Theoretical Population Biology **57**, 109–119 (2000)
doi:10.1006/tpbi.1999.1442, available online at http://www.idealibrary.com on IDEAL®

Times on Trees, and the Age of an Allele

Matthew Stephens

Department of Statistics, University of Oxford, Oxford OX1 3TG, United Kingdom
E-mail: stephens@stats.ox.ac.uk

Received November 12, 1998

In this paper we consider the genealogy of a random sample of *n* chromosomes from a panmictic population which has evolved with constant size *N* over many generations. We address two related problems. First we describe how genealogical information may be usefully partitioned into information on the events (mutations and coalescences) which occur in the genealogy, and the times between these events. We show that the distribution of the times given information on the events is particularly simple and describe how this can considerably reduce the computational burden when performing inference for these times. Second we investigate the effect on the genealogy of conditioning on a single mutation having occurred during the ancestry of the sample. In particular we use results from the first part of the paper to derive explicit formulae for the density of the age of a mutant allele, conditional on its frequency in either a sample or the population.

Key Words: **age of a mutation; ancestral inference; coalescent; genealogy; population genetics.**

1. INTRODUCTION

Consider a random sample of n chromosomes from a panmictic population which has evolved with constant (haploid) size N over many generations. For sufficiently large N the ancestral relationships between the chromosomes in such a sample are well modelled by the Kingman coalescent (Kingman 1982), in which lineages are considered backwards in time and merge (coalesce) whenever two lineages share a common ancestor. If time is measured in units of N generations then each pair of lineages coalesces independently as a Poisson process with rate 1, and so when there are k ancestral lines coalescences occur as a Poisson process with total rate $k(k-1)/2$.

We consider a general mutation mechanism for a single locus, in which the probability that a mutation occurs in a given chromosome in a given generation is μ. We denote the set of possible genetic types (alleles) by E and assume that all alleles are neutral (no selection) and that the mutation mechanism is a Markov process on E. This model thus includes the *infinite alleles* model, the *infinite sites* and *finite sites* models for sequence data, and the *k-alleles* model (in particular the stepwise mutation model for microsatellite data). If we consider the limits $N \to \infty$ and $\mu \to 0$, with $\theta = 2N\mu$ held constant, then mutations occur independently on different ancestral lines of the coalescent tree as a Poisson process of rate $\theta/2$. For further information on the use of the coalescent to model genealogies see one of the reviews by Donnelly and Tavaré (1995), Hudson (1991), or (for a more mathematical treatment) Tavaré (1984).

This paper considers the problem of inference for the times at which events in a genealogy occur, which includes inference for the ages of alleles observed in a sample and inference for the time since the most recent common ancestor of the sample. Such problems have had much attention in the literature, both in practical data analysis (see for example Harding *et al.*, 1997; Hammer *et al.*, 1998), and from a more theoretical viewpoint (including Kimura and Ohta, 1973; Watterson, 1977; Watterson and Guess, 1977; Donnelly, 1986; Donnelly and Tavaré, 1986; Griffiths and Tavaré, 1998; Wiuf and Donnelly, 1999).

0040-5809/00 $35.00

In Section 2 we describe how genealogical information may be usefully partitioned into information on the events (mutations and coalescences) which occur in the genealogy and the times between these events. We show that the distribution of the times given the events is particularly simple and describe how this can be used to reduce the computational burden when using certain computationally intensive methods (including the method of Griffiths and Tavaré, 1998, and any Markov chain Monte Carlo scheme in which mutations are considered explicitly) to estimate the times at which events in the genealogy occurred.

In Section 3 we consider the distribution of the age of a neutral mutation persisting in a finite population, first studied by Kimura and Ohta (1973), and subsequently by many authors including Watterson (1977), Griffiths and Tavaré (1998) and Wiuf and Donnelly (2000). In particular we use results from Section 2 to derive a parsimonious representation of the distribution of the age of a mutant allele conditional on its relative frequency in a sample under a general mutation model and obtain an explicit formula for its density.

2. TIMES ON TREES

Recall that we model the genealogical history of our sample by the Kingman coalescent, with mutations superposed independently on different ancestral lines as a Poisson process of rate $\theta/2$. Let T_k denote interchangeably both the time interval during which there are k ancestral lines and the length of this time interval. The coalescent process is usually considered backwards in time, starting with n ancestral lines at the time the sample was taken. Each pair of lines is assumed to coalesce independently as a Poisson process of rate 1, and so during T_k the total rate at which coalescences occur is $k(k-1)/2$ and T_k is exponentially distributed with this rate parameter:

$$T_k \sim \text{Exp}(k(k-1)/2). \tag{1}$$

The process continues backwards in time, reducing the number of lines by one each time a coalescence occurs, until the final two lines coalesce at a point which represents the *most recent common ancestor* (MRCA) of the sample.

If we know the distribution π of the type of the MRCA of the sample (which is the stationary distribution of the mutation mechanism if the population is assumed to have reached stationarity and we have no other information) then the genealogical history $\mathcal{G}$ of a random sample of size n may be sampled by simulating the ancestral lineages backwards in time according to the coalescent process, simulating the type of the MRCA from π, and then superposing mutations on different lineages as independent Poisson processes of rate $\theta/2$ (see for example Donnelly and Tavaré, 1995). An example of a simulated genealogical history for a sample of size 5 is shown in Fig. 1. This method of sampling $\mathcal{G}$ separates the independent Poisson processes governing ancestral lineages and mutation. An alternative method can be obtained by combining these two processes into a single Poisson process evolving forwards in time, which results in the following "urn model" method for sampling $\mathcal{G}$:

ALGORITHM 2.1. Urn model for sampling $\mathcal{G}$. This is a continuous-time version of the discrete-time urn model discussed by Ethier and Griffiths (1987); its properties are investigated by Donnelly and Kurtz (1996).

1. Choose a type in E at random according to the distribution π. Start the ancestry with a gene of this type, which splits immediately into two lines of this type.

2. If there are currently k lines in the ancestry, wait a random amount of time which is exponentially distributed with rate parameter $\lambda_k = k(k-1+\theta)/2$ and then choose an ancestral line at random from the k. Split this line into two with probability $(k-1)/(k-1+\theta)$; otherwise mutate it (according to the Markov mutation mechanism).

