes and let $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d)$ be the vector of parameters of interest (for example, mutation and migration rates), where $\boldsymbol{\theta} \in D \subset \mathbb{R}^d$, $d \geqslant 1$. Suppose we have evaluated the likelihood at a set of N parameter values $\{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_N\}$. Each likelihood evaluation can be extremely computer intensive and can require a long running time. We are interested in estimating the likelihood surface, that is, to predict $y(\boldsymbol{\theta})$ at any $\boldsymbol{\theta} \in D$ given the values $\{y(\boldsymbol{\theta}_1), \ldots, y(\boldsymbol{\theta}_N)\}$. The unknown function $y(\boldsymbol{\theta})$ is assumed to be a realization of a Gaussian process $Y = \{Y(\boldsymbol{\theta}), \boldsymbol{\theta} \in D\}$ (see Ripley (1981)). A Gaussian process is defined by its mean function and covariance function:

$$\mathbb{E}[Y(\boldsymbol{\theta})] = \mu(\boldsymbol{\theta}), \tag{A.1}$$

$$\operatorname{cov}[Y(\boldsymbol{\theta}_i), Y(\boldsymbol{\theta}_j)] = C(\boldsymbol{\theta}_i, \boldsymbol{\theta}_j). \tag{A.2}$$

Moreover, normality of the finite-dimensional distributions is assumed, that is for every finite subset of points $S = \{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_k\} \subset D$, the joint distribution of $(Y(\boldsymbol{s}_1), \ldots, Y(\boldsymbol{s}_k))$ is a multivariate normal. In a situation where the likelihood is computed from independent sampling realizations, the multivariate normal assumption will hold approximately because of the central limit theorem, but the assumption is not critical because prediction of interpolated points in the local surface explained below really only depends on best prediction with minimum variance consideration.

Given the value of the likelihood at $\{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_N\}$, we wish to predict the likelihood surface at a new point $\widetilde{\boldsymbol{\theta}}$. Let y_N be the column vector defined as

$$y_N = \begin{pmatrix} y(\boldsymbol{\theta}_1) \\ \vdots \\ y(\boldsymbol{\theta}_N) \end{pmatrix}.$$

Then the minimum MSE unbiased predictor of $y(\widetilde{\boldsymbol{\theta}})$ is given by

$$\widehat{y}(\widetilde{\boldsymbol{\theta}}) = \mathbb{E}(Y(\widetilde{\boldsymbol{\theta}}) \mid y_N) = \mu(\widetilde{\boldsymbol{\theta}}) + k(\widetilde{\boldsymbol{\theta}})K^{-1}(y_N - \mu_N) \tag{A.3}$$

and its variance is

$$\operatorname{var}(\widehat{y}(\widetilde{\boldsymbol{\theta}})) = \operatorname{var}(Y(\widetilde{\boldsymbol{\theta}}) \mid y_N) = C(\widetilde{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\theta}}) - k(\widetilde{\boldsymbol{\theta}})K^{-1}k(\widetilde{\boldsymbol{\theta}})', \tag{A.4}$$

where

$$\mu_N = \begin{pmatrix} \mu(\boldsymbol{\theta}_1) \\ \vdots \\ \mu(\boldsymbol{\theta}_N) \end{pmatrix}, \quad k(\widetilde{\boldsymbol{\theta}}) = \begin{pmatrix} C(\widetilde{\boldsymbol{\theta}}, \boldsymbol{\theta}_1) \\ \vdots \\ C(\widetilde{\boldsymbol{\theta}}, \boldsymbol{\theta}_N) \end{pmatrix}$$

and $K = (K_{ij})$ is $N \times N$ matrix whose elements are given by

$$K_{ij} = C(\boldsymbol{\theta}_i, \boldsymbol{\theta}_j).$$

The linear prediction in (A.3)–(A.4) is known in geostatistics and spatial statistics as kriging (Ripley, 1981; Cressie, 1993; Stein, 1999) and has previously been applied to surface estimation in different contexts, including interpolation, image restoration (Geman and Geman, 1984; Ripley, 1988), prediction of deterministic functions (Currin et al., 1991) and analysis of computer experiments (Currin et al., 1991; Sacks et al., 1989a,b).

