Theoretical Population Biology **52**, 216–223 (1997)
Article No. TP971332

On the Genealogy of a Rare Allele

Bruce Rannala*

Department of Integrative Biology, University of California, Berkeley, California 94720-3140

Received October 31, 1996

The gene genealogy is derived for a rare allele that is descended from a mutant ancestor that arose at a fixed time in the past. Following Thompson (1976, *Amer. J. Human Genet.* 28, 442–452), the fractional linear branching process is used as a model of the demography of a rare allele. The model does not require the total population size to be constant or the mutant class to be neutral; so long as individuals in the class are selectively equivalent, the class as a whole may have a selective advantage, or disadvantage, relative to other alleles in the population. An exact result is given for the joint probability distribution of the coalescence times among a sample of alleles descended from the mutant. A method is described for rapidly simulating these coalescence times. The relationship between the genealogical structure of a discrete generation branching process and a continuous generation birth–death process is elucidated. The theory may be applied to the problem of estimating the ages of rare nonrecurrent mutations. © 1997 Academic Press

INTRODUCTION

The dynamics of a newly-arisen mutant allele in a large population may often be accurately modeled using a branching process (BP) in which each individual gives rise to an independent and identically distributed number of offspring in each generation (Fisher, 1922, 1930; Haldane, 1927). Theoretical studies of BPs in population genetics have focused mainly on the distribution of the number of copies of a particular (nonrecurrent) mutant type that are maintained in a population over time. In particular the probability of ultimate extinction, or the frequency distribution conditional on non-extinction, has been a focus of interest (see Ewens, 1979). Thompson (1976) studied the fractional linear BP as a model of the dynamics of a rare allele, for human populations, and derived a maximum likelihood method for inferring allele age based on the number of copies found in a population.

* E-mail: bruce@mws4.biol.berkeley.edu.

Much of the information about the history of a mutant allele is contained in the intraallelic genealogy of the allele (Slatkin, 1996) rather than in its frequency. If the genealogy of a mutant allele that is descended from a single ancestral copy may be reconstructed from the configuration of mutations (at completely linked sites) among present-day descendents, it is possible to make inferences about parameters related to the demographic history of the allelic class such as the time at which the mutant first arose, the population growth rate, or the level of selection affecting the mutant (Slatkin and Rannala, 1997). A necessary first step in this process is the derivation of the distribution of the times, in the past, at which present-day lineages coalesce to common ancestral lineages for a BP.

The structure of the gene coalescence process has previously been studied for a random sample of alleles from a large population conforming to a class of exchangeable models that includes the Wright–Fisher model as a special case (Kingman, 1982). The usual coalescent

0040-5809/97 $25.00

theory does not apply, however, when we consider a particular allelic class and condition on the time at which the mutant ancestor from which the class is descended first arose. The birth–death process provides a good approximation to this distribution (Slatkin and Rannala, 1997), but may be inaccurate when the mutant arose very recently and the population size is increasing, or the allele has experienced strong positive selection. In such cases, multiple coalescence events might be expected to occur within a single generation. The BP, which allows for multiple coalescence events in any generation, may provide a better approximation of the allelic dynamics generating the distribution of coalescence times in such cases.

The distribution of the number of offspring for a supercritical BP (i.e., with mean offspring number greater than one), in each generation, that ultimately survive to leave descendents appears to have been first studied by Harris (1948). This distribution was further considered by Fleischmann and Siegmund–Schultze (1977) who derived the distribution of offspring in each generation that leave descendents at a fixed time, t, in the future for a critical BP (i.e., with mean offspring number equal to one); they referred to this as the "reduced process." These authors showed that the genealogy of a critical branching process converges to the genealogy of a continuous-time Yule pure-birth process (Yule, 1925), with an appropriate rescaling of time, in the limit as the time of origin of the process tends to infinity. O'Connell (1993, 1995) extended these results to consider the limiting Yule process associated with slightly supercritical BPs. In this paper, I describe an exact result for the probability distribution of the times at which a sample of lineages coalesces to common ancestral lineages under a BP with a modified geometric offspring distribution, and conditional on the time in the past at which the ultimate ancestor of the sampled lineages first arose.