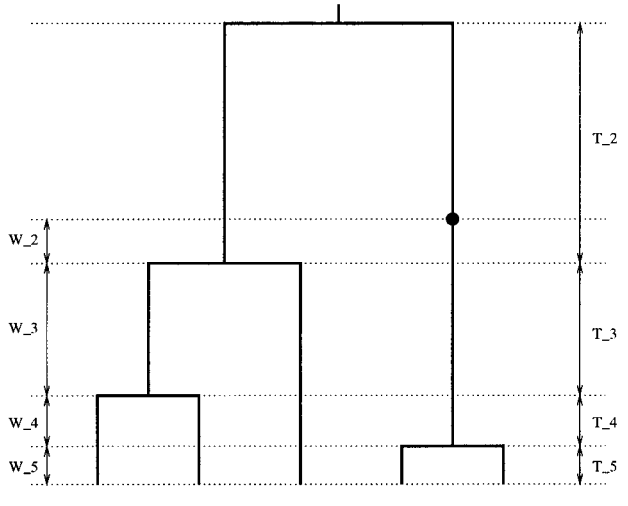

FIG. 1. A simulated genealogical history for a sample of size 5, with a single mutation (denoted by a black circle). The times T_k ($k = 2, ..., 5$), during which there are k ancestral lines, are *a priori* exponentially distributed with rate parameter $k(k-1)/2$. The conditional distribution of these times given the history shown is given by Theorem 2.1. Similarly, the conditional distribution of the age of the mutation is given by expressions (29) and (30) as $W_2 + \cdots + W_5$, where W_i is exponentially distributed with parameter $i(i-1+\theta)/2$.

3. If there are fewer than $n+1$ lines in the ancestry return to step 2. Otherwise go back to the last time at which there were n lines in the ancestry and stop.

This urn model suggests a partition of the genealogical history $\mathscr{G}$ into two parts: the *history* $\mathscr{H}$ which records the type of the ancestral gene and the details of the mutation and split events and the order in which they occur, and the *time information* $\mathscr{T}$ which records the times between these events. The following proposition gives the conditional distribution of $\mathscr{T}$ given $\mathscr{H}$.

PROPOSITION 2.1. *Given the history $\mathscr{H}$ of a random sample of n chromosomes from a panmictic constant-sized population, the times between consecutive events in the history are independent, and when there are k lines in the ancestry of the sample the time to the next event has an exponential distribution with rate parameter* $\lambda_k = k(k-1+\theta)/2$.

COROLLARY 2.2. *The distribution of $\mathscr{T}$ given both $\mathscr{H}$ and the observed types A_n is the same as the distribution of $\mathscr{T}$ given only $\mathscr{H}$, and so is given by Proposition* 2.1.

Proof. We note that (by definition) $\mathscr{H}$ records the states visited by the Markov process described by Algorithm 2.1, beginning with the ancestral gene, and ending with the genetic types A_n observed in our sample. Correspondingly, $\mathscr{T}$ records the time spent in each state visited by this process. Consideration of Algorithm 2.1 shows that the fact that $\mathscr{H}$ ends at A_n is equivalent to information that the next event in the process is a split event, and the Markov nature of the process means that this contains no information about $\mathscr{T}$. The distribution of $\mathscr{T}$ given $\mathscr{H}$ is therefore simply the conditional distribution of the jump times of the Markov process given the jump chain of the process. These jump times are independent, and when there are k lines in the ancestry events occur at rate $\lambda_k = k(k-1+\theta)/2$, so the time to the next event has an exponential distribution with rate parameter λ_k.

The Corollary follows directly from Proposition and the fact that $\mathscr{H}$ includes A_n as the final state of the jump chain, and so conditioning on $\mathscr{H}$ and A_n is equivalent to conditioning on just $\mathscr{H}$. ∎

Remark 2.3. Unfortunately, extension of Proposition 2.1 to scenarios where the population size is not constant (see Griffiths and Tavaré, 1994c, for example) seems impossible, since the process governing the genealogical history is no longer Markov. However, it is straightforward to extend Proposition 2.1 and its corollary to other contexts where the process governing the genealogical history *is* Markov, such as the constant-sized population cases of the structured coalescent (Herbots, 19997), the Ancestral Recombination Graph (Griffiths and Marjoram, 1997), and the Ancestral Selection Graph (Krone and Neuhauser, 1997).

2.1. Application of Proposition 2.1.

Proposition 2.1 allows us to perform inference for times on genealogical trees, given the history $\mathscr{H}$. For example, consider the time intervals T_k during which there are k lines in the ancestry of our sample ($k = 2, ..., n$). Suppose that the history $\mathscr{H}$ is such that there are m_k mutations during the time interval T_k, so there are $m_k + 1$ events (m_k mutations and 1 split) during T_k. Then (by Proposition 2.1) conditional on $\mathscr{H}$, $T_2, ..., T_n$ are independent, and T_k is distributed as the sum of $m_k + 1$ independent exponential random variables, each with rate parameter λ_k. That is,

$$T_k \mid A_n, \mathscr{H} \sim \Gamma(m_k + 1, \lambda_k), \tag{2}$$

which has density

$$f(t) = \frac{t^{m_k} \lambda_k^{m_k+1} \exp(-\lambda_k t)}{m_k!}, \tag{3}$$

mean $(m_k+1)/\lambda_k$, and variance $(m_k+1)/\lambda_k^2$. In particular, given that no mutations occurred during the history of a sample the times $T_2, ..., T_n$ are independent and T_k is exponentially distributed with rate parameter λ_k; a fact noted by Donnelly *et al.* (1996) for example (see also Tavaré *et al.*, 1997).

More generally, let S denote a time of interest, such as the time to the most recent common ancestor T_{MRCA}, or the age $\mathscr{A}$ of a particular mutation in $\mathscr{H}$. Proposition 2.1 allows us to write S as the sum of independent gamma distributions, $S = X_a + \cdots + X_b$, say, where the X_i are independent and $X_i \sim \Gamma(n_i, \lambda_i)$ for some integers $n_a, ..., n_b$. Although complex, the density of S can be found explicitly (as in Mathai, 1982), and its mean and variance are given by

$$\mathsf{E}(S \mid \mathscr{H}) = \sum_{i=a}^{b} n_i/\lambda_i, \qquad \mathsf{Var}(S \mid \mathscr{H}) = \sum_{i=a}^{b} n_i/\lambda_i^2.$$

Inference for $\mathscr{T}$ given $\mathscr{H}$ is of limited interest in itself as usually both $\mathscr{T}$ *and* $\mathscr{H}$ are unknown, and we wish to perform inference for $(\mathscr{T}, \mathscr{H})$ given A_n. In certain simple cases some progress can be made analytically, as we will see in Section 3 where we consider the case where a single mutation is assumed to have occurred in the ancestry of

the sample. However, in most cases of interest little analytic progress has been made, and so numerical methods are used to perform inference for quantities of interest, such as the distribution of T_{MRCA}, or the distribution of the age $\mathscr{A}$ of a mutation which is known to have occurred in the ancestry of the sample. Recent methodological advances, including the methods developed by Griffiths and Tavaré (see for example Griffiths and Tavaré, 1994b, 1999), and the Markov chain Monte Carlo (MCMC) schemes developed by Kuhner *et al.* (1995, 1998), Wilson and Balding (1998), and Beaumont (1999), allow efficient inference of such quantities to be performed for a range of mutation models under different demographic scenarios. These methods are efficient in that they make use of all the information contained in the data, rather than using some summary of the data, such as the sample homozygosity or the set of pairwise differences (see the section on inference in Donnelly and Tavaré, 1995 for a discussion of the drawbacks of using such summary statistics). The methods are also highly computationally intensive, and it would be helpful if we could use Proposition 2.1 to ease the computational burden. This is straightforward in the case of Griffiths and Tavaré's method applied to a constant-sized population, as we describe and illustrate below. MCMC methods which model the mutations which occur in the genealogy explicitly, which includes the method of Beaumont (1999), but not the methods of Kuhner *et al.* (1995, 1998) and Wilson and Balding (1998), can also make use of Proposition 2.1, as described in Section 2.1.2.