We are looking for a general method to estimate the likelihood surface. The surface is assumed to be smooth, continuous and differentiable. In particular, we require that the mean is constant for all $\boldsymbol{\theta} \in D$, i.e. $\mu(\boldsymbol{\theta}) = \mu$, for all $\boldsymbol{\theta} \in D$. As covariance function we use

$$C(\boldsymbol{\theta}_i, \boldsymbol{\theta}_j) = \sigma^2 \exp\left\{ -\lambda \sum_{k=1}^{d} (\theta_{ik} - \theta_{jk})^2 \right\}. \tag{A.5}$$

where $\lambda > 0$ determines the correlation structure of Y and σ^2 is a scale factor. Of course, other choices of mean and covariance functions are possible. See, for example, Currin et al. (1991), Stein (1999). For example, we could incorporate a linear (or polynomial) model for Y through the mean function. In our experience there is no need for this, especially when d is quite large (greater than 4), as predictions based on a constant mean function are quite good. Determination of μ, λ and σ^2 is usually achieved by iterative search methods (Ripley, 1988; Mardia and Marshall, 1984).

The last issue we need to address is the choice of the N values of the parameter vector $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d)$ at which to evaluate the likelihood. Ideally we would like all the areas of the space D to be represented. We have applied *Latin Hypercube Sampling* to determine $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_N$ (McKay et al., 1979). Briefly, *Latin Hypercube Sampling* consists of dividing the range of each θ_k, $k = 1, \ldots, d$, into N intervals of equal marginal probability $1/N$ and then sampling a point at random from each interval. In this way a sample $\theta_{kj}, j = 1, \ldots, N$ is obtained, and these sampled values form the kth component in $\boldsymbol{\theta}_i$, $i = 1, \ldots, N$. The components of the various $\boldsymbol{\theta}_i$ are then matched at random. Therefore, there are N intervals on the range of each element of $\boldsymbol{\theta}_i$ and they combine to form N^d cells which cover the space D. The bigger N is, the better the estimation of the likelihood surface is. Of course, the computational cost increases.

The results presented in this paper have been obtained using the `R` package `fields` (`http://www.cgd.ucar.edu/stats/Software/Fields/index.shtml`). This package contains a collection of functions for curve and function fitting with an emphasis on spatial data. In particular, the function `krig` implements spatial process estimates through kriging. The `R` web site is `http://www.R-project.org` where information and software is available, and the `R` manual is *R: A language and environment for statistical computing* (R Development Core Team, 2004).

References

Bahlo, M., Griffiths, R.C., 2001. Coalescence time for two genes from a subdivided population. J. Math. Biol. 43, 397–410.

Beaumont, M.A., Zhang, W., Balding, D.J., 2002. Approximate Bayesian computation in population genetics. Genetics 162, 2025–2035.

Beerli, P., Felsenstein, J., 1999. Maximum likelihood estimation of migration rates and effective population numbers in two populations using a coalescent approach. Genetics 152, 763–773.

Beerli, P., Felsenstein, J., 2001. Maximum likelihood estimation of migration matrix and effective population size in n subpopulations by using a coalescent approach. PNAS 98, 4563–4568.

Chen, Y., Liu, J.S., 2000. Discussion on the paper of Stephens and Donnelly (2000). J. Roy. Statist. Soc. B 62, 644–645.

Chen, Y., Xie, J., Liu, J.S., 2005. Stopping-time resampling for sequential Monte Carlo methods. J. Roy. Statist. Soc. B 67, 199–217.

Cressie, N.A.C., 1993. Statistics for Spatial Data. Wiley, New York.

Currin, C., Mitchell, T., Morris, M., Ylvisaker, D., 1991. Bayesian prediction of deterministic functions, with applications to the design and analysis of computer experiments. J. Amer. Statist. Assoc. 86, 953–963.

De Iorio, M., Griffiths, R.C., 2004a. Importance sampling on coalescent histories, I. Adv. Appl. Probab. 36, 417–433.

De Iorio, M., Griffiths, R.C., 2004b. Importance sampling on coalescent histories, II. Subdivided population models. Adv. Appl. Probab. 36, 434–454.