1. BRANCHING PROCESSES

I consider a Galton–Watson BP, initiated from a single lineage at time t_1 in the past, and observed at time present. In any generation, each lineage gives rise to an independent and identically distributed number of offspring and then dies. If n_t is the number of lineages that exist at generation t in the past, then the process may be described using a Markov chain with one-step transition probabilities:

$$P_{ij} = \mathsf{Pr}(n_{t-1} = j | n_t = i) \qquad \text{for} \quad n_t = 0, 1, ... \tag{1}$$

Let p_k be the probability that a single lineage has k descendents after one generation. Then the probability generating function (pgf) is defined as

$$f(s) = \sum_{k=0}^{\infty} p_k s^k \qquad |s| \leqslant 1. \tag{2}$$

Lineages reproduce independently of one another, and the pgf for P_{ij} is then a convolution over the offspring of the i lineages,

$$f(s)^{(i)}. \tag{3}$$

The pgf for the offspring distribution of a lineage born t generations in the past is

$$f_t(s) = f(f_{t-1}(s)) = f(f(\cdots f(s))), \tag{4}$$

where $f_t(s)$ is the tth iterate of the pgf (see Harris, 1963). In general, Eq. (4) will have no known solution, although for certain offspring distributions (such as the modified geometric; see below) explicit solutions are available. In this paper, I focus on the structure of the genealogy of the present-day descendents of a BP initiated t_1 generations in the past.

1.1. Rare Alleles in the Fisher–Wright Model

The Fisher–Wright demographic model is characterized by the property that, in any generation, each individual gives rise to a single descendent, on average, and then dies; the total population size remains constant. With this reproductive scheme, the distribution of the number of descendent copies for each of $2N$ alleles at the next generation is

$$\mathsf{Pr}(\mathbf{n}) = \binom{2N}{n_1, n_2, ..., n_{2N}} \prod_{i=1}^{2N} \frac{1}{\{2N\}^{n_i}}, \tag{5}$$

where $\mathbf{n} = \{n_1, n_2, ..., n_{2N}\}$ is a vector of the number of offspring for each allele. Note that n_i is the number of copies of the ith allele that survive to reproduce at the next generation and $0 \leqslant n_i \leqslant 2N$. The marginal distribution of the offspring number for any particular allele follows a binomial distribution and for large N an excellent approximation is the Poisson probability

$$\mathsf{Pr}(n_i) = \frac{e^{-1}}{n_i!}. \tag{6}$$

A BP in which each allele has an independent offspring distribution of this form was first considered as a model

of the dynamics of a rare allele by Fisher (1922). It is easy to determine the conditions under which this approximation is valid: suppose that n alleles are descended from a single mutant allele that arose at time t in the past; if at any generation $t' \leqslant t$ each allele gives birth to an independent number of offspring, as required for a BP model, then the total number of descendents of the mutant, in the next generation, must be a convolution over the offspring distributions of the n alleles. If each has an offspring distribution described by the Poisson probability of Eq. (6) then, by a standard result, the total number of descendent alleles in the next generation follows a Poisson distribution with parameter n.

In other words, the BP approximation is valid if, in any generation t' since the allele first arose, the offspring distribution is Poisson($n_{t'}$). For the Fisher–Wright model, the number of descendents of a particular mutant class at generation $t+1$, given that there are n_t copies of the mutant allele at generation t, is

$$\mathsf{Pr}(n_{t+1}|n_t) = \binom{2N}{n_{t+1}} \left[\frac{n_t}{2N}\right]^{n_{t+1}} \left[\frac{2N-n_t}{2N}\right]^{2N-n_{t+1}}. \quad (7)$$

This distribution is approximately Poisson(n_t) if the following conditions are satisfied:

$$\left(1-\frac{n_t}{2N}\right)^{2N} \approx e^{-n_t} \quad (8)$$

$$\frac{2N!}{(2N-n_{t+1})!(2N)^{n_{t+1}}} \approx 1 \quad (9)$$

$$\left(1-\frac{n_t}{2N}\right)^{n_{t+1}} \approx 1. \quad (10)$$

These equalities are satisfied in the limit of large N and small n_t, n_{t+1} and this is the basis for a BP approximation of the dynamics of a rare allele under the Fisher–Wright model.

1.2. Modified Geometric Distribution of Offspring

If the probability of k offspring for an individual of a particular allelic class is of the modified geometric form:

$$p_k = bc^{k-1}, \qquad k = 1, 2, ..., \quad (11)$$

for $b, c > 0$ and $b \leqslant 1-c$, with $p_0 = 1 - b/(1-c)$, the probability distribution of the number of offspring at an arbitrary number of generations later may be explicitly derived using probability generating functions (see, e.g., Harris, 1963). A BP constructed using this so-called fractional linear offspring distribution has previously been used to model the dynamics of a rare type (such as a surname, or a mutant allele) in a human population (see Lotka, 1931; Thompson, 1976) and appears to provide a reasonable approximation. In this paper, I consider explicit results for the genealogy of a BP using an offspring distribution of this form.