2.1.1. Griffiths and Tavaré's Method

Griffiths and Tavaré (1994a) developed a method of estimating the likelihood $L(\theta) = \mathsf{P}_\theta(A_n)$ (which in the sequel we write as $\mathsf{P}(A_n)$, dropping the explicit dependence on θ for convenience) under a variety of models, by writing down a set of recurrence relations for the likelihood and solving them numerically. Their method (as described by Felsenstein *et al.*, 1999) involves randomly sampling a history consistent with the data A_n, by starting at A_n and randomly sampling events backwards in time. At each stage the next event backwards in time is chosen by assigning each possible event a probability which is proportional to the probability it has under the forwards simulation method underlying Algorithm 2.1. In the original formulation this process continues until a single ancestral line remains (the MRCA of the sample) and so the history is complete. (Refinements of this algorithm stop when two or three ancestral lines remain; all that follows may be applied to these refined algorithms, but for simplicity we concentrate on the original formulation.) Let $Q(\mathscr{H})$ denote the distribution on histories defined by this sampling scheme. Griffiths and Tavaré (1994a) find a function on history space whose expected value under $Q(\mathscr{H})$ is the likelihood $\mathsf{P}(A_n)$ (see their Eq. (15)). For convenience, let $\mathscr{F}(\mathscr{H})$ denote this function. Then the strong law of large numbers gives (almost surely)

$$\mathsf{P}(A_n) = \lim_{M\to\infty} \frac{1}{M} \sum_{i=1}^{M} \mathscr{F}(\mathscr{H}^{(i)}), \tag{4}$$

where $\mathscr{H}^{(1)}, \ldots, \mathscr{H}^{(M)}$ are independent samples from $Q(\mathscr{H})$. A natural approximation to this limit is obtained by taking M to be large:

$$\mathsf{P}(A_n) \approx \frac{1}{M} \sum_{i=1}^{M} \mathscr{F}(\mathscr{H}^{(i)}). \tag{5}$$

In the sequel we will make similar approximations using large M without comment.

Felsenstein *et al.* (1999) point out that Griffiths and Tavaré's method can be viewed as importance sampling (see Ripley, 1987, for background). Indeed we have

$$\begin{aligned}\mathsf{P}(A_n) &= \int \mathsf{P}(A_n \mid \mathscr{H}) \frac{\mathsf{P}(\mathscr{H})}{Q(\mathscr{H})} Q(\mathscr{H})\, d\mathscr{H} \\ &= \int \frac{\mathsf{P}(\mathscr{H})}{Q(\mathscr{H})} Q(\mathscr{H})\, d\mathscr{H} \\ &\qquad [\mathsf{P}(A_n \mid \mathscr{H}) = 1 \text{ on support of } Q] \\ &\approx \frac{1}{M} \sum_{i=1}^{M} \frac{\mathsf{P}(\mathscr{H}^{(i)})}{Q(\mathscr{H}^{(i)})},\end{aligned} \tag{6}$$

where $\mathscr{H}^{(1)}, \ldots, \mathscr{H}^{(M)}$ are independent samples from $Q(\mathscr{H})$. Careful calculation shows that $\mathsf{P}(\mathscr{H})/Q(\mathscr{H})$ is exactly the function $\mathscr{F}(\mathscr{H})$ used by Griffiths and Tavaré, and so (5) and (6) are equivalent. As an aside we note that this interpretation of Q as an importance sampling distribution raises the question of whether Griffiths and Tavaré's scheme is an optimal choice of Q; we show elsewhere (Stephens and Donnelly, 2000) that in many cases alternative choices of Q can substantially reduce the variance of the estimator (6).

Given this importance sampling interpretation of the Griffiths and Tavaré scheme, standard theory (Ripley, 1987) allows us to interpret the discrete distribution with atoms of mass

$$w_i = \frac{\mathsf{P}(\mathscr{H}^{(i)})}{Q(\mathscr{H}^{(i)})} \Big/ \sum_{j=1}^{M} \frac{\mathsf{P}(\mathscr{H}^{(j)})}{Q(\mathscr{H}^{(j)})} = \mathscr{F}(\mathscr{H}^{(i)}) \Big/ \sum_{j=1}^{M} \mathscr{F}(\mathscr{H}^{(j)}) \tag{7}$$

at history $\mathscr{H}^{(i)}$ ($i = 1, ..., M$) as an approximation to the distribution of $\mathscr{H}$ given A_n. The w_i are known as the *importance weights*. If we sample $\mathscr{T}^{(i)}$ from $\mathsf{P}(\mathscr{T} \mid \mathscr{H}^{(i)})$, as given by Proposition 2.1, for $i = 1, ..., M$, then the discrete distribution with atoms of mass w_i at $(\mathscr{T}^{(i)}, \mathscr{H}^{(i)})$ approximates the distribution of $(\mathscr{T}, \mathscr{H})$ given A_n. This in turn induces a distribution on quantities of interest such as T_{MRCA} and $\mathscr{A}$, which is exactly the distribution arrived at by different means by Griffiths and Tavaré (1999) (their Eqs. (4.9) and (4.10)). Estimation of quantities relating to the distribution of T_{MRCA} and $\mathscr{A}$ is then straightforward. For example, the posterior mean of T_{MRCA} may be estimated by

$$\mathsf{E}(T_{\mathrm{MRCA}} \mid A_n) \approx \sum_{i=1}^{M} w_i T_{\mathrm{MRCA}}^{(i)}, \tag{8}$$

where $T_{\mathrm{MRCA}}^{(i)}$ is the time to the most recent common ancestor in the genealogical history $(\mathscr{T}^{(i)}, \mathscr{H}^{(i)})$. Similarly, the probability that T_{MRCA} is in any given set $\mathscr{B}$ may be estimated by

$$\mathsf{P}(T_{\mathrm{MRCA}} \in \mathscr{B}) \approx \sum_{i=1}^{M} w_i \mathrm{I}(T_{\mathrm{MRCA}}^{(i)} \in \mathscr{B}) \tag{9}$$

where $\mathrm{I}(\cdot)$ is an indicator function.

