Adv. Appl. Prob. **33,** 600–616 (2001)
Printed in Northern Ireland

POST-DATA INFERENCE OF COALESCENCE TIMES AND SEGREGATING-SITE DISTRIBUTION IN A TWO-ISLAND MODEL WITH SYMMETRIC MIGRATION

PRAKASH GORROOCHURN,* *University of Mauritius*

Abstract

In this paper, we present the distribution of the coalescence time of two DNA sequences (or genes) subject to symmetric migration between two islands, and conditional on the observed number of segregating sites in the sequences. The distribution for the segregating-site pattern is also obtained. Some surprising results emerge when both genes are initially on the same island. First, the post-data mean coalescence time is shown to be dependent on the migration parameter, as opposed to the pre-data mean. Second, both the post-data density and expectation for the coalescence time are shown to converge, in the weak-migration limit, to the corresponding panmictic results, as opposed to the pre-data situation where there is convergence in the density but not in the expectation. Finally, it is shown that there is convergence in the weak-migration limit in the distribution of the number of segregating sites but not in the expectation and variance. Numerical and graphical results for samples of size greater than two are also presented.

Keywords: Structured coalescent; segregating sites; two-island model; symmetric migration; monomorphic

AMS 2000 Subject Classification: Primary 92D25
Secondary 60J85

1. Introduction

Much of the recent work on genealogical processes in population genetics has drawn from the vast amounts of molecular genetic data that are rapidly becoming available. The data normally consist of homologous DNA sequences obtained from various contemporary humans (Griffiths and Tavaré (1994)). With the availability of these data, it has become important that we revise many of our genealogical inferences which have hitherto been based on pre-data analysis. However, recent post-data work has tended to concentrate exclusively or mainly on panmictic models of population structure (Slatkin and Hudson (1991), Tavaré *et al.* (1997)). See Marjoram and Donnelly (1994), Beerli and Felsenstein (1999) and Bahlo and Griffiths (2000), though, for some recent simulation-based post-data work in subdivided populations.

In this paper, we consider a two-island model of population structure with symmetric migration and make use of the coalescent in a population with geographic structure (Notohara (1990), Nath and Griffiths (1993), Herbots (1997), Wilkinson-Herbots (1998), Gorroochurn (1999); note that the different authors use different scalings for the rate matrix). The essential details of the model are given in Section 2.2 below. We obtain exact results for the distribution of the coalescence time of two genes, conditional on the observed number of segregating sites

Received 5 February 2001; revision received 12 June 2001.
* Postal address: Department of Mathematics, University of Mauritius, Réduit, Mauritius.
Email address: gorroocp@uom.ac.mu

between them. We also obtain the distribution for the segregating-site pattern and point out some anomalies when the two genes are initially on the same island.

2. The coalescent

2.1. Panmictic population

As first put forward by Kingman (1982a), (1982b), (1982c), the coalescent describes the genealogical process of a random sample of n genes drawn from a large haploid population of constant size $2N$. We assume a panmictic population structure. We can trace back the ancestry of the sample by considering a coalescent tree in which the leaves represent the sample gene (or DNA) sequences (see Figure 1). Time is measured continuously up the tree (i.e. backwards) in units of $2N$ generations. Coalescences are represented by the vertices of the tree, whereby two of its ancestral lines (branches) merge to form a single ancestral line. Starting from time $t = 0$, as we move up the tree (i.e. go back in time), more and more coalescences occur until the entire ancestry coalesces to a single individual, known as the most recent common ancestor (MRCA) of all the genes in the sample. The total time taken for this to happen is called the time to the most recent common ancestor (TMRCA). For reviews of the coalescent process, see Ewens (1989), Hudson (1991), Donnelly and Tavaré (1995) and Fu and Li (1998), (1999).

The times $T_{(j)}$ during which the sample has j distinct ancestors are independently distributed with an exponential distribution of mean $2/\{j(j-1)\}$, $j = 2, 3, \ldots$. This is equivalent to saying that, conditional on there being j ancestral lines at some time in the tree, coalescences occur according to a Poisson process of (total) rate $j(j-1)/2$. Also,

$$\mathrm{E(TMRCA)} = \sum_{j=2}^{n} \frac{2}{j(j-1)} = 2\left(1 - \frac{1}{n}\right).$$

Mutations can be superimposed on the coalescent tree. We assume no recombination and an infinitely-many-sites model (Watterson (1975)). If the probability of a (neutral) mutation per gene per generation is constant at u, then we set $\theta = 4Nu$. This means that the mutation rate per gene is $\theta/2$. Moreover, conditional on a branch of length l, the number of mutations on that branch has a Poisson distribution of rate $\theta l/2$, independently of other branches of the tree.

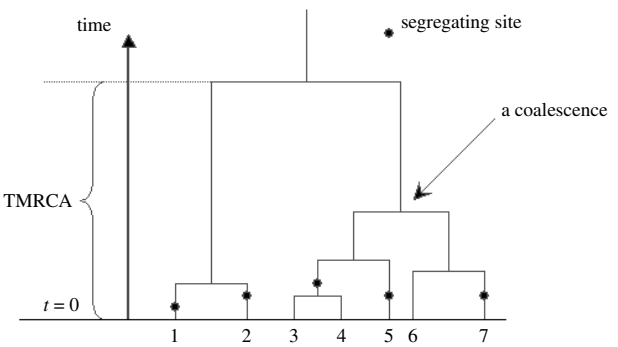

FIGURE 1: A typical coalescent tree with superimposed mutations. The numbers 1–7 denote the labelled genes in the sample. Time is measured up the tree (i.e. backwards). A coalescence is said to take place when two ancestral lines (branches) merge to form a single ancestral line as we move up the tree.

2.2. Subdivided population

In many situations an assumption of geographic structure is necessary and the coalescent is then approximated by the structured coalescent, as first put forward in its general form by Notohara (1990). The next paragraph is adapted from the latter. Further applications of the coalescent approach in subdivided populations include Takahata (1988), (1991), Notohara (1993), (1997), Nath and Griffiths (1993), Donnelly and Tavaré (1995), Herbots (1997), Wilkinson-Herbots (1998) and Gorroochurn (1999).

We consider a (haploid) population consisting of K colonies (or subpopulations) such that colony i contains $2c_iN$ genes, $i = 1, \ldots, K$. We define $\alpha_N^i(\tau)$ to stand for the number of haploid genes sampled without replacement from colony i at generation τ. The K-tuple $\boldsymbol{\alpha}_N(\tau) = (\alpha_N^1(\tau), \alpha_N^2(\tau), \ldots, \alpha_N^K(\tau))$ thus denotes the geographical composition of the sample at generation τ in the past and is called the ancestral process. With migration taking place between the colonies and Wright-type reproduction within each colony, and a re-scaling of time t in units of $2N$ generations (i.e. $\tau = \lfloor 2Nt \rfloor$, where $\lfloor x \rfloor$ denotes the greatest integer less than or equal to x), the ancestral process can be well approximated by the structured coalescent $\{\boldsymbol{\alpha}(t), P_{\boldsymbol{\alpha}}\}$. Here $\boldsymbol{\alpha}(t) = (\alpha^1(t), \alpha^2(t), \ldots, \alpha^K(t))$, where $\alpha^i(t)$ is the number of haploid genes in colony i, and the initial number of sampled genes is $|\boldsymbol{\alpha}(0)|$. We now introduce the backward migration matrix $\boldsymbol{M} = \{m_{i,j}/2\}$, where $m_{i,j}/2$ is the scaled backward migration rate such that the probability that a gene now in colony i actually came from colony j in the time interval Δt is $(m_{i,j}/2)\Delta t + o(\Delta t)$, where $\Delta t = 1/(2N)$ and $i \neq j$. Further, $\{\boldsymbol{\alpha}(t), P_{\boldsymbol{\alpha}}\}$ defines a continuous-time Markov chain with rates

$$q_{\boldsymbol{\alpha},\boldsymbol{\beta}} = \begin{cases} \dfrac{\alpha_i m_{i,j}}{2} & \text{if } \boldsymbol{\beta} = \boldsymbol{\alpha} - \boldsymbol{\varepsilon}^i + \boldsymbol{\varepsilon}^j,\ i \neq j; \\ \dfrac{\alpha_i(\alpha_i - 1)}{2c_i} & \text{if } \boldsymbol{\beta} = \boldsymbol{\alpha} - \boldsymbol{\varepsilon}^i; \\ -\left\{ \displaystyle\sum_{i \in S} \frac{\alpha_i(\alpha_i - 1)}{2c_i} + \sum_{i \in S} \frac{\alpha_i |m_{i,i}|}{2} \right\} & \text{if } \boldsymbol{\beta} = \boldsymbol{\alpha}; \\ 0 & \text{otherwise.} \end{cases} \tag{2.1}$$