Here, I review some standard results, to be used later in the paper, for the evolution over time of a BP with a modified geometric offspring distribution. More detailed descriptions may be found in Harris (1963) and Thompson (1976). The mean and variance of the number of offspring per individual, in any generation, are

$$\mathsf{E}(k) = \frac{b}{(1-c)^2}, \quad (12)$$

$$\mathsf{V}(k) = \frac{b(1-b-c^2)}{(1-c)^4}. \quad (13)$$

The probability that a lineage arising at time t in the past has k descendents at present is (see Thompson, 1976)

$$p_t(k) = m^t M_t^2 (1-M_t)^{k-1} \qquad \text{for} \quad k = 1, 2, ..., \quad (14)$$

and

$$p_t(0) = 1 - m^t M_t, \quad (15)$$

where

$$M_t = \frac{1-s_0}{m^t - s_0}, \quad (16)$$

$$m = \frac{b}{(1-c)^2}, \quad (17)$$

$$s_0 = 1 - \frac{(m-1)(1-c)}{c}. \quad (18)$$

If $m = 1$ then $s_0 = 1$, $M_t = h/(h+t)$, and these equations become

$$p_t(k) = \frac{t^{k-1}h^2}{(h+t)^{k+1}} \qquad \text{for} \quad k = 1, 2, ... \quad (19)$$

and

$$p_t(0) = \frac{t}{h+t}, \tag{20}$$

where $h = (1-c)/c$. These results apply to the collection of descendent lineages in the population as a whole. In most cases of interest, the observed lineages are a sample from the total population. In the next section, I derive the sampling theory for a BP of this type.

1.3. Branching Process with Sampling

I now take into account the effects of sampling in calculating the transition probabilities for a BP. Let $\rho = n/2N$ be the fraction of lineages that are sampled at time present, where $0 \leqslant n \leqslant 2N$ is the sample size and N is the (diploid) population size when considering a rare mutant allele in a large population. It is assumed that present-day lineages are sampled at random, with replacement, or sampling is without replacement and $n \ll 2N$. The probability that y lineages in the sample are descended from a lineage that first arose at time t in the past is

$$p_t^*(y) = \sum_{i=y}^{\infty} \{\mathsf{Pr}(y|i) p_t(i)\}, \qquad \text{for all} \quad i, y \geqslant 1, \tag{21}$$

where

$$\mathsf{Pr}(y|i) = \binom{i}{y} \rho^y (1-\rho)^{i-y}, \qquad \text{for all} \quad y \leqslant i. \tag{22}$$

In the special case of the modified geometric offspring distribution considered above, this gives

$$\begin{aligned} p_t^*(y) &= \sum_{i=y}^{\infty} \left\{ m^t M_t^2 (1-M_t)^{i-1} \binom{i}{y} \rho^y (1-\rho)^{i-y} \right\}, \\ &\qquad \text{for all} \quad i, y \geqslant 1, \\ &= \frac{M_t^2 m^t \rho^y (1-M_t)^{y-1}}{(M_t + \rho(1-M_t))^{y+1}}, \\ &\qquad \text{for all} \quad y = 1, 2, \ldots, \end{aligned} \tag{23}$$

and

$$p_t^*(0) = \frac{M_t + \rho[1 - M_t(1+m^t)]}{M_t + \rho[1-M_t]}. \tag{24}$$

It is easily shown that if $\rho = 1$, Eqs. (23) and (24) reduce to Eqs. (14) and (15) as required. In the remainder of this paper, I make use of the transition probabilities for the BP with sampling as this is the most general case; the special case of a complete population sample is obtained from these results by setting $\rho = 1$.

2. GENEALOGY OF A BRANCHING PROCESS

Let $t = 0, 1, 2, \ldots, t_1$ denote the generations of a BP, where $t = 0$ at present, $t = 1$ at the previous generation (i.e., when lineages alive at present were born), and $t = t_1$ at the time when the lineage ancestral to all present-day lineages first arose (see Fig. 1). Let $L(t)$ be the number of lineages that coalesce at generation t, where $0 \leqslant L(i) \leqslant N-1$, and $N = \sum_{j=1}^{t_1} L(j) + 1$ is the number of lineages sampled. The configuration of lineages that coalesce at times $t = 1, 2, \ldots, t_1$ in the past may be represented as a vector $\boldsymbol{L} = \{\mathrm{L}(1), \mathrm{L}(2), \ldots, \mathrm{L}(\mathrm{t}_1)\}$. The total number of possible configurations, given N and t_1, is $1/(t_1-1)! \prod_{i=1}^{t_1-1} (N+i-1)$, which may be very large. I now derive the joint distribution of $\mathbf{L}$.