However, simulating $\mathscr{T}^{(i)}$ from $\mathsf{P}(\mathscr{T} \mid \mathscr{H}^{(i)})$ seems unnecessary and wasteful, since we know the form of this distribution (given by Proposition 2.1), and can use this to reduce the sampling variance of estimators for the distribution of T_{MRCA} and $\mathscr{A}$. For example, the distribution of T_{MRCA} can be estimated by

$$\mathsf{P}(T_{\mathrm{MRCA}} \mid A_n) \approx \sum_{i=1}^{M} w_i \mathsf{P}(T_{\mathrm{MRCA}} \mid \mathscr{H}^{(i)}), \tag{10}$$

which by Proposition 2.1 is a mixture of sums of independent gamma distributions. Its density can thus (in theory) be found explicitly (Mathai, 1982), and it is certainly easy to estimate summaries such as the mean and variance using

$$\begin{aligned} \mathsf{E}(T_{\mathrm{MRCA}} \mid A_n) &\approx \sum_{i=1}^{M} w_i \mathsf{E}(T_{\mathrm{MRCA}} \mid \mathscr{H}^{(i)}) \\ &= \sum_{i=1}^{M} w_i \sum_{k=2}^{n} \frac{m_k^{(i)} + 1}{\lambda_k}, \end{aligned} \tag{11}$$

$$\begin{aligned} \mathsf{E}(T_{\mathrm{MRCA}}^2 \mid A_n) &\approx \sum_{i=1}^{M} w_i \mathsf{E}(T_{\mathrm{MRCA}}^2 \mid \mathscr{H}^{(i)}) \\ &= \sum_{i=1}^{M} w_i \sum_{k=2}^{n} \frac{(m_k^{(i)} + 1)(m_k^{(i)} + 2)}{\lambda_k^2}, \end{aligned} \tag{12}$$

and

$$\mathsf{Var}(T_{\mathrm{MRCA}} \mid A_n) = \mathsf{E}(T_{\mathrm{MRCA}}^2 \mid A_n) - [\mathsf{E}(T_{\mathrm{MRCA}} \mid A_n)]^2, \tag{13}$$

where $m_k^{(i)}$ is the number of mutations which occur during time interval T_k in history $\mathscr{H}^{(i)}$.

The estimator (11) will clearly have a smaller sampling error than (8), and requires no additional computation. We compared the variability of these estimators on the Nuu-Chah-Nulth data analysed by Griffiths and Tavaré (1994b) and found that the variance of the estimator (11) was approximately 10 times smaller than that of (2.8), which suggests that up to 10 times fewer iterations are required to obtain the same accuracy when using this improved estimator—a substantial computational saving.

The estimators (11) and (12) thus provide a convenient and efficient way of estimating the mean and variance of T_{MRCA}, and similar ideas can be used to improve efficiency when using the Griffiths–Tavaré algorithm to estimate the means and variances of ages of mutations known to occur in the history of the sample, as in Harding *et al.* (1997) and Hammer *et al.* (1998). It must, however, be admitted that the explicit density given by Mathai (1982) is somewhat daunting, and in practice simulation-based estimators such as (2.9) provide a much more attractive method of estimating more detailed quantities than simply the mean or variance, relating to the distribution of T_{MRCA} or ages of mutations. It is useful then to note that Proposition 2.1 may be used to reduce the simulation variance of estimates such as (9), by simulating several values $(\mathscr{T}^{(i,1)}, ..., \mathscr{T}^{(i,R)})$ from $\mathsf{P}(\mathscr{T} \mid \mathscr{H}^{(i)})$, and replacing estimators such as (9) with

$$\mathsf{P}(T_{\mathrm{MRCA}} \in \mathscr{B}) \approx \sum_{i=1}^{M} w_i \frac{1}{R} \sum_{j=1}^{R} \mathrm{I}(T_{\mathrm{MRCA}}^{(i,j)} \in \mathscr{B}), \tag{14}$$

where $T_{\mathrm{MRCA}}^{(i,j)}$ is the time to the most recent common ancestor in the genealogical history $(\mathscr{T}^{(i,j)}, \mathscr{H})$. The estimator (14) will have lower sampling variance than (9) at the expense of more computational expense in generating the R times for each history. This raises the question of how most efficiently to divide computer time between simulating histories, and simulating times conditional on these histories. Rough calculations suggest that approximately the same amount of computer time should be devoted to simulating times as is devoted to simulating histories; results from simulations (Table I) suggest that

TABLE I

Example of How Efficiency of Estimators such as (14) Can Vary with Choice of M and R.

M:R	180,000:1	100,000:5	65,000:10
Mean squared error	0.0045	0.0037	0.0077

Note. The table is based on estimates of the probability that the mutation numbered 3 in Griffiths and Tavaré (1994b) (their Fig. 2) is between 0 and 0.5 coalescent units old. An accurate estimate of this probability was found using $M = 10{,}000{,}000$ and $R = 20$. This estimate (0.78) was then used as a benchmark to estimate the mean squared error of the shorter runs shown here. Each figure in the table was based on 10 runs of the Griffiths–Tavaré algorithm with different seeds, and each took a very similar amount of time to compute. Of the three columns, the estimated mean squared error is smallest in the middle column (with $M = 100{,}000$ and $R = 5$), for which approximately equal amounts of computer time were spent simulating histories and simulating times, lending support to the rule of thumb given in the text.

this is a reasonable rule of thumb. Further efficiency gains could be obtained (at the expense of greater complexity in the computer code) by devoting more computational effort to simulating times for those histories with the larger weights, that is, letting $R = R_i$ depend on w_i (rough calculations suggest $R_i \propto w_i$ would be a good rule of thumb).

Note that since Proposition 2.1 is easily extended to other contexts where the process governing the genealogical history is Markov, similar methods also apply when analogues of the Griffiths–Tavaré algorithm (or more generally, the kinds of importance sampling method discussed in Stephens and Donnelly (2000)) are used to perform inference for constant-sized structured populations (as in Bahlo and Griffiths, 2000) and for constant-sized populations in the presence of recombination (Griffiths and Marjoram, 1997) or selection (Slade, 2000). Unfortunately, extension to scenarios where the population size is not constant (as in Griffiths and Tavaré, 1994c) is more difficult (see Remark 2.3).