Here, $\boldsymbol{\varepsilon}^i$ is the vector $\boldsymbol{\alpha}$ for which $\alpha_i = 1$ and $\alpha_j = 0,\ j \neq i$, that is, $\boldsymbol{\varepsilon}^i = (0, 0, \ldots, 1, \ldots, 0)$ with the 1 in the ith position. Also, since $\sum_{j=1}^K m_{i,j} = 0$, we have $m_{i,i} = -\sum_{j=1, j\neq i}^K m_{i,j} = -m_i$, where $m_i/2$ is the net emigration rate per gene from island i. In (2.1), a change of $\boldsymbol{\alpha}$ to $\boldsymbol{\beta} = \boldsymbol{\alpha} - \boldsymbol{\varepsilon}^i + \boldsymbol{\varepsilon}^j,\ i \neq j$, corresponds to a migration backward in time of an ancestral lineage from colony i to colony j; the rate of such a migration is then $\alpha_i m_{i,j}/2$. A change of $\boldsymbol{\alpha}$ to $\boldsymbol{\beta} = \boldsymbol{\alpha} - \boldsymbol{\varepsilon}^i$ corresponds to a coalescence between two lineages in colony i; the rate of such a coalescence is thus

$$\frac{1}{c_i}\binom{\alpha_i}{2}.$$

A coalescence between two lineages can occur if and only if those two lineages are in the same colony at a particular point in time (see Figure 2, below).

A special case of the above is the symmetric island model of population structure (Wright (1931)). In this model, we assume that all K subpopulations are of the same size $2N$ (i.e. $c_i = 1$) and the migration rate between any two subpopulations is the same, i.e. $m_i = m$ for $i = 1, \ldots, K$ and $m_{i,j} = m/(K-1)$ for $j \neq i$.

3. Post-data inference in a panmictic population

We first consider the single-island case with two genes. We let $T_{(2)}$ denote the TMRCA for the sample and $S_{(2)}$ the number of segregating sites in the sample. The conditional distribution of $T_{(2)}$ given $S_{(2)} = k$ has been obtained by Tajima (1983); with a suitable re-scaling, it can be written as

$$f_{T_{(2)} \mid S_{(2)}}\{t \mid S_{(2)} = k\} = \frac{(1+\theta)^{1+k}}{k!} t^k \mathrm{e}^{-(1+\theta)t}, \qquad t > 0, \tag{3.1}$$

which is Gamma$(1 + k, 1/(1 + \theta))$. We note, from (3.1), that

$$\mathrm{E}\{T_{(2)} \mid S_{(2)} = k\} = \frac{1+k}{1+\theta}. \tag{3.2}$$

This formula is reasonable intuitively, since if $k \gg \theta$ then this means that a lot more segregating sites have been observed than the value of θ warrants, so the expected time to the most recent common ancestor must be large. Moreover, when θ is known, it suggests the use of $\hat{T}_{(2)} = \{1 + S_{(2)}\}/(1 + \theta)$ as an estimator of $T_{(2)}$. If we compare $\hat{T}_{(2)}$ with $S_{(2)}/\theta$ (which is sometimes used, because $\mathrm{E}\{S_{(2)} \mid T_{(2)}\} = \theta T_{(2)}$), we observe that the first estimator is biased whereas the second is not. Nevertheless, $\hat{T}_{(2)}$ is superior to $S_{(2)}/\theta$ since the latter ignores post-data information and is also inappropriate when $S_{(2)} = 0$. For a review of some of the previous methods used to estimate $T_{(2)}$, see Tavaré *et al.* (1997).

4. Post-data inference in a two-island model

We attempt a parallel analysis in a two-island model of population structure with symmetric migration. Before we proceed to the actual derivation of densities, we give a brief explanation on the nature of the segregating sites. Consider Figure 2, which shows a typical coalescent tree with two subpopulations. From the figure, we can define the three quantities L_1, L_2, L_{12}, where

$$\begin{aligned} L_1 &= \text{total length of edges subtended to island 1 at time 0,} \\ L_2 &= \text{total length of edges subtended to island 2 at time 0,} \\ L_{12} &= \text{total length of edges subtended to both islands 1 and 2 at time 0.} \end{aligned}$$

If we denote by S_1, S_2 and S_{12} the number of segregating sites on island 1, on island 2 and on both islands 1 and 2 respectively, at time 0, then, conditional on L_1, L_2 and L_{12}, S_1, S_2 and S_{12} are independent Poisson random variables with means $\theta L_1/2$, $\theta L_2/2$ and $\theta L_{12}/2$ respectively. Application of Bayes' theorem then shows that the post-data distribution of the time T to ultimate coalescence, conditional on the starting state $\boldsymbol{\alpha}(0)$ and on the observed segregating site pattern $\boldsymbol{S} = \boldsymbol{k}$, where $\boldsymbol{S} = (S_1, S_2, S_{12})$, is given by the following probability density function:

$$f_T\{t \mid \boldsymbol{S} = \boldsymbol{k}, \boldsymbol{\alpha}(0)\} = f_T\{t \mid \boldsymbol{\alpha}(0)\} \frac{\Pr\{\boldsymbol{S} = \boldsymbol{k} \mid T = t, \boldsymbol{\alpha}(0)\}}{\Pr\{\boldsymbol{S} = \boldsymbol{k} \mid \alpha(0)\}} \tag{4.1}$$

(Tavaré *et al.* (1997)). We now consider some special cases of the above formula.

4.1. Starting state $\boldsymbol{\alpha}(0) = (2, 0)$

We first consider the case when both genes are initially (i.e. at time $t = 0$) on island 1 and assume a rate of symmetric migration between the two islands of $m/2$ per gene. Since

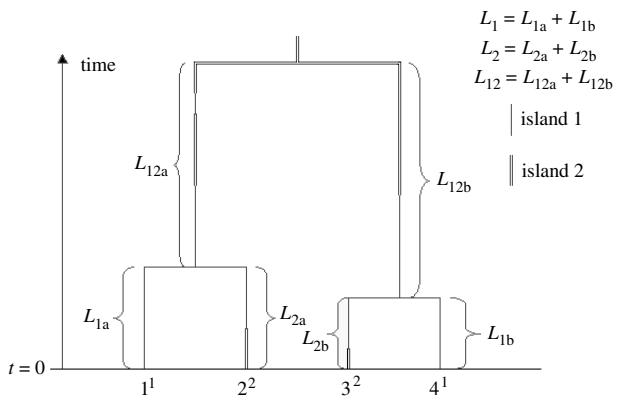

FIGURE 2: The different types of edges in a two-island situation. The numbers 1, 2, 3, 4 denote the labelled genes in the sample and the superscripts 1, 2 stand for the island on which a particular gene is at time $t = 0$. Migrations occur when the double vertical lines are replaced by single ones, or vice versa. A coalescence between two lineages can occur if and only if those two lineages are in the same colony at a particular point in time.