THEOREM 1. *For a BP with a modified geometric offspring distribution, the probability of the configuration of coalescence events,* $\mathbf{L}$, *conditional on the number of lineages sampled,* N, *is*

$$\begin{aligned} &\mathsf{Pr}(\mathbf{L}|N) \\ &= (N-1)! \prod_{j=1}^{t_1} \left\{ \frac{1}{L(j)!} \left[\frac{p_j(2|1) - p_{j-1}(2|1)}{p_{t_1}(2|1)} \right]^{L(j)} \right\}, \end{aligned} \tag{25}$$

where

$$\begin{aligned} p_t(2|1) &= \frac{p_t(2)}{p_t(1)}, \\ &= \frac{\rho(1-M_t)}{M_t + \rho(1-M_t)}. \end{aligned} \tag{26}$$

Proof. I first derive the joint probability of $\mathbf{L}$ and N. An initial lineage that arose at time t_1 in the past has exactly one descendent in the sample with probability $p_{t_1}(1)$. An additional descendent is born to the lineage, at generation t, and leaves exactly one descendent lineage at present with probability

$$p_t(2|1) - p_{t-1}(2|1), \tag{27}$$

where $p_t(2|1)$, defined above, is the probability that a lineage leaving a single descendent at present gives rise to

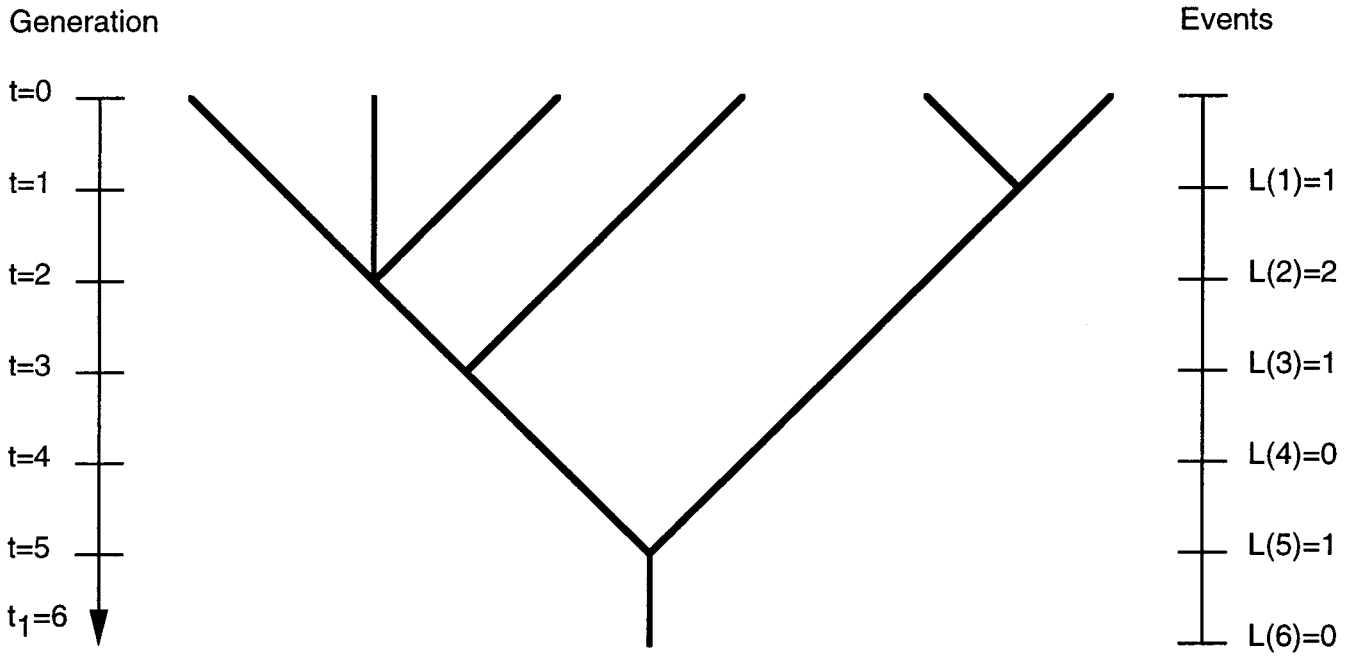

FIG. 1. The genealogy for a sample of lineages from a branching process. The generation times are listed on the left, with generations increasing as the process moves backward in time. The configuration of coalescence events for the genealogy depicted is given at the right. The complete configuration of the coalescence times for this genealogy is $\mathbf{L} = \{1, 2, 1, 0, 1, 0\}$.