2.1.2. MCMC Methods

The MCMC methods developed by Kuhner *et al.* (1995, 1998) (for sequence data) and Wilson and Balding (1998) (for microsatellite data) do not consider the mutational events occurring in the history of the sample explicitly, and so these methods cannot take advantage of Proposition 2.1 in the way described above for Griffiths and Tavaré's method. However, an MCMC approach which considers these mutational events explicitly by constructing an ergodic Markov chain with stationary distribution $\mathsf{P}(\mathscr{T}, \mathscr{H} \mid A_n)$, as in (for example) Beaumont (1999), *can* take advantage of the proposition. If

$$(\mathscr{T}^{(1)}, \mathscr{H}^{(1)}), \ldots, (\mathscr{T}^{(M)}, \mathscr{H}^{(M)})$$

is a sampled realisation of this Markov chain, then standard theory (see for example Gilks *et al.*, 1996) allows us to interpret the discrete distribution with atoms of equal mass $1/M$ at $(\mathscr{T}^{(1)}, \mathscr{H}^{(1)}), \ldots, (\mathscr{T}^{(M)}, \mathscr{H}^{(M)})$ as an approximation to the distribution of $(\mathscr{T}, \mathscr{H})$ given A_n. This induces a distribution on quantities of interest, such as T_{MRCA} and $\mathscr{A}$, which may be used to estimate the distribution of these quantities given A_n. However, the sampling variability of this Monte Carlo estimator can be reduced using Proposition 2.1 just as with the Griffiths and Tavaré method described in the previous section. For example, we have

$$\mathsf{P}(T_{\mathrm{MRCA}} \mid A_n) \approx \sum_{i=1}^{M} \frac{1}{M} \mathsf{P}(T_{\mathrm{MRCA}} \mid \mathscr{H}^{(i)}), \quad (15)$$

where $\mathsf{P}(T_{\mathrm{MRCA}} \mid \mathscr{H}^{(i)})$ is given by Proposition 2.1 as the sum of independent exponential distributions.

3. THE AGE OF AN ALLELE

We now turn to the problem of inference for $(\mathscr{T}, \mathscr{H})$ conditional on the event $\mathscr{E}$ that a single mutation has occurred at a particular locus of interest during the ancestry of the sample. The genealogical history in Fig. 1 is an example of such an ancestry. Given $\mathscr{E}$, the locus will be biallelic, with one allele being the original *ancestral* allele, and the other being the *mutant* allele. We assume that we know which allele in our sample is ancestral and which is the mutant. Questions of interest include the conditional distribution of the age of the mutant allele given its frequency in a sample or a population, a problem first considered by Kimura and Ohta (1973) and more recently by Griffiths and Tavaré (1998) and Wiuf and Donnelly (1999).

Griffiths and Tavaré (1998) consider the conditional distribution of the age of a unique mutation under general models for the underlying genealogical process. Their analysis proceeds by considering the distribution of the time interval T_k in which the mutation occurred. They derive a very general expression for the conditional density of the age of a mutation and obtain an expression for its mean in the special case of a constant-sized population in the limit as $\theta \to 0$.

Wiuf and Donnelly (1999) study the conditional distribution of the age of a unique mutation, in the limit as

$\theta \to 0$, first in a constant-sized population and then in an exponentially expanding population. In the case of a constant-sized population they describe a method of finding the full conditional distribution of the age, and in the case of an exponentially expanding population they provide expressions for the conditional mean and variance of the age. Their analysis proceeds by considering the conditional distribution of the *genealogy* of a random sample from the population, conditioning separately on (i) the fact that the set of individuals with the mutant allele must have coalesced with each other before any of them coalesces with any individual without the mutant allele; and (ii) the fact that the (unique) mutation occurred. They then use the conditional distribution of the genealogy to infer the distribution of the age of the mutation. However, as Wiuf and Donnelly note in the discussion of their paper, their approach also allows consideration of other aspects of the genealogy which may be of interest. In particular, their approach allows the study of the distribution of the T_{MRCA} of the individuals who carry the mutant allele and the distribution of the length of the chromosomal segment around the site of the mutation that these individuals share identical by descent. In many contexts (such as when developing theory relating to locating genes which play a rôle in complex traits) these quantities will be of more interest than the actual age of the mutant allele.

The limit $\theta \to 0$ considered by both sets of authors is appropriate when considering a mutation which arose at a single pre-specified site at which mutations are assumed to occur at negligible rate. Here we consider the analysis for a single pre-specified site at which the mutation parameter θ is assumed to be non-negligible (note that this is not inconsistent with a unique mutation having occurred at this site in the history of the sample) and treat $\theta \to 0$ as a special case. This analysis is also appropriate when considering a pre-specified region of the genome for which the infinite sites model with non-negligible total mutation rate $\theta > 0$ (and negligible recombination rate) is assumed to apply, and in which we observe only one segregating site. In Section 3.3 we illustrate on a simple example how an analysis which assumes a fixed non-zero θ will tend to give smaller estimates for the age of the allele (in units of $2N$ generations) than an analysis using $\theta \to 0$. Our analysis is based on the use of Proposition 2.1 to derive results for the distribution of the age of the mutant and thus applies only in the special case of a constant-sized panmictic population. Results for very much more general settings can be obtained from Griffiths and Tavaré (1998).

As an aside we note that all the analyses discussed in this section are appropriate for a unique mutation discovered by observation of a single *pre-specified* site; the correct analysis for a mutant allele which was found through more complex ascertainment processes, such as scanning an area of the genome for single nucleotide polymorphisms (SNPs) in a panel of individuals, will depend on the details of the ascertainment process (which may be rather difficult to model).

3.1. Conditional Distribution of Age for $\theta > 0$

We now show how we can use Proposition 2.1 to obtain a parsimonious representation of the full conditional distribution of the age of a mutant allele in the context of a constant-sized panmictic population, and find an explicit formula for its density. The following results may also be obtained using Eq. (5.3) in Griffiths and Tavaré (1998).

THEOREM 3.1 (The Full Conditional Distribution of the Age). *Suppose that of a random sample of n chromosomes from a panmictic constant-sized population, d of them possess a particular mutation which is assumed to have occurred only once during the ancestry of the sample. The full conditional distribution of the age $\mathscr{A}$ of the allele given d has density*

$$f_{\mathscr{A}\mid d}(a) = \sum_{k=2}^{n-d+1} p(k \mid d, \theta) \sum_{i=k}^{n} \alpha_i^{(k,n)} \exp(-\lambda_i a), \quad (16)$$

where

$$p(k \mid d, \theta) = \frac{\dfrac{1}{k-1+\theta}\binom{n-1}{k-1}^{-1}\binom{n-d-1}{k-2}}{\displaystyle\sum_{k_0=2}^{n-d+1} \frac{1}{k_0-1+\theta}\binom{n-1}{k_0-1}^{-1}\binom{n-d-1}{k_0-2}} \quad (17)$$

is the conditional probability that the unique mutation occured while there were k distinct lineages in the ancestry of the sample, and

$$\alpha_i^{(k,n)} = \lambda_k \cdots \lambda_n [(\lambda_k - \lambda_i)\cdots(\lambda_{i-1} - \lambda_i)(\lambda_{i+1} - \lambda_i) \cdots (\lambda_n - \lambda_i)]^{-1} \qquad (2 \leqslant k \leqslant i \leqslant n). \quad (18)$$