$\boldsymbol{\alpha}(0) = (2, 0)$, we have

$$\boldsymbol{S} = (S_1, S_2, S_{12}) = (S_1, 0, 0),$$
$$\boldsymbol{k} = (k_1, k_2, k_{12}) = (k_1, 0, 0),$$

and (4.1) can be written as

$$f_T\{t \mid \boldsymbol{S} = (k_1, 0, 0), \boldsymbol{\alpha}(0) = (2, 0)\}$$
$$= f_T\{t \mid \boldsymbol{\alpha}(0) = (2, 0)\} \frac{\Pr\{\boldsymbol{S} = (k_1, 0, 0) \mid T = t, \boldsymbol{\alpha}(0) = (2, 0)\}}{\Pr\{\boldsymbol{S} = (k_1, 0, 0) \mid \boldsymbol{\alpha}(0) = (2, 0)\}}. \tag{4.2}$$

Now, $f_T\{t \mid \boldsymbol{\alpha}(0) = (2, 0)\}$ is the density of the time to ultimate coalescence conditional on $\boldsymbol{\alpha}(0) = (2, 0)$:

$$f_T\{t \mid \boldsymbol{\alpha}(0) = (2, 0)\} = \frac{1}{2\xi}\{(\xi - 1)\mathrm{e}^{-(m+1/2-\xi/2)t} + (\xi + 1)\mathrm{e}^{-(m+1/2+\xi/2)t}\},$$

where $\xi = (1 + 4m^2)^{1/2}$ (Takahata (1988), Nath and Griffiths (1993), Herbots (1997), Gorroochurn (1999)). Also,

$$\Pr\{\boldsymbol{S} = (k_1, 0, 0) \mid T = t, \boldsymbol{\alpha}(0) = (2, 0)\} = \mathrm{e}^{-\theta t} \frac{(\theta t)^{k_1}}{k_1!}, \qquad t > 0,\ k_1 = 0, 1, 2, \ldots.$$

Finally, since $\Pr\{\boldsymbol{S} = (k_1, 0, 0) \mid \boldsymbol{\alpha}(0) = (2, 0)\}$ is constant with respect to t, we write

$$\frac{1}{\Pr\{\boldsymbol{S} = (k_1, 0, 0) \mid \boldsymbol{\alpha}(0) = (2, 0)\}} = K^{(s)},$$

where we calculate $K^{(s)}$ below. From (4.2), we have

$$f_T\{t \mid \boldsymbol{S} = (k_1, 0, 0), \boldsymbol{\alpha}(0) = (2, 0)\}$$
$$= K^{(s)} \frac{(\xi - 1)\mathrm{e}^{-(m+1/2-\xi/2)t} + (\xi + 1)\mathrm{e}^{-(m+1/2+\xi/2)t}}{2\xi} \mathrm{e}^{-\theta t} \frac{(\theta t)^{k_1}}{k_1!}.$$

After simplification, we can re-write the above density as

$$f_T\{t \mid \boldsymbol{S} = (k_1, 0, 0), \boldsymbol{\alpha}(0) = (2, 0)\} = K^{(s)}\left\{\frac{(\xi-1)\theta^{k_1}}{2\xi\phi_1^{k_1+1}} f_{T_1^{(s)}}(t) + \frac{(\xi+1)\theta^{k_1}}{2\xi\phi_2^{k_1+1}} f_{T_2^{(s)}}(t)\right\}, \tag{4.3}$$

where

$$T_1^{(s)} \sim \text{Gamma}\left(1+k_1, \frac{1}{\phi_1}\right), \qquad T_2^{(s)} \sim \text{Gamma}\left(1+k_1, \frac{1}{\phi_2}\right), \tag{4.4}$$
$$\phi_1 = m + \tfrac{1}{2} - \tfrac{1}{2}\xi + \theta, \qquad \phi_2 = m + \tfrac{1}{2} + \tfrac{1}{2}\xi + \theta.$$

We now integrate (4.3) with respect to t from 0 to ∞ and obtain

$$K^{(s)} = \frac{2\xi(\phi_1\phi_2)^{k_1+1}}{\theta^{k_1}\{(\xi-1)\phi_2^{k_1+1} + (\xi+1)\phi_1^{k_1+1}\}}. \tag{4.5}$$

Hence, by substituting (4.5) back into (4.3), we have

$$f_T\{t \mid \boldsymbol{S} = (k_1, 0, 0), \boldsymbol{\alpha}(0) = (2, 0)\} = \frac{(\xi-1)\phi_2^{k_1+1} f_{T_1^{(s)}}(t) + (\xi+1)\phi_1^{k_1+1} f_{T_2^{(s)}}(t)}{(\xi-1)\phi_2^{k_1+1} + (\xi+1)\phi_1^{k_1+1}}. \tag{4.6}$$

We see that $f_T\{t \mid \boldsymbol{S} = (k_1, 0, 0), \boldsymbol{\alpha}(0) = (2, 0)\}$ is a linear combination of two gamma densities. Straightforward calculations show that

$$\text{E}\{T \mid \boldsymbol{S} = (k_1, 0, 0), \boldsymbol{\alpha}(0) = (2, 0)\} = \frac{1+k_1}{\phi_1\phi_2}\left\{\frac{(\xi-1)\phi_2^{k_1+2} + (\xi+1)\phi_1^{k_1+2}}{(\xi-1)\phi_2^{k_1+1} + (\xi+1)\phi_1^{k_1+1}}\right\}. \tag{4.7}$$

Now, $\xi \to 2m$, $\phi_1 \to \theta + \frac{1}{2}$, $\phi_2 \to 2m$ and $\phi_1\phi_2 \to m + 2m\theta$, as $m \to \infty$; thus

$$\lim_{m\to\infty} \text{E}\{T \mid \boldsymbol{S} = (k_1, 0, 0), \boldsymbol{\alpha}(0) = (2, 0)\} = \frac{2(1+k_1)}{1+2\theta}.$$

The case when there are no mutations present in the sequences is also of interest since this implies that all gametes in the sample are then monomorphic (Watterson (1975), Dorit *et al.* (1995), Donnelly *et al.* (1996), Tavaré *et al.* (1997)). We have

$$\text{E}\{T \mid \boldsymbol{S} = (0, 0, 0), \boldsymbol{\alpha}(0) = (2, 0)\} = \frac{1}{\phi_1\phi_2}\left\{\frac{(\xi-1)\phi_2^2 + (\xi+1)\phi_1^2}{(\xi-1)\phi_2 + (\xi+1)\phi_1}\right\}.$$

From (4.6), we can also easily show that

$$\begin{aligned}\text{var}\{T \mid \boldsymbol{S} = (k_1, 0, 0), \boldsymbol{\alpha}(0) = (2, 0)\} &= \frac{1}{\{(\xi-1)\phi_2^{k_1+1} + (\xi+1)\phi_1^{k_1+1}\}^2\phi_1^2\phi_2^2} \\ &\quad\times ((\xi+1)^2\phi_1^{2k_1+4} + (\xi-1)^2\phi_2^{2k_1+4} \\ &\qquad + 4m^2(\phi_1\phi_2)^{k_1+1}\{2(\phi_1^2 + \phi_2^2 - \phi_1\phi_2) + k_1(\phi_1 - \phi_2)^2\}) \\ &\quad\times (1+k_1).\end{aligned}$$