a second lineage, born no earlier than generation t, that also leaves a single descendent at present. The difference $p_t(2|1) - p_{t-1}(2|1)$ is the probability that the additional lineage arose precisely at generation t. A total of $L(t)$ additional lineages arise at generation t with probability

$$\{p_t(2|1) - p_{t-1}(2|1)\}^{L(t)}. \tag{28}$$

If $L(t)$ additional lineages arise at generation t, each leaving a single descendent at present, all lineages that leave descendents at present and are alive at $t+1$ are equally likely to have produced the new lineages. The joint probability of the coalescence events is then

$$\Pr(\mathbf{L}, N) = p_{t_1}(1) \binom{N-1}{L(1), L(2), ..., L(t_1)} \times \prod_{j=1}^{t_1} \{p_j(2|1) - p_{j-1}(2|1)\}^{L(j)}. \tag{29}$$

The probability that N lineages occur in the present-day sample is

$$\Pr(N) = \frac{M_{t_1}^2 m^{t_1} \rho^N (1 - M_{t_1})^{N-1}}{(M_{t_1} + \rho(1 - M_{t_1}))^{N+1}} = p_{t_1}(1) \times p_{t_1}(2|1)^{N-1}, \tag{30}$$

and the probability of observing configuration $\mathbf{L}$, conditional on N, is then

$$\Pr(\mathbf{L}|N) = \frac{\Pr(\mathbf{L}, N)}{\Pr(N)} = \binom{N-1}{L(1), L(2), ..., L(t_1)} \prod_{j=1}^{t_1} \left\{\frac{p_j(2|1) - p_{j-1}(2|1)}{p_{t_1}(2|1)}\right\}^{L(j)}, \tag{31}$$

which is the required distribution. ∎

Note that the distribution of coalescence events, $\mathbf{L}$, may be represented as the joint distribution of the order statistics (i.e., the rank-ordered observations) of $N-1$ independent and identically distributed random variables with the common distribution

$$g(j|t_1) = \frac{p_j(2|1) - p_{j-1}(2|1)}{p_{t_1}(2|1)} \qquad j = 1, 2, ..., t_1. \tag{32}$$

It is easy to show that $\sum_{j=1}^{t_1} g(j|t_1) = 1$ as required. The probabilities for this distribution with $b = 0.3$, $c = 0.62$, and $t_1 = 10$, which gives a mean offspring number per lineage of 2, and $\rho = 0.01$ are shown in Fig. 2a, and with $\rho = 1$ in Fig. 2b.