The distribution has mean

$$\mathsf{E}(\mathscr{A} \mid d) = \sum_{k=2}^{n-d+1} p(k \mid d, \theta) \sum_{i=k}^{n} \frac{1}{\lambda_i} \quad (19)$$

and variance

$$\mathrm{Var}(\mathscr{A} \mid d) = \sum_{k=2}^{n-d+1} p(k|d,\theta) \left[\left(\sum_{i=k}^{n} \frac{1}{\lambda_i} \right)^2 + \sum_{i=k}^{n} \frac{1}{\lambda_i^2} \right] - \mathsf{E}(\mathscr{A} \mid d)^2. \quad (20)$$

COROLLARY 3.2 (The Large Sample Limit). *The conditional distribution of the age of a mutation, given that it occurs at frequency f in the population, has density*

$$f_{\mathscr{A}|f}(a) = \lim_{n\to\infty} \sum_{i=2}^{n} \sum_{k=2}^{i} \tilde{p}(k \mid f, \theta)\, \alpha_i^{(k,n)} \exp(-\lambda_i a), \quad (21)$$

where

$$\tilde{p}(k \mid f, \theta) = \frac{\frac{k-1}{k-1+\theta}(1-f)^{k-2}}{\sum_{k_0=2}^{\infty} \frac{k_0-1}{k_0-1+\theta}(1-f)^{k_0-2}} \quad (22)$$

is the conditional probability that the unique mutation occured while there were k distinct lineages in the ancestry of the sample, and the $\alpha_i^{(k,n)}$ $(2 \leqslant k \leqslant i \leqslant n)$ *are given by* (18). *In practice* (21) *may be approximated by taking n to be large. The conditional mean and variance are given by*

$$\mathsf{E}(\mathscr{A} \mid f) = \sum_{k=2}^{\infty} \tilde{p}(k \mid f, \theta) \sum_{i=k}^{\infty} \frac{1}{\lambda_i}, \quad (23)$$

$$\mathrm{Var}(\mathscr{A} \mid f) = \sum_{k=2}^{\infty} \tilde{p}(k \mid f, \theta) \left[\left(\sum_{i=k}^{\infty} \frac{1}{\lambda_i} \right)^2 + \sum_{i=k}^{\infty} \frac{1}{\lambda_i^2} \right] - \mathsf{E}(\mathscr{A} \mid f)^2. \quad (24)$$

This follows from Theorem 3.1 *by letting* $d, n \to \infty$ *while* $d/n \to f$, *as in Griffiths and Tavaré* (1998) *and Wiuf and Donnelly* (1999).

Proof of Theorem 3.1. For notational convenience we will write probability conditional on $\mathscr{E}$ as $\mathsf{P}_{\mathscr{E}}$. Given $\mathscr{E}$, the mutation must have occurred during one of the time intervals $T_2, ..., T_n$. Let $\mathscr{D}_k$ denote the event that a single mutation occurred during T_k and no mutation occurred in any of the other time intervals. It is clear from Algorithm 2.1 (see also Watterson, 1975, Eq. (2.16)) that the number of mutations in time intervals $T_2, ..., T_k$ are (a priori) independent, with the number in T_k having geometric distribution with parameter $(k-1)/(k-1+\theta)$:

$$\mathsf{P}(m \text{ mutations in } T_k) = \left(\frac{k-1}{k-1+\theta}\right)\left(\frac{\theta}{k-1+\theta}\right)^m. \quad (25)$$

Thus we have

$$\begin{aligned} \mathsf{P}_{\mathscr{E}}(\mathscr{D}_k) &\propto \mathsf{P}(\mathscr{D}_k \cap \mathscr{E}) \\ &= \mathsf{P}(\mathscr{D}_k) \\ &= \frac{1}{(1+\theta)} \frac{2}{(2+\theta)} \cdots \frac{k-1}{(k-1+\theta)} \\ &\quad \times \frac{\theta}{(k-1+\theta)} \frac{k}{(k+\theta)} \cdots \frac{n-1}{(n-1+\theta)} \\ &\propto \frac{1}{k-1+\theta}. \end{aligned} \quad (26)$$

Let $\mathscr{M}_d$ be the event that d of the n individuals in the sample have the mutant gene. Given $\mathscr{D}_k$, the mutation occurs on a line chosen at random from the k in T_k, and the number of mutants is given by the number of descendants of this line in our sample. The joint distribution of the number of descendents of the ancestral lines during T_k was studied by Kingman (1982) and (as noted by Griffiths and Tavaré, 1998) can be characterised as the distribution of the number of balls in k boxes when n balls are dropped into these boxes uniformly at random, conditional on at least one ball being in each box. The number of ways of assigning balls to boxes in such a way is

$$\binom{n-1}{k-1}$$

(see Feller, 1968, pp. 38–39), and it follows that the number of ways of doing it with d in a given box is

$$\binom{n-d-1}{k-2}.$$

Since all assignments are equally likely, the probability there are d balls in a box chosen at random, and hence the probability that d of the sample have the mutant gene given that the mutation occurred in T_k, is given by

$$\mathsf{P}_{\mathscr{E}}(\mathscr{M}_d \mid \mathscr{D}_k) = \binom{n-1}{k-1}^{-1} \binom{n-d-1}{k-2}, \quad 1 \leqslant d \leqslant n-k+1, \quad (27)$$

which is Eq. (1.9) of Griffiths and Tavaré (1998). Combining (26) and (27) we obtain the conditional distribution of the number of ancestral lines when the mutation occurs, given d,

$$\begin{aligned} p(k|d,\theta) &= \mathsf{P}_{\mathscr{E}}(\mathscr{D}_k \mid \mathscr{M}_d) \\ &\propto \mathsf{P}_{\mathscr{E}}(\mathscr{D}_k \cap \mathscr{M}_d) \\ &\propto \mathsf{P}_{\mathscr{E}}(\mathscr{D}_k)\, \mathsf{P}_{\mathscr{E}}(\mathscr{M}_d \mid \mathscr{D}_k) \\ &\propto \frac{1}{k-1+\theta}\binom{n-1}{k-1}^{-1}\binom{n-d-1}{k-2}, \\ &\qquad 2 \leqslant k \leqslant n-d+1, \end{aligned} \tag{28}$$

which gives (17).