This implies that, in the strong-migration limit (Nagylaki (1980), (2000)), we have the following result:

$$\lim_{m\to\infty} \operatorname{var}\{T \mid \boldsymbol{S} = (k_1, 0, 0), \boldsymbol{\alpha}(0) = (2, 0)\} = \frac{4(1+k_1)}{(1+2\theta)^2}$$

and also

$$\begin{aligned}&\operatorname{var}\{T \mid \boldsymbol{S} = (0, 0, 0), \boldsymbol{\alpha}(0) = (2, 0)\}\\ &\quad = \frac{(\xi+1)^2\phi_1^4 + (\xi-1)^2\phi_2^4 + 8m^2\phi_1\phi_2(\phi_1^2+\phi_2^2-\phi_1\phi_2)}{\{(\xi-1)\phi_2 + (\xi+1)\phi_1\}^2\phi_1^2\phi_2^2}.\end{aligned}$$

Finally, we consider the probability mass function $\Pr\{\boldsymbol{S} = (k_1, 0, 0) \mid \boldsymbol{\alpha}(0) = (2, 0)\}$. From (4.5), we have

$$\Pr\{\boldsymbol{S} = (k_1, 0, 0) \mid \boldsymbol{\alpha}(0) = (2, 0)\} = \frac{\theta^{k_1}}{2\xi}\left(\frac{\xi-1}{\phi_1^{k_1+1}} + \frac{\xi+1}{\phi_2^{k_1+1}}\right), \tag{4.8}$$

which is a linear combination of two geometric-like densities. The probability that the sample is monomorphic is then

$$\Pr\{\boldsymbol{S} = (0, 0, 0) \mid \boldsymbol{\alpha}(0) = (2, 0)\} = \frac{1}{2\xi}\left(\frac{\xi-1}{\phi_1} + \frac{\xi+1}{\phi_2}\right).$$

From (4.8), we obtain the probability generating function (PGF), $G_{\boldsymbol{S}}(z)$, for the number of segregating sites,

$$\begin{aligned}G_{\boldsymbol{S}}(z) &= \sum_{k_1=0}^{\infty} \frac{(z\theta)^{k_1}}{2\xi}\left(\frac{\xi-1}{\phi_1^{k_1+1}} + \frac{\xi+1}{\phi_2^{k_1+1}}\right)\\ &= \frac{(m+\theta) - \theta z}{(\theta+m+2m\theta+\theta^2) - \theta(2m+2\theta+1)z + \theta^2 z^2},\end{aligned}$$

from which it follows that

$$\mathrm{E}\{\boldsymbol{S} \mid \boldsymbol{\alpha}(0) = (2, 0)\} = 2\theta, \tag{4.9}$$

$$\operatorname{var}\{\boldsymbol{S} \mid \boldsymbol{\alpha}(0) = (2, 0)\} = 2\theta\left\{(2\theta+1) + \frac{\theta}{m}\right\}. \tag{4.10}$$

Note that the expectation in (4.9) is independent of the migration parameter (Slatkin (1987), Notohara (1997)).

4.2. Starting state $\boldsymbol{\alpha}(0) = (1, 1)$

Now, if $\boldsymbol{\alpha}(0) = (1, 1)$, then $\boldsymbol{S} = (S_1, S_2, 0)$, where S_1 and S_2 are independent and have the same Poisson distribution conditional on the TMRCA T. The equation (4.1) becomes

$$\begin{aligned}&f_T\{t \mid \boldsymbol{S} = (k_1, k_2, 0), \boldsymbol{\alpha}(0) = (1, 1)\}\\ &\quad = f_T\{t \mid \boldsymbol{\alpha}(0) = (1, 1)\}\frac{\Pr\{\boldsymbol{S} = (k_1, k_2, 0) \mid T = t, \boldsymbol{\alpha}(0) = (1, 1)\}}{\Pr\{\boldsymbol{S} = (k_1, k_2, 0) \mid \boldsymbol{\alpha}(0) = (1, 1)\}},\end{aligned} \tag{4.11}$$

with

$$f_T\{t \mid \boldsymbol{\alpha}(0) = (1, 1)\} = \frac{m}{\xi}\{\mathrm{e}^{-(m+1/2-\xi/2)t} - \mathrm{e}^{-(m+1/2+\xi/2)t}\} \tag{4.12}$$

(see Takahata (1988), Nath and Griffiths (1993), Herbots (1997), Gorroochurn (1999)). Also,

$$\Pr\{\boldsymbol{S} = (k_1, k_2, 0) \mid T = t, \boldsymbol{\alpha}(0) = (1, 1)\} = \frac{\mathrm{e}^{-\theta t/2}(\theta t/2)^{k_1}}{k_1!}\frac{\mathrm{e}^{-\theta t/2}(\theta t/2)^{k_2}}{k_2!}$$
$$= \mathrm{e}^{-\theta t}\frac{(\theta t)^{k_1+k_2}}{2^{k_1+k_2}k_1!k_2!}, \qquad k_1, k_2 = 0, 1, 2, \ldots \tag{4.13}$$

and we write

$$\frac{1}{\Pr\{\boldsymbol{S} = (k_1, k_2, 0) \mid \boldsymbol{\alpha}(0) = (1, 1)\}} = K^{(d)}. \tag{4.14}$$

By substituting (4.12), (4.13) and (4.14) into (4.11) and integrating with respect to t, we obtain

$$K^{(d)} = \frac{2^{k_1+k_2}\xi k_1!k_2!}{m(k_1+k_2)!\theta^{k_1+k_2}}\frac{(\phi_1\phi_2)^{k_1+k_2+1}}{(\phi_2^{k_1+k_2+1} - \phi_1^{k_1+k_2+1})}, \tag{4.15}$$

where ϕ_1 and ϕ_2 are as defined in (4.4).

If we substitute (4.12), (4.13) and the expression for $K^{(d)}$ back into (4.11), we obtain

$$f_T\{t \mid \boldsymbol{S} = (k_1, k_2, 0), \boldsymbol{\alpha}(0) = (1, 1)\} = \frac{\phi_2^{k_1+k_2+1} f_{T_1^{(d)}}(t) - \phi_1^{k_1+k_2+1} f_{T_2^{(d)}}(t)}{\phi_2^{k_1+k_2+1} - \phi_1^{k_1+k_2+1}}, \tag{4.16}$$

where

$$T_1^{(d)} \sim \text{Gamma}\left(k_1 + k_2 + 1, \frac{1}{\phi_1}\right),$$
$$T_2^{(d)} \sim \text{Gamma}\left(k_1 + k_2 + 1, \frac{1}{\phi_2}\right).$$

Note that $\phi_2 > \phi_1$, so that (as required) the density function in (4.16) is always positive and bounded (see Figures 3 and 4).

From (4.16), we can prove in the same way as before that

$$\mathrm{E}\{T \mid \boldsymbol{S} = (k_1, k_2, 0), \boldsymbol{\alpha}(0) = (1, 1)\} = \frac{k_1 + k_2 + 1}{\phi_1\phi_2}\left\{\frac{\phi_2^{k_1+k_2+2} - \phi_1^{k_1+k_2+2}}{\phi_2^{k_1+k_2+1} - \phi_1^{k_1+k_2+1}}\right\}$$

and

$$\lim_{m\to\infty} \mathrm{E}\{T \mid \boldsymbol{S} = (k_1, k_2, 0), \boldsymbol{\alpha}(0) = (1, 1)\} = \frac{2(k_1 + k_2 + 1)}{1 + 2\theta}.$$

The expected TMRCA for the monomorphic case is given by

$$\mathrm{E}\{T \mid \boldsymbol{S} = (0, 0, 0), \boldsymbol{\alpha}(0) = (1, 1)\} = \frac{1}{\phi_1} + \frac{1}{\phi_2}.$$