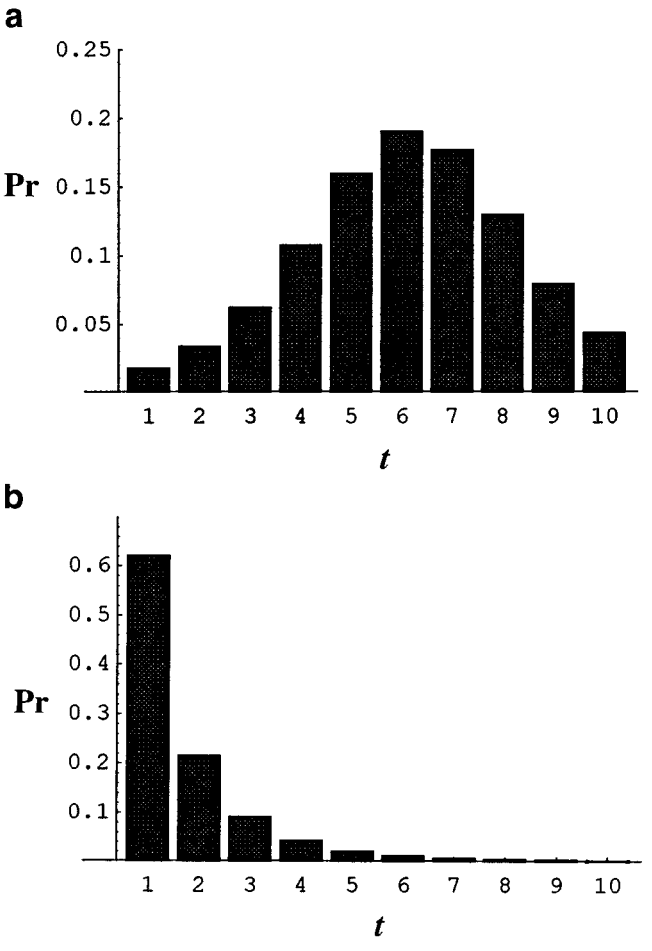

FIG. 2. The probability distribution of the coalescence times for a sample of $N = 10$ lineages of a branching process with a modified geometric offspring distribution with parameters $b = 0.3$, $c = 0.62$, and $t_1 = 10$ (see description in text). The mean offspring number per lineage in each generation is, in this case, equal to 2. The random variable t denotes the time at which a sampled lineage coalesces to a shared ancestral lineage and Pr denotes the probability that this coalescence event occurs at time t. (a) shows the probability distribution for a sample of 1% of the lineages in a population descended from a common ancestor. In this case, the distribution is nearly symmetrical with a mode of $t = 6$. (b) shows the probability distribution for a complete sample of all lineages in a population. In this case, the distribution is skewed toward more recent coalescence times with a mode of $t = 1$. These results illustrate the general effect of sampling a smaller fraction of the lineages descended from a most recent common ancestor which is to increase the average coalescence time relative to that for a complete sample of the same number of lineages.

3. RELATION TO THE BIRTH–DEATH PROCESS

The birth–death (BD) process (Feller, 1939; Kendall, 1949) is a continuous-time Markov process closely related to the discrete BP. The linear BD process can be described as follows: during an infinitesimal interval δt, any lineage gives birth to a single descendent lineage with probability $\lambda\delta t$, or dies with probability $\mu\delta t$. The probability of two or more events during δt is of order $o(\delta t)$ and may be neglected. The genealogy of the present-day (surviving) lineages under a BD process has been studied previously (see e.g., Thompson, 1975; Nee *et al.*, 1994; Rannala, 1997). If a population is initiated by a single lineage, arising at time t in the past, the probability that the population is extinct at present is (from Kendall, 1949)

$$p_0(t) = \frac{\mu(e^{(\lambda-\mu)t} - 1)}{\lambda e^{(\lambda-\mu)t} - \mu}. \tag{33}$$

The probability that the population consists of N descendent lineages at present is

$$p_N(t) = (1 - p_0(t))(1 - v_t)\, v_t^{N-1}, \tag{34}$$

where $v_t = p_0(t)(\lambda/\mu)$. If t is large relative to N, a BP with a modified geometric offspring distribution is equivalent to a linear BD process with parameters satisfying (Thompson, 1975)

$$m = e^{(\lambda-\mu)}, \qquad s_0 = \frac{\mu}{\lambda}. \tag{35}$$

The relationship between the distributions of the coalescence times under the two models may be demonstrated as follows: we consider observations from a BD process that are contained within adjacent disjoint intervals of the form $i = (T_i, T_{i-1})$, where $0 \leqslant T_i \leqslant t_1$ for all $i = 1, 2, ..., k$. Let t_1 be the time at which the ancestor of all present-day lineages arose, and let k be the number of intervals considered. The probability that $L(i)$ lineages in a sample coalesce to j ancestral lineages during the ith interval is

$$\Pr(L(i)\,|\,j) = \int_{x_1 = T_{i-1}}^{T_i} \cdots \int_{x_{L(i)} = T_{i-1}}^{x_{L(i)-1}} \lambda^{L(i)} p_1(x_1) \cdots p_1(x_{L(i)}) \times \frac{(L(i)+j)!}{j!}\, dx_1 \cdots dx_{L(i)}, \tag{36}$$

$$= \frac{(L(i)+j)!}{j!} (v_{T_i} - v_{T_{i-1}})^{L(i)}. \tag{37}$$

The joint probability of the configuration $\mathbf{L} = \{L(1), ..., L(k)\}$ and N is

$$\Pr(\mathbf{L}, N) = (N-1)! \left(\frac{\lambda}{\mu}\right)^{N-1} p_{t_1}(1) \times \prod_{j=1}^{k} \left\{ \frac{1}{L(j)!} (v_{T_j} - v_{T_{j-1}})^{L(j)} \right\}. \quad (38)$$