By Proposition 2.1, the distribution of the age $\mathscr{A}$ of the mutation given $\mathscr{D}_k$ can be written as

$$\mathscr{A} \mid \mathscr{D}_k \sim W_k + \cdots + W_n, \tag{29}$$

where the W_i are independent and

$$W_i \sim \mathrm{Exp}(\lambda_i), \tag{30}$$

where $\lambda_i = i(i-1+\theta)/2$. For each i, W_i may be interpreted as the time before the split from i lines to $i+1$ ancestral lines, as illustrated in Fig. 1. Given $\mathscr{D}_k$, the number of mutants depends on only which lines in the ancestry split, which is independent of the times between splits. Therefore $\mathscr{A}$ and $\mathscr{M}_d$ are conditionally independent given $\mathscr{D}_k$, and

$$\mathscr{A} \mid \mathscr{D}_k, \mathscr{M}_d \sim W_k + \cdots + W_n. \tag{31}$$

Thus the distribution of $\mathscr{A}$ is affected by d only through its effect on the distribution of the time interval in which the mutation occurred. Expressions (28) and (31), combined with

$$\mathsf{P}_{\mathscr{E}}(\mathscr{A} \mid \mathscr{M}_d) = \sum_{k=2}^{n} \mathsf{P}_{\mathscr{E}}(\mathscr{D}_k \mid \mathscr{M}_d)\, \mathsf{P}_{\mathscr{E}}(\mathscr{A} \mid \mathscr{D}_k, \mathscr{M}_d), \tag{32}$$

give a parsimonious representation of the distribution of $\mathscr{A}$ given $\mathscr{E}$ and $\mathscr{M}_d$. In particular it is a mixture of sums of independent exponential distributions, and so its density is easy to compute (using the formula for the density of a sum of independent exponential distributions given by Feller, 1970, p. 40, Problem 12) which gives (16). (Alternatively, this density may be obtained from results in Section 5.1 of Tavaré (1984).)

Expressions (19) and (20) follow from

$$\begin{aligned} \mathsf{E}(\mathscr{A} \mid \mathscr{M}_d) &= \sum_{k=2}^{n-d+1} \mathsf{P}_{\mathscr{E}}(\mathscr{D}_k \mid \mathscr{M}_d)\, \mathsf{E}(\mathscr{A} \mid \mathscr{D}_k, \mathscr{M}_d) \\ &= \sum_{k=2}^{n-d+1} p(k \mid d, \theta) \sum_{i=k}^{n} \frac{1}{\lambda_i}, \\ \mathsf{E}(\mathscr{A}^2 \mid \mathscr{M}_d) &= \sum_{k=2}^{n-d+1} \mathsf{P}_{\mathscr{E}}(\mathscr{D}_k \mid \mathscr{M}_d)\, \mathsf{E}_{\mathscr{E}}(\mathscr{A}^2 \mid \mathscr{D}_k, \mathscr{M}_d) \\ &= \sum_{k=2}^{n-d+1} p(k \mid d, \theta) \sum_{i=k}^{n} \sum_{j=k}^{n} \mathsf{E}_{\mathscr{E}}(W_i W_j) \\ &= \sum_{k=2}^{n-d+1} p(k \mid d, \theta) \left[\left(\sum_{i=k}^{n} \frac{1}{\lambda_i} \right)^2 + \sum_{i=k}^{n} \frac{1}{\lambda_i^2} \right], \end{aligned}$$

and

$$\mathsf{Var}(\mathscr{A} \mid \mathscr{M}_d) = \mathsf{E}(\mathscr{A}^2 \mid \mathscr{M}_d) - \mathsf{E}(\mathscr{A} \mid \mathscr{M}_d)^2. \quad \blacksquare$$

3.2. The Limit $\theta \to 0$

Results for the conditional distribution of the age of the mutation in the limit $\theta \to 0$ considered by Griffiths and Tavaré (1998) and Wiuf and Donnelly (1999), can be obtained (for both the sample and population cases) by setting $\theta = 0$ in the expressions (16)–(24).

In particular, (17) becomes

$$\lim_{\theta \to 0} p(k \mid d, \theta) = \binom{n-1}{d}^{-1}\binom{n-k}{d-1}, \tag{33}$$

and we have

$$\begin{aligned} \lim_{\theta \to 0} \sum_{i=k}^{n} \frac{1}{\lambda_i} &= \sum_{i=k}^{n} \frac{2}{i(i-1)} \\ &= 2\,\frac{n-k+1}{n(k-1)}, \end{aligned} \tag{34}$$

from which we obtain

$$\begin{aligned} &\lim_{\theta \to 0} \mathsf{E}_{\mathscr{E}}(\mathscr{A} \mid \mathscr{M}_d) \\ &\quad = 2 \sum_{k=2}^{n-d+1} \binom{n-1}{d}^{-1}\binom{n-k}{d-1}\frac{n-k+1}{n(k-1)}. \end{aligned} \tag{35}$$

This is the same as the expression obtained by Griffiths and Tavaré (1998) (the limits on their sum are given as 2 to n, but the terms of the sum corresponding to $k > n-d+1$ are zero)

In the large sample limit case, as $\theta \to 0$, (22) gives a geometric distribution for the number of lines in the ancestry at the time of the mutation,

$$\lim_{\theta \to 0} \tilde{p}(k \mid f, \theta) = f(1-f)^{k-2} \qquad (k \geqslant 2), \qquad (36)$$

and (34) gives

$$\lim_{n \to \infty} \lim_{\theta \to 0} \sum_{i=k}^{n} \frac{1}{\lambda_i} = \frac{2}{k-1}, \qquad (37)$$

from which we obtain

$$\begin{aligned} \mathsf{E}_{\mathscr{E}}(\mathscr{A} \mid \mathscr{M}_f) &= \sum_{k=2}^{\infty} f(1-f)^{k-2} \frac{2}{k-1} \\ &= -\frac{2f}{1-f} \log f, \end{aligned} \qquad (38)$$

which is the celebrated result of Kimura and Ohta (1973).

3.3 The Effect of Using $\theta > 0$

In order to illustrate the effect of assuming different values of θ, consider the distribution of the age of a mutation which is present in 10 individuals of a random sample of size 20 from a large panmictic population which has evolved with constant size over many generations. Figure 2 compares the density of this age, as given by the above analyses, for $\theta = 4.0$ (solid line), $\theta = 2.0$ (dotted line), and $\theta \to 0.0$ (dashed line). We can see that our beliefs about the value of θ have a substantial effect on the distribution of the age of the mutation. In general, taking $\theta \to 0.0$ will tend to move the mutation further up the tree (see Eq. (28)), and make the times between events longer (Eq. (30)), thus causing us to believe the mutation is older than if we use larger values of θ. The situation is further complicated by the effect that different estimates of θ may have on estimates of N (since $\theta = 2N\mu$), which is the time scale on which times are being measured. In practice there is considerable uncertainty in any estimate of θ, and it is thus important that any analysis assesses how sensitive the results are to different assumptions for θ (see Tavaré *et al.*, 1997, for further discussion).

ACKNOWLEDGMENTS

I thank Peter Donnelly, David Balding, Robert Griffiths, Jonathan Pritchard, Simon Tavaré, and an anonymous reviewer for helpful comments on earlier versions of this paper. The author was supported by a grant from the University of Oxford. Part of this paper was written while the author was resident at the Isaac Newton Institute for Mathematical Sciences, Cambridge, UK.

REFERENCES

Bahlo, M., and Griffiths, R. C. 2000. Gene trees in subdivided populations, *Theoret. Popul. Biol.*, to appear.

Beaumont, M. 1999. Detecting population expansion and decline using microsatellites, *Genetics* **153**, 2013–2029.