Also,

$$\mathrm{var}\{T \mid \boldsymbol{S} = (k_1, k_2, 0), \boldsymbol{\alpha}(0) = (1, 1)\}$$
$$= \frac{1}{\{\phi_2^{k_1+k_2+1} - \phi_1^{k_1+k_2+1}\}^2\phi_1^2\phi_2^2}$$
$$\times (\phi_2^{2(k_1+k_2+2)} - \phi_1^{2(k_1+k_2+2)}$$
$$- (\phi_1\phi_2)^{k_1+k_2+1}\{2(\phi_1^2 + \phi_2^2 - \phi_1\phi_2) + (k_1 + k_2)(\phi_1 - \phi_2)^2\})$$
$$\times (k_1 + k_2 + 1).$$

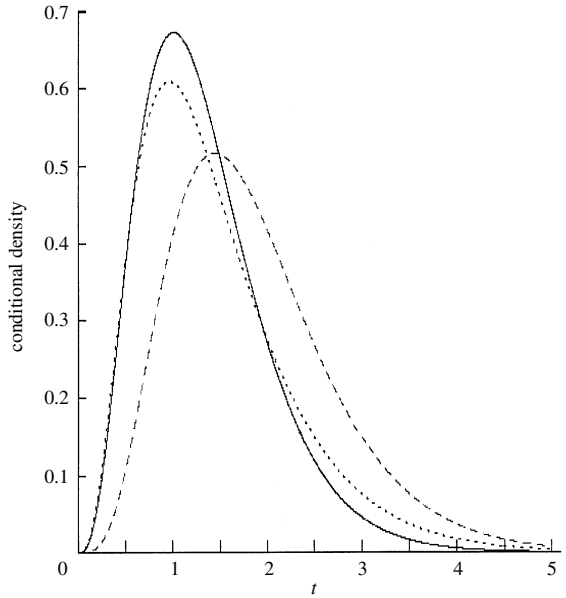

FIGURE 3: Plots of the various densities $f_T\{t \mid \boldsymbol{S} = (k, 0, 0), \boldsymbol{\alpha}(0) = (2, 0)\}$ ($\boldsymbol{\alpha}(0) = (2, 0)$ case, dotted line), $f_T\{t \mid \boldsymbol{S} = (k_1, k_2, 0), \boldsymbol{\alpha}(0) = (1, 1)\}$ ($\boldsymbol{\alpha}(0) = (1, 1)$ case, dashed line) and $f_{T_{(2)}}\{t \mid S_{(2)} = k\}$ (panmictic case, solid line) with $m = 0.5$, $\theta = 2$ and $k_1 + k_2 = k = 3$.

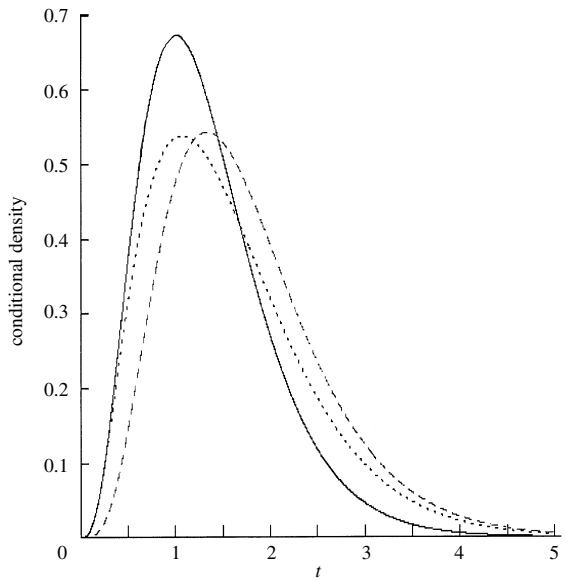

FIGURE 4: Plots of the various densities $f_T\{t \mid \boldsymbol{S} = (k, 0, 0), \boldsymbol{\alpha}(0) = (2, 0)\}$ ($\boldsymbol{\alpha}(0) = (2, 0)$ case, dotted line), $f_T\{t \mid \boldsymbol{S} = (k_1, k_2, 0), \boldsymbol{\alpha}(0) = (1, 1)\}$ ($\boldsymbol{\alpha}(0) = (1, 1)$ case, dashed line) and $f_{T_{(2)}}\{t \mid S_{(2)} = k\}$ (panmictic case, solid line) with $m = 1$, $\theta = 2$ and $k_1 + k_2 = k = 3$.

We note that the conditional density, expectation and variance of the TMRCA depend on the observed segregation site pattern $\boldsymbol{k} = (k_1, k_2, 0)$ only through $k_1 + k_2$, a consequence of the fact that the associated Markov chain $\{\boldsymbol{\alpha}(t), P_{\boldsymbol{\alpha}}\}$ (see Section 2.2) is time-reversible. Also,

$$\lim_{m\to\infty} \operatorname{var}\{T \mid \boldsymbol{S} = (k_1, k_2, 0), \boldsymbol{\alpha}(0) = (1, 1)\} = \frac{4(k_1 + k_2 + 1)}{(2\theta + 1)^2},$$

$$\operatorname{var}\{T \mid \boldsymbol{S} = (0, 0, 0), \boldsymbol{\alpha}(0) = (1, 1)\} = \frac{\phi_2^4 - \phi_1^4 - 2\phi_1\phi_2(\phi_1^2 + \phi_2^2 - \phi_1\phi_2)}{(\phi_2 - \phi_1)^2\phi_1^2\phi_2^2}.$$

Finally, from (4.15), we have

$$\Pr\{\boldsymbol{S} = (k_1, k_2, 0) \mid \boldsymbol{\alpha}(0) = (1, 1)\} = \binom{k_1 + k_2}{k_1} \frac{m}{\xi} \left(\frac{\theta}{2}\right)^{k_1+k_2} \left(\frac{1}{\phi_1^{k_1+k_2+1}} - \frac{1}{\phi_2^{k_1+k_2+1}}\right).$$

The probability of observing a monomorphic sample is

$$\Pr\{\boldsymbol{S} = (0, 0, 0) \mid \boldsymbol{\alpha}(0) = (1, 1)\} = \frac{m}{\xi}\left(\frac{1}{\phi_1} - \frac{1}{\phi_2}\right).$$

5. Results for the case when both genes are initially on the same island

5.1. Dependence of the expected TMRCA on *m*

We note that the expectation in (4.7) is very dependent on m (see Figures 5 and 6), as opposed to $\mathrm{E}\{T \mid \boldsymbol{\alpha}(0) = (2, 0)\}$ which is constant at 2 (Notohara (1990), Nath and Griffiths (1993), Wakeley (1998)); thus, including the effects of mutation and conditioning on the number of segregating sites results in the expected TMRCA being dependent on m, when the starting state is $\boldsymbol{\alpha}(0) = (2, 0)$.

5.2. Convergence in both distribution and mean of the TMRCA

We can take the limit of the density in (4.6) as $m \to 0^+$ and check if there is convergence to the panmictic density. As $m \to 0^+$, $\xi \to 1^+$, $\phi_1 \to \theta$, $\phi_2 \to 1 + \theta$, so that

$$\lim_{m\to 0^+} f_T\{t \mid \boldsymbol{S} = (k_1, 0, 0), \boldsymbol{\alpha}(0) = (2, 0)\} = f_{T_2^{(s)}}(t),$$

where $T_2^{(s)} \sim \text{Gamma}(1 + k_1, 1/(1 + \theta))$ for $m \to 0^+$. Also, if we take the weak-migration limit for the corresponding expectation, we obtain

$$\lim_{m\to 0^+} \mathrm{E}\{T \mid \boldsymbol{S} = (k_1, 0, 0), \boldsymbol{\alpha}(0) = (2, 0)\} = \frac{1 + k_1}{1 + \theta}.$$

Thus there is convergence to the single-island case in the conditional distribution as well as in the conditional expectation as $m \to 0^+$ (see (3.1) and (3.2)). This is quite different from the situation in which forces of mutation are not taken into account, where there is panmictic convergence in the density but not in the expectation (Nath and Griffiths (1993)).