The probability of $\mathbf{L}$, conditional on N, is then

$$\Pr(\mathbf{L} \mid N) = \frac{\Pr(\mathbf{L}, N)}{\Pr(N)} = \binom{N-1}{L(1), ..., L(k)} \prod_{j=1}^{k} \left\{ \frac{v_{T_j} - v_{T_{j-1}}}{v_{t_1}} \right\}^{L(j)}. \quad (39)$$

With an appropriate substitution of parameters (i.e., setting $v_T = 1 - M_T$), this is identical to the distribution of $\mathbf{L}$ for the discrete BP with a complete population sample. Therefore, if the observations for a BD process are taken on disjoint intervals (each equivalent in length to a single generation of the discrete BP), rather than at distinct points, the joint distribution of coalescence events within each interval is identical to that for a BP with a modified geometric offspring distribution. An equivalent relationship may be established between the BD process with sampling and the fractional linear BP with sampling. The BD process may then be expected to provide a good approximation for the fractional linear BP when the intervals between observations are small and t_1 is large.

4. SIMULATING COALESCENCE TIMES

The joint distribution of the number of coalescence events at each generation of a BP with a modified geometric offspring distribution, as noted above, is equivalent to the joint distribution of the order statistics of $N-1$ random variables that are independent and identically distributed with density

$$g(x) = \frac{p_x(2 \mid 1) - p_{x-1}(2 \mid 1)}{p_{t_1}(2 \mid 1)}, \qquad 1 \leqslant x \leqslant t_1. \quad (40)$$

To generate observations from this density by simulation it is straightforward to apply the inverse transformation method. The basic algorithm is as follows (see Fishman, 1996): set $p = g(1)$, $q = p$, and $z = 1$. Generate a uniform $(0, 1)$ pseudorandom variable U. While $U \geqslant q$ set $z = z + 1$, $p = p \times h(z)$ and $q = q + p$, where $h(z) = g(z)/g(z-1)$. When $U \leqslant q$ the iteration is terminated and z is a pseudorandom observation from the distribution $g(1)$, $g(2)$, ..., $g(t_1)$. In many applications one may be interested in some function of the coalescence times, such as the total tree length. The moments and the distribution of such functions may often be determined by Monte Carlo integration using coalescence times simulated by the above procedure; this is useful when no analytical solutions exist.

5. DEMOGRAPHIC PARAMETERS

If one considers a mutant allele in low frequency in a population, the gene genealogy may be used to estimate parameters related to the history of the allele. For example, the age of the mutant allele might be of interest (i.e., the time at which the mutation first occurred; Slatkin and Rannala, 1997). The probability distribution of the coalescence times for a BP with a modified geometric distribution of offspring depends on the demographic parameters b and c, in addition to the parameter t_1. In principle, these parameters might be estimated using the genealogy, but often they may be more easily determined by a direct study of the population demography. Furthermore, if the demographic parameters can be estimated independently of the genetic data this will provide more power for estimating other parameters, such as t_1. A first step is then to estimate b and c using demographic data.

The demographic data usually consist of counts of the number of offspring per individual, per generation, that survive to reproduce at the next generation. Let N be the number of individuals observed and let n_i be the number of offspring for the ith individual producing offspring, where $1 \leqslant i \leqslant N$. Suppose that Z individuals have no offspring, where $0 \leqslant Z < N$. The joint offspring distribution for the remaining $N - Z$ individuals that have one, or more, offspring is

$$\Pr(\mathbf{n} \mid \neq 0) = p_k/(1 - p_0), = \prod_{j=1}^{N-Z} (1-c)\, c^{n_j - 1}, \quad (41)$$

where p_k and p_0 are given by Eq. (13) and $\mathbf{n} = \{n_1, n_2, ..., n_{N-Z}\}$. The notation $\neq 0$ indicates that $n_i > 0$ for all $i = 1, 2, ..., N - Z$. Taking logarithms we obtain

$$\ell = \sum_{j=1}^{N-Z} \{(n_j - 1)\log(c) + \log(1-c)\}. \quad (42)$$

This function does not depend on the parameter b and so by differentiation we have the maximum likelihood estimator (MLE) of $\hat{c}$,

$$\hat{c} = \frac{\bar{n} - 1}{\bar{n}}, \tag{43}$$

where $\bar{n}$ is the average number of offspring per individual. If we now consider the distribution of the number of individuals having no offspring we obtain

$$\Pr(Z) = \binom{N}{Z} \left\{\frac{1-c-b}{1-c}\right\}^{Z} \left\{\frac{b}{1-c}\right\}^{N-Z}, \tag{44}$$

so that an approximate MLE of $\hat{b}$ is

$$\hat{b} = (1-\hat{c})\left(\frac{N-Z}{N}\right). \tag{45}$$

Clearly, if $N > Z > 0$ then the required constraints on the parameters are satisfied, namely $\hat{c}, \hat{b} > 0$ and $\hat{c} + \hat{b} \leqslant 1$.