Excoffier, L., 2001. Analysis of population subdivision. In: Balding, D.J., Bishop, M., Cannings, C. (Eds.), Handbook of Statistical Genetics. Wiley, Chichester, pp. 271–307.

Fearnhead, P., Donnelly, P., 2001. Estimating recombination rates from population genetics data. Genetics 159, 1299–1318.

Geman, S., Geman, D., 1984. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. IEEE Trans. Pattern Anal. Mach. Intelli. 6, 721–741.

Griffiths, R.C., 2001. Ancestral inference from gene trees. In: Donnelly, P., Foley, R. (Eds.), Genes, Fossils, and Behaviour: an Integrated Approach to Human Evolution, NATO Science Series, Series A: Life Sciences, vol. 310. IOS Press, pp. 137–172.

Griffiths, R.C., Tavaré, S., 1994. Simulating probability distributions in the coalescent. Theoret. Popul. Biol. 46, 131–159.

Herbots, H.M., 1997. The structured coalescent. In: Donnelly, P., Tavaré, S. (Eds.), Progress in Population Genetics and Human Evolution. Springer, Berlin, 330pp, IMA Volumes in Mathematics and its Applications, vol. 87, pp. 231–255.

Kingman, J.F.C., 1982. The coalescent. Stochastic Processes Appl. 13, 235–248.

Lade, J.A., Murray, N.D., Marks, C.A., Robinson, N.A., 1996. Microsatellite differentiation between Philip Island and mainland Australian populations of the red fox Vulpes vulpes. Mol. Ecol. 5, 81–87.

Leblois, R., Estoup, A., Rousset, F., 2003. Influence of mutational and sampling factors on the estimation of demographic parameters in a continuous population under isolation by distance. Mol. Biol. Evol. 20, 491–502.

Mardia, K.V., Marshall, R.J., 1984. Maximum likelihood estimation of models for residual covariance in spatial regression. Biometrika 71, 135–146.

McKay, M.D., Conover, W.J., Beckman, R.J., 1979. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics 21, 239–245.

Moran, P.A.P., 1975. Wandering distributions and the electrophoretic profile, I. Theoret. Popul. Biol. 8, 318–330.

Moran, P.A.P., 1976. Wandering distributions and the electrophoretic profile, II. Theoret. Popul. Biol. 10, 145–149.

Nagylaki, T., 1983. The robustness of neutral models of geographical variation. Theoret. Popul. Biol. 24, 268–294.

Nielsen, R., 1997. A likelihood approach to population samples of microsatellite alleles. Genetics 146, 711–716.

Nordborg, M., 2001. Coalescent theory. In: Balding, D.J., Bishop, M., Cannings, C. (Eds.), Handbook of Statistical Genetics. Wiley, Chichester, pp. 179–208.

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

Ohta, T., Kimura, M., 1973. A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. Genet. Res. 22, 201–204.

R Development Core Team, 2004. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-00-3.

Ripley, B.D., 1981. Spatial Statistics. Wiley, New York.

Ripley, B.D., 1988. Statistical Inference for Spatial Processes. Cambridge University Press, Cambridge.

Rousset, F., 1996. Equilibrium values of measures of population subdivision for stepwise mutation processes. Genetics 142, 1357–1362.

Rousset, F., 2001. Inferences from spatial population genetics. In: Balding, D.J., Bishop, M., Cannings, C. (Eds.), Handbook of Statistical Genetics. Wiley, Chichester, pp. 239–269.

Sacks, J., Schiller, S.B., Welch, W.J., 1989a. Designs for computer experiments. Technometrics 31, 41–47.

Sacks, J., Welch, W.J., Mitchell, T.J., Wynn, H.P., 1989b. Design and analysis of computer experiments (with comments). Statist. Sci. 4, 409–435.

Stein, M.L., 1999. Interpolation of Spatial Data. Springer, New York.

Stephens, M., 2001. Inference under the coalescent. In: Balding, D.J., Bishop, M., Cannings, C. (Eds.), Handbook of Statistical Genetics. Wiley, Chichester, pp. 213–238.

Stephens, M., Donnelly, P., 2000. Inference in molecular population genetics. J. Roy. Statist. Soc. B 62, 605–655.

Wilson, I.J., Balding, D.J., 1998. Genealogical inference from microsatellite data. Genetics 150, 499–510.