5.3. Convergence problems in the distribution of segregating sites

The results in (4.9) and (4.10) are surprising and point towards *convergence problems* in the PGF $G_{\boldsymbol{S}}(z)$ in the weak-migration limit. We have

$$\lim_{m\to 0^+} G_{\boldsymbol{S}}(z) = \frac{1}{1 + \theta - \theta z}, \tag{5.1}$$

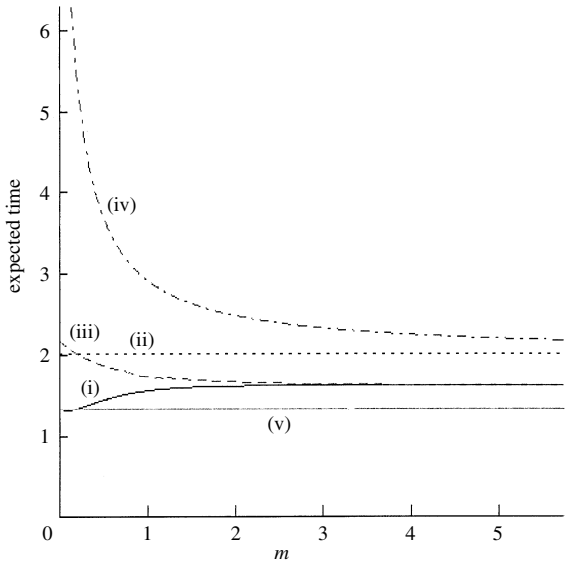

FIGURE 5: Variation of various expected times with $m(\theta = 2$ and $k_1 + k_2 = k = 3)$. (i) $\mathrm{E}^*\{T \mid \boldsymbol{S} = (k_1, 0, 0), \boldsymbol{\alpha}(0) = (2, 0)\}$ (two-island situation); (ii) $\mathrm{E}\{T \mid \boldsymbol{\alpha}(0) = (2, 0)\}$ (two-island situation); (iii) $\mathrm{E}\{T \mid \boldsymbol{S} = (k_1, k_2, 0), \boldsymbol{\alpha}(0) = (1, 1)\}$ (two-island situation); (iv) $\mathrm{E}\{T \mid \boldsymbol{\alpha}(0) = (1, 1)\}$ (two-island situation); (v) $\mathrm{E}\{T_{(2)} \mid S_{(2)} = k\}$ (panmictic situation).

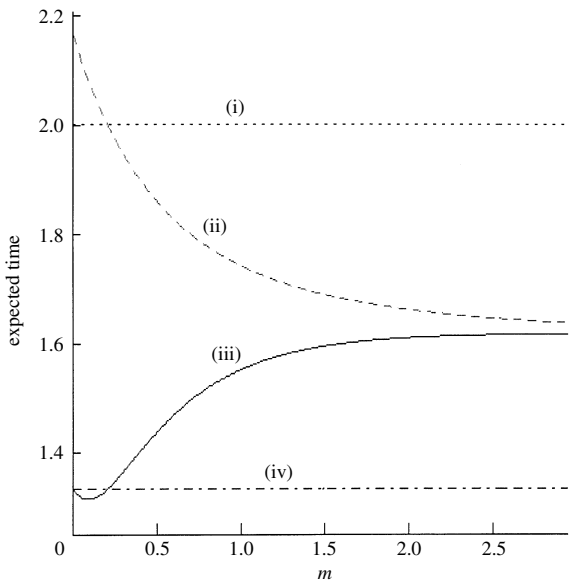

FIGURE 6: A magnification of the graph in Figure 5. (i) $\mathrm{E}\{T \mid \boldsymbol{\alpha}(0) = (2, 0)\}$ (two-island situation); (ii) $\mathrm{E}\{T \mid \boldsymbol{S} = (k_1, k_2, 0), \boldsymbol{\alpha}(0) = (1, 1)\}$ (two-island situation); (iii) $\mathrm{E}\{T \mid \boldsymbol{S} = (k_1, 0, 0), \boldsymbol{\alpha}(0) = (2, 0)\}$ (two-island situation); (iv) $\mathrm{E}\{T_{(2)} \mid S_{(2)} = k\}$ (panmictic situation).

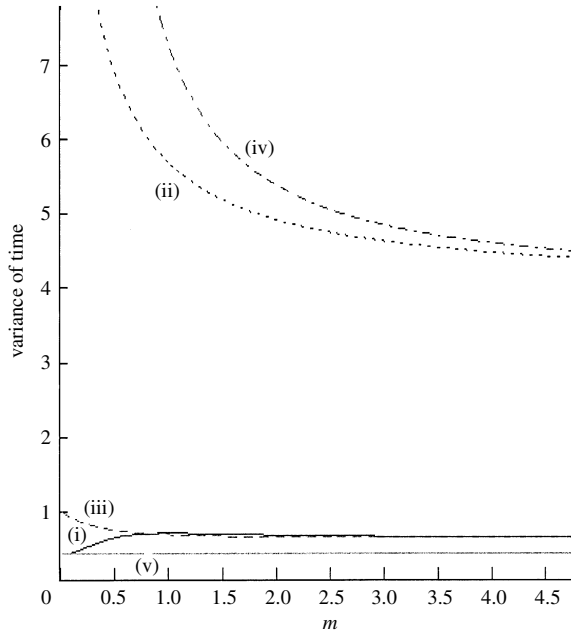

FIGURE 7: Variation of various variances with m ($\theta = 2$ and $k_1 + k_2 = k = 3$). (i) $\mathrm{var}\{T \mid \boldsymbol{S} = (k_1, 0, 0), \boldsymbol{\alpha}(0) = (2, 0)\}$ (two-island situation); (ii) $\mathrm{var}\{T \mid \boldsymbol{\alpha}(0) = (2, 0)\}$ (two-island situation); (iii) $\mathrm{var}\{T \mid \boldsymbol{S} = (k_1, k_2, 0), \boldsymbol{\alpha}(0) = (1, 1)\}$ (two-island situation); (iv) $\mathrm{var}\{T \mid \boldsymbol{\alpha}(0) = (1, 1)\}$ (two-island situation); (v) $\mathrm{var}\{T_{(2)} \mid S_{(2)} = k\}$ (panmictic situation).

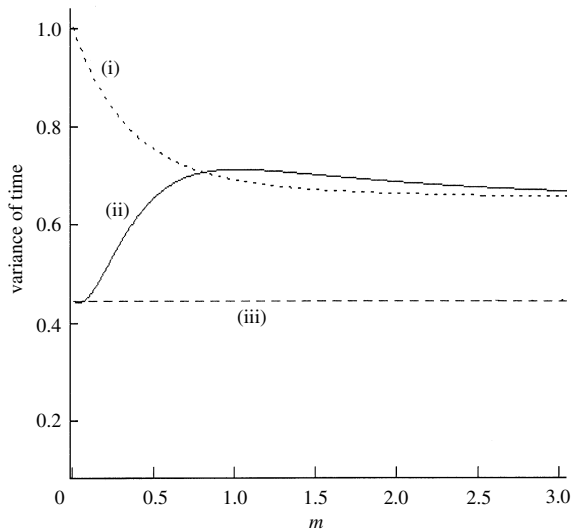

FIGURE 8: A magnification of the graph in Figure 7. (i) $\mathrm{var}\{T \mid \boldsymbol{S} = (k_1, k_2, 0), \boldsymbol{\alpha}(0) = (1, 1)\}$ (two-island situation); (ii) $\mathrm{var}\{T \mid \boldsymbol{S} = (k_1, 0, 0), \boldsymbol{\alpha}(0) = (2, 0)\}$ (two-island situation); (iii) $\mathrm{var}\{T_{(2)} \mid S_{(2)} = k\}$ (panmictic situation).