Donnelly, P. 1986. Partition structures, Polya urns, the Ewens sampling formula, and the age of alleles, *Theoret. Popul. Biol.* **30**, 271–288.

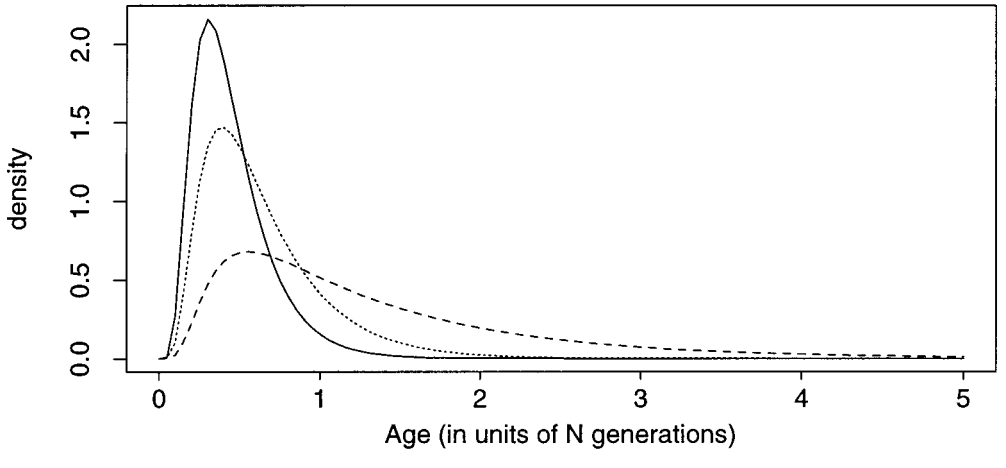

FIG. 2. Distribution of the age of a unique mutation, conditional on it being observed in 10 individuals of a random sample of size 20 from a large panmictic constant-sized population. The three lines show densities for *Solid line*: $\theta = 4.0$, *Dotted line*: $\theta = 2.0$, and *Dashed line*: $\theta \to 0.0$, as given by Eqs. (16)–(18).

Donnelly, P., and Kurtz, T. G. 1996. The asymptotic behaviour of an urn model arising in population genetics, *Stochst. Process. Appl.* **64**, 1–16.

Donnelly, P., and Tavaré, S. 1986. The ages of alleles and a coalescent, *Adv. Appl. Probab.* **18**, 1–19.

Donnelly, P., and Tavaré, S. 1995. Coalescents and genealogical structure under neutrality, *Ann. 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.

Ethier, S. N., and Griffiths, R. C. 1987. The infinitely-many-sites model as a measure-valued diffusion, *Ann. Probab.* **15**, 515–545.

Feller, W. 1968. "An Introduction to Probability Theory and Its Applications," 3rd ed., Vol. 1, Wiley, New York.

Feller, W. 1970. "An Introduction to Probability and Its Applications," 3rd ed., Vol. II, Wiley, New York.

Felsenstein, J., Kuhner, M. K., Yamato, J., and Beerli, P. 1999. "Likelihoods on Coalescents: A Monte Carlo Sampling Approach to Inferring Parameters from Population Samples of Molecular Data," *in* "Statistics in Molecular Biology and Genetics" (F. Seillier-Moiseiwitsch, Ed.), pp. 163–185, Institute of Mathematical Statistics and American Mathematics Society, Haywood, California.

Gilks, W. R., Richardson, S., and Spiegelhalter, D. J. 1996. Introducing Markov chain Monte Carlo, *in* "Markov Chain Monte Carlo in Practice" (W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, Eds.), Chapman & Hall, London.

Griffiths, R. C., and Marjoram, P. 1997. An ancestral recombination graph, *in* "Progress in Population Genetics" (P. Donnelly and S. Tavaré, Eds.), Springer-Verlag, New York.

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

Griffiths, R. C., and Tavaré, S. 1994b. Ancestral inference in population genetics, *Statist. Sci.* **9**, 307–319.

Griffiths, R. C., and Tavaré, S. 1994c. Sampling theory for neutral alleles in a varying environment, *Philos. Trans. R. Soc. London B* **344**, 403–410.

Griffiths, R. C., and Tavaré, S. 1998. The age of a mutation in a general coalescent tree, *Stochast. Models* **14**, 273–295.

Griffiths, R. C., and Tavaré, S. 1999. The ages of mutations in gene trees, *Ann. Appl. Prob.* **9**, 567–590.

Hammer, M., Karafet, T., Rasanayagam, A., Wood, E., Altheide, T., Jenkins, T., Griffiths, R., Templeton, A. R., and Zegura, S. 1998. Out of Africa and back again: Nested cladistic analysis of human Y chromosome variation, *Mol. Biol. Evol.* **15**(4), 427–441.

Harding, R., Fullerton, S. M., Griffiths, R., Bond, J., Cox, M., Schneider, J., Moulin, D., and Clegg, J. 1997. Archaic African and Asian lineages in the genetic ancestry of Modern humans, *Am. J. Hum. Gen.* **60**, 772–789.

Herbots, H. 1997. The structured coalescent, *in* "Progress in Population Genetics" (P. Donnelly and S. Tavaré, Eds.), Springer-Verlag, New York.

Hudson, R. R. 1991. Gene genealogies and the coalescent process, *in* "Oxford Surveys in Evolutionary Biology" (D. Futuyma and J. Antonovics, Eds.), Vol. 7, pp. 1–44, Oxford Univ. Press, Oxford.

Kimura, M., and Ohta, T. 1973. The age of a neutral mutation persisting in a finite population, *Genetics* **75**, 199–212.

Kingman, J. F. C. 1982. The coalescent, *Stochast. Process. Appl.* **13**, 235–248.

Krone, S. M., and Neuhauser, C. 1997. Ancestral processes with selection, *Theor. Popul. Biol.* **51**, 210–237.

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

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

Mathai, A. M. 1982. Storage capacity of a dam with gamma type inputs, *Ann. Inst. Statist. Math.* **34**, 591–597.

Ripley, B. D. 1987. "Stochastic Simulation," Wiley, New York.

Slade, P. 2000. Simulation of selected genealogies, submitted for publication.

Stephens, M., and Donnelly, P. 2000. Inference in molecular population genetics, *J. Roy. Statist. Soc. Ser. B*, to appear.

Tavaré, S. 1984. Line-of-descent and genealogical processes, and their applications in population genetics models, *Theor. Popul. Biol.* **26**, 119–164.

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

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

Watterson, G. A. 1977. Reversibility and the age of an allele II. Two-allele models, with selection and mutation, *Theor. Popul. Biol.* **12**, 179–196.

Watterson, G. A., and Guess, H. A. 1977. Is the most frequent allele the oldest? *Theor. Popul. Biol.* **11**, 141–160.

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

Wiuf, C., and Donnelly, P. J. 1999. Conditional genealogies and the age of a neutral mutant, *Theor. Popul. Biol.* **56**, 183–201.