6. DISCUSSION

A genealogical perspective has proven useful for studying many problems in population genetics (see Kingman, 1982; Tavaré, 1984). The existing theory in population genetics concerning coalescent processes and gene genealogy focuses on a random sample of genes from a large population. The theory is valid under conditions similar to those necessary for a diffusion approximation of the process of gene frequency change. A case of some importance that is not easily treated using the generic coalescent theory is that of a rare allele descended from a unique (nonrecurrent) mutant ancestor. The dynamics of a rare allele, under these conditions, are well approximated using a BP model and it is therefore important to study the genealogy that arises for such a model. The distribution of the coalescence times for a BP presented in this paper allows genealogical approaches to be applied to study the demographic histories of rare mutant alleles in large populations. Similar approaches, based on a birth–death process approximation, have recently been applied to study the age of two putatively novel mutant alleles that cause cystic fibrosis, $\Delta F508$ and $G542X$, based on intraallelic genealogies reconstructed using closely-linked microsatellite loci (Slatkin and Rannala, 1997).

ACKNOWLEDGMENTS

Financial support for this work was provided by a postdoctoral fellowship from the Natural Sciences and Engineering Research Council (NSERC) of Canada, and in part by a grant from the National Institutes of Health (GM-40282) to M. Slatkin. I thank M. Slatkin and two anonymous reviewers for helpful suggestions.

REFERENCES

Ewens, W. J. 1979. "Mathematical Population Genetics," Springer-Verlag, New York.

Feller, W. 1939. Die Grundlagen der Volterraschen Theorie des Kampfes um das Sein in wahrscheinlichkeits theoretischen Behandlung, *Acta Biotheoretica* **5**, 1–40.

Fisher, R. A. 1922. On the dominance ratio, *Proc. R. Soc. Edinburgh* **42**, 321–341.

Fisher, R. A. 1930. "The Genetical Theory of Natural Selection," Clarendon, Oxford.

Fishman, G. S. 1996. "Monte Carlo: Concepts, Algorithms, and Applications," Springer-Verlag, New York.

Fleischmann, K., and Siegmund–Schultze, R. 1977. The structure of reduced critical Galton-Watson processes, *Math. Nachr.* **79**, 233–241.

Haldane, J. B. S. 1927. A mathematical theory of natural and artificial selection. V. Selection and mutation, *Proc. Cambridge Philos. Soc.* **23**, 838–844.

Harris, T. E. 1948. Branching processes, *Ann. Math. Statist.* **19**, 474–494.

Harris, T. H. 1963. "The Theory of Branching Processes," Dover, New York.

Kendall, D. G. 1949. Stochastic processes and population growth, *J. R. Statist. Soc. B* **11**, 230–264.

Kingman, J. C. 1982. On the genealogy of large populations, *J. Appl. Probab. A* **19**, 27–43.

Lotka, A. J. 1931. The extinction of families, I, *J. Washington Acad. Sci.* **21**, 377–380.

Nee, S., May, R. M., and Harvey, P. H. 1994. The reconstructed evolutionary process, *Philos. Trans. R. Soc. London B* **344**, 305–311.

O'Connell, N. 1993. Yule process approximation for the skeleton of a branching process, *J. Appl. Prob.* **30**, 725–729.

O'Connell, N. 1995. The genealogy of branching processes and the age of our most recent ancestor, *Adv. Appl. Prob.* **27**, 418–442.

Rannala, B. 1997. Gene genealogy in a population of variable size, *Heredity* **78**, 417–423.

Slatkin, M. 1996. Gene genealogies within mutant allelic classes, *Genetics* **143**, 579–587.

Slatkin, M., and Rannala, B. 1997. Estimating the age of alleles using intraallelic variability, *Am. J. Human Genet.* **60**, 447–458.

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

Thompson, E. A. 1975. "Human Evolutionary Trees," Cambridge Univ. Press, Cambridge.

Thompson, E. A. 1976. Estimation of age and rate of increase of rare variants, *Am. J. Human Genet.* **28**, 442–452.

Yule, G. U. 1925. A mathematical theory of evolution, based on the conclusions of Dr. J. C. Willis, F.R.S., *Philos. Trans. R. Soc. London B* **213**, 21–87.