which is the panmictic PGF for $S_{(2)}$ (Watterson (1975)), so there is convergence in the (conditional) PGF of $\boldsymbol{S} = (S_1, 0, 0)$, as $m \to 0^+$, to the panmictic PGF. However, there is clearly divergence in both the (conditional) mean and variance:

$$\lim_{m\to 0^+} \mathrm{E}\{\boldsymbol{S} \mid \boldsymbol{\alpha}(0) = (2,0)\} = 2\theta,$$
$$\lim_{m\to 0^+} \mathrm{var}\{\boldsymbol{S} \mid \boldsymbol{\alpha}(0) = (2,0)\} = +\infty,$$

compared to the panmictic values:

$$\mathrm{E}\{S_{(2)}\} = \theta,$$
$$\mathrm{var}\{S_{(2)}\} = \theta + \theta^2.$$

6. Discussion

Conditioning on the observed segregating site pattern leads to marked changes in the behaviour of the expected TMRCA, especially, for a sample of size two, when both genes are initially on the same island. Figures 3 and 4 show the graphs of the densities in (3.1) (panmictic case), (4.6) ($\boldsymbol{\alpha}(0) = (2, 0)$ case) and (4.16) ($\boldsymbol{\alpha}(0) = (1, 1)$ case) for when $m = 0.5$ and $m = 1$. Note the increase in variability due to subpopulation division, particularly when migration rates are small.

TABLE 1: The variation of the 'unconditional' expected time $\mathrm{E}\{T \mid \boldsymbol{\alpha}(0) = (5,5)\}$ and the 'unconditional' variance $\mathrm{var}\{T \mid \boldsymbol{\alpha}(0) = (5,5)\}$ with m. Here $\theta = 3$.

migration parameter, m	'unconditional' expected time, $\mathrm{E}\{T \mid \boldsymbol{\alpha}(0) = (5,5)\}$	'unconditional' variance of time, $\mathrm{var}\{T \mid \boldsymbol{\alpha}(0) = (5,5)\}$
0.1	11.781	102.72
0.6	3.4619	3.6554
1.1	2.6927	1.4314
1.6	2.4118	0.87260
2.1	2.2620	0.62990
2.6	2.1756	0.49796
3.1	2.1118	0.41213
3.6	2.0682	0.35319
4.1	2.0349	0.30958
4.6	2.0140	0.27704
5.1	1.9833	0.24877
5.6	1.9690	0.18430
6.1	1.9566	0.20971
6.6	1.9459	0.19458
7.1	1.9348	0.18119
7.6	1.9298	0.17022
8.1	1.9226	0.16016
8.6	1.9087	0.15058
9.1	1.9025	0.14260
9.6	1.9057	0.13630

TABLE 2: The variation of the 'conditional' expected time $\mathrm{E}\{T \mid \boldsymbol{S} = (3,3,2), \boldsymbol{\alpha}(0) = (5,5)\}$ and the 'conditional' variance $\mathrm{var}\{T \mid \boldsymbol{S} = (3,3,2), \boldsymbol{\alpha}(0) = (5,5)\}$ with m. Here $\theta = 3$.

migration parameter, m	'conditional' expected time, $\mathrm{E}\{T \mid \boldsymbol{S} = (3,3,2), \boldsymbol{\alpha}(0) = (5,5)\}$	'conditional' variance of time, $\mathrm{var}\{T \mid \boldsymbol{S} = (3,3,2), \boldsymbol{\alpha}(0) = (5,5)\}$
0.1	2.0071	0.51299
0.6	1.8633	0.43024
1.1	1.7337	0.37682
1.6	1.6544	0.34444
2.1	1.6015	0.32312
2.6	1.5701	0.31654
3.1	1.5388	0.31523
3.6	1.5168	0.30526
4.1	1.5038	0.29993
4.6	1.4932	0.30605
5.1	1.4787	0.29010
5.6	1.4667	0.29323
6.1	1.4595	0.29307
6.6	1.4491	0.28628
7.1	1.4466	0.28574
7.6	1.4450	0.29173
8.1	1.4437	0.29206
8.6	1.4328	0.28366
9.1	1.4287	0.28512
9.6	1.4312	0.28364

Figure 5 shows a comparison of various expected times as they vary with m. The graph of $\mathrm{E}\{T \mid \boldsymbol{S} = (k_1, 0, 0), \boldsymbol{\alpha}(0) = (2,0)\}$ clearly shows its dependence on m. If we compare the graph of $\mathrm{E}\{T \mid \boldsymbol{S} = (k_1, 0, 0), \boldsymbol{\alpha}(0) = (2,0)\}$ with that of $\mathrm{E}\{T \mid \boldsymbol{\alpha}(0) = (2,0)\}$ and the graph of $\mathrm{E}\{T \mid \boldsymbol{S} = (k_1, k_2, 0), \boldsymbol{\alpha}(0) = (1,1)\}$ with that of $\mathrm{E}\{T \mid \boldsymbol{\alpha}(0) = (1,1)\}$, we see that conditioning on $\boldsymbol{S}$ results in a reduction in the expected TMRCA in general. This is not always true in panmixia. The behaviour of $\mathrm{E}\{T \mid \boldsymbol{S} = (k_1, 0, 0), \boldsymbol{\alpha}(0) = (1,1)\}$ comes as no surprise: as m decreases, the expected TMRCA increases. This is because, with decreasing migration rates, the two genes have to wait longer before they can be brought, through migration, to the same colony so that they can coalesce. The graph of $\mathrm{E}\{T \mid \boldsymbol{S} = (k_1, 0, 0), \boldsymbol{\alpha}(0) = (2,0)\}$ gives another interesting observation, which is more clearly seen in Figure 6: a minimum when m is very small.

In Figures 7 and 8, different variances are compared. The notable thing is the marked decrease in the variance of the TMRCA when the observed segregating-site pattern is taken into account.

We now consider samples of size greater than two. Using an adaptation of the rejection algorithm (Tavaré *et al.* (1997)), it is possible to simulate the TMRCA and the segregating-site pattern, and calculate their corresponding means and variances over a large number of runs (100 000 typically) for $n > 2$, and different starting states, values of m and θ, and *observed* segregating-site pattern $\boldsymbol{k}$. In Tables 1 and 2 we list various means and variances for increasing values of m, and we use the term 'conditional' to mean conditional on the observed segregating-site pattern $\boldsymbol{k} = (k_1, k_2, k_{12})$. Also, we assume $\boldsymbol{\alpha}(0) = (5,5)$, $\theta = 3$ and $\boldsymbol{k} = (3,3,2)$. We

TABLE 3: The variation of the mean segregating pattern $\mathrm{E}\{\boldsymbol{S} \mid \boldsymbol{\alpha}(0) = (5,5)\}$ with m. Here $\theta = 3$.

migration parameter, m	mean segregating pattern, $\mathrm{E}\{\boldsymbol{S} \mid \boldsymbol{\alpha}(0) = (5,5)\}$
0.1	(18.017, 17.982, 3.1640)
0.6	(5.3423, 5.3338, 3.4066)
1.1	(4.0754, 4.0740, 3.5405)
1.6	(3.5448, 3.5503, 3.6937)
2.1	(3.2608, 3.2503, 3.7772)
2.6	(3.0692, 3.0604, 3.8562)
3.1	(2.9325, 2.9170, 3.1938)
3.6	(2.8168, 2.8169, 3.9760)
4.1	(2.7295, 2.7339, 4.0248)
4.6	(2.6715, 2.6716, 4.0655)
5.1	(2.6034, 2.6013, 4.0903)
5.6	(2.5565, 2.5559, 4.1235)
6.1	(2.5300, 2.5006, 4.1534)
6.6	(2.4837, 2.4777, 4.1796)
7.1	(2.4451, 2.4515, 4.2030)
7.6	(2.4259, 2.4297, 4.2185)
8.1	(2.3965, 2.4072, 4.2404)
8.6	(2.3765, 2.3731, 4.2422)
9.1	(2.3590, 2.3501, 4.2586)
9.6	(2.3425, 2.3423, 4.2877)

do not, in general, expect the behaviour of the conditional means and variances to be greatly different from that in the $\boldsymbol{\alpha}(0) = (1,1)$ case. In Table 3, we give the variation of the mean segregating-site pattern $\mathrm{E}\{\boldsymbol{S} \mid \boldsymbol{\alpha}(0) = (5,5)\}$ with m ($\theta = 3$). Figures 9 and 10 also illustrate the results obtained.

We note that the graphs in Figures 9 and 10 are very similar to the $n = 2$ case, in particular when $\boldsymbol{\alpha}(0) = (1,1)$. In Table 3, as m increases, both $\mathrm{E}\{S_1 \mid \boldsymbol{\alpha}(0) = (5,5)\}$ and $\mathrm{E}^*\{S_2 \mid \boldsymbol{\alpha}(0) = (5,5)\}$ decrease, while $\mathrm{E}\{S_{12} \mid \boldsymbol{\alpha}(0) = (5,5)\}$ increases and, overall, the expected total number of segregating sites decreases as well. The reason is intuitively clear: as m increases, because of the increased number of migrations, there will be fewer segregating sites on island 1 or 2 alone and more on both islands; the TMRCA also decreases, so that the total number of mutations decreases as well.

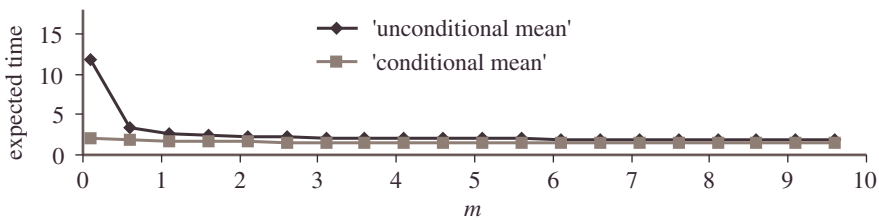

FIGURE 9: The variation of the 'unconditional' expected time $\mathrm{E}\{T \mid \boldsymbol{\alpha}(0) = (5,5)\}$ and the 'conditional' expected time $\mathrm{E}\{T \mid \boldsymbol{S} = (3,3,3), \boldsymbol{\alpha}(0) = (5,5)\}$ with m. Here $\theta = 3$.

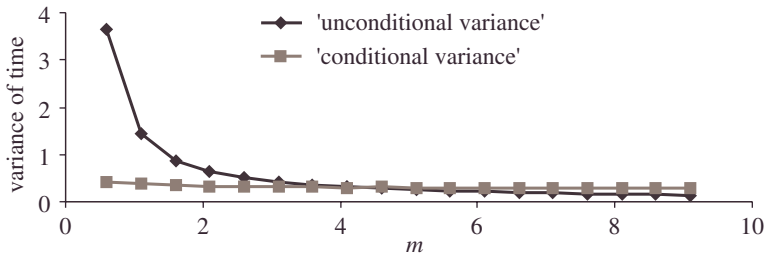

FIGURE 10: The variation of the 'unconditional' variance var$\{T \mid \alpha(0) = (5,5)\}$ and the 'conditional' variance var$\{T \mid \boldsymbol{S} = (3,3,3), \boldsymbol{\alpha}(0) = (5,5)\}$ with m. Here $\theta = 3$.

References

BAHLO, M. AND GRIFFITHS, R. C. (2000). Inference from gene trees in a subdivided population. *Theoret. Popn Biol.* **57,** 79–95.

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

DONNELLY, P. AND TAVARÉ, S. (1995). Coalescents and genealogical structure under neutrality. *Annual Rev. Genet.* **29,** 401–421.

DONNELLY, P., TAVARÉ, S., BALDING, D. J. AND GRIFFITHS, R. C. (1996). Estimating the age of the common ancestor of men from the ZFY intron. *Science* **272,** 1357–1359.

DORIT, R. L., AKASHI, H. AND GILBERT, W. (1995). Absence of polymorphism at the ZFY locus on the human Y chromosome. *Science* **268,** 1183–1185.

EWENS, W. J. (1989). Population genetics—the past and the future. In *Mathematical and Statistical Problems of Evolutionary Theory*, ed. S. Lessard. Kluwer, Dordrecht, pp. 177–227.

FU, Y. X. AND LI, W. H. (1998). Coalescent theory and its applications in population genetics. In *Statistics in Genetics*, ed. E. Halloran. Springer, Berlin, pp. 45–79.

FU, Y. X. AND LI, W. H. (1999). Coalescing into the 21st century: an overview and prospects of coalescent theory. *Theoret. Popn Biol.* **56,** 1–10.

GORROOCHURN, P. (1999). Population genetics in subdivided populations. Doctoral Thesis, Monash University.

GRIFFITHS, R. C. AND TAVARÉ, S. (1994). Ancestral inference in population genetics. *Statist. Sci.* **9,** 307–319.

HERBOTS, H. M. (1997). The structured coalescent. In *Progress in Population Genetics and Human Evolution* (IMA Vols Math. Appl. **87**), eds P. Donnelly and S. Tavaré. Springer, New York, pp. 231–255.

HUDSON, R. R. (1991). Gene genealogies and the coalescent process. In *Oxford Surveys in Evolutionary Biology* **7,** eds D. Futuyma and J. Antonovics. Oxford University Press, pp. 1–44.

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

KINGMAN, J. F. C. (1982b). Exchangeability and the evolution of large populations. In *Exchangeability in Probability and Statistics*, eds G. Koch and F. Spizzichino. North-Holland, Amsterdam, pp. 97–112.

KINGMAN, J. F. C. (1982c). On the genealogy of large populations. *Adv. Appl. Prob.* **19A,** 27–43.

MARJORAM, P. AND DONNELLY, P. (1994). Pairwise comparisons of mitochondrial DNA sequences in subdivided populations and implications for early human evolution. *Genetics* **136,** 673–683.

NAGYLAKI, T. (1980). The strong-migration limit in geographically structured populations. *J. Math. Biol.* **9,** 101–114.

NAGYLAKI, T. (2000). Geographical invariance and the strong-migration limit in subdivided populations. *J. Math. Biol.* **41,** 123–142.

NATH, H. B. AND GRIFFITHS, R. C. (1993). The coalescent in two colonies with symmetric migration. *J. Math. Biol.* **31,** 841–852.

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

NOTOHARA, M. (1993). The strong-migration limit for the genealogical process in geographically structured populations. *J. Math. Biol.* **31,** 115–122.

NOTOHARA, M. (1997). The number of segregating sites in a sample of DNA sequences from a geographically structured population. *J. Math. Biol.* **36,** 188–200.

SLATKIN, M. (1987). The average number of sites separating DNA sequences drawn from a subdivided population. *Theoret. Popn Biol.* **32,** 42–49.

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

Tajima, F. (1983). Evolutionary relationships of DNA sequences in finite populations. *Genetics* **105,** 437–460.

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

Takahata, N. (1991). Genealogy of neutral genes and spreading of selected mutations in a geographically structured population. *Genetics* **129,** 585–595.

Tavaré, S., Balding, D. J., Griffiths, R. C. and Donnelly, P. (1997). Inferring coalescence times from DNA sequence data. *Genetics* **145,** 505–518.

Wakeley, J. (1998). Segregating sites in Wright's island model. *Theoret. Popn Biol.* **53,** 166–174.

Watterson, G. A. (1975). On the number of segregating sites in genetic models without recombination. *Theoret. Popn 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.

Wright, S. (1931). Evolution in Mendelian populations. *Genetics* **16,** 97–159.