Theoretical Population Biology 93 (2014) 14–29

Contents lists available at ScienceDirect

Theoretical Population Biology

journal homepage: www.elsevier.com/locate/tpb

Theory and applications of a deterministic approximation to the coalescent model

Ethan M. Jewett*, Noah A. Rosenberg

Department of Biology, Stanford University, 371 Serra Mall, Stanford, CA 94305-5020, USA

ARTICLE INFO

Article history:
Received 14 November 2013
Available online 7 January 2014

Keywords:
Approximation
Coalescent
Computational complexity

ABSTRACT

Under the coalescent model, the random number n_t of lineages ancestral to a sample is nearly deterministic as a function of time when n_t is moderate to large in value, and it is well approximated by its expectation $E[n_t]$. In turn, this expectation is well approximated by simple deterministic functions that are easy to compute. Such deterministic functions have been applied to estimate allele age, effective population size, and genetic diversity, and they have been used to study properties of models of infectious disease dynamics. Although a number of simple approximations of $E[n_t]$ have been derived and applied to problems of population-genetic inference, the theoretical accuracy of the resulting approximate formulas and the inferences obtained using these approximations is not known, and the range of problems to which they can be applied is not well understood. Here, we demonstrate general procedures by which the approximation $n_t \approx E[n_t]$ can be used to reduce the computational complexity of coalescent formulas, and we show that the resulting approximations converge to their true values under simple assumptions. Such approximations provide alternatives to exact formulas that are computationally intractable or numerically unstable when the number of sampled lineages is moderate or large. We also extend an existing class of approximations of $E[n_t]$ to the case of multiple populations of time-varying size with migration among them. Our results facilitate the use of the deterministic approximation $n_t \approx E[n_t]$ for deriving functionally simple, computationally efficient, and numerically stable approximations of coalescent formulas under complicated demographic scenarios.

1. Introduction

Many coalescent distributions and expectations can be obtained by conditioning on the random number n_t of lineages at time t in the past that are ancestral to a sample of n_0 lineages at time $t = 0$ in the present (Fig. 1). Quantities that can be obtained by conditioning on n_t include Wakeley and Hey's (1997) formula for the joint allele frequency spectrum between two populations, Takahata's (1989) formula for the probability of concordance between a gene tree and a species tree, Griffiths and Tavaré's (1998) formula for the distribution of the age of a neutral allele, Rosenberg's (2003) formulas for the probabilities of monophyly, paraphyly, and polyphyly in two populations, and many others (Takahata and Nei, 1985; Hudson and Coyne, 2002; Rosenberg, 2002; Rosenberg and Feldman, 2002; Degnan and Salter, 2005; Efromovich and Kubatko, 2008; Degnan, 2010; Bryant et al., 2012; Helmkamp et al., 2012; Jewett and Rosenberg, 2012; Wu, 2012).

When many lineages are sampled (and n_0 is large), summing over all possible values of n_t can be computationally expensive. As a result, evaluating formulas that condition on n_t can be computationally difficult or intractable for modern genomic datasets with hundreds or thousands of sampled alleles. In addition, formulas for the probability distribution $\mathbb{P}(n_t)$ of the number of ancestors at time t (Griffiths, 1980; Donnelly, 1984; Tavaré, 1984) involve sums of terms of alternating sign that produce round-off error when t is small and n_0 is large (e.g. $t \lesssim 10^{-2}$ coalescent time units and $n_0 \gtrsim 50$), further complicating the evaluation of formulas that condition on n_t (Griffiths, 1984).

When computing formulas that depend on the distribution $\mathbb{P}(n_t)$, round-off error can be eliminated by using asymptotic approximations of $\mathbb{P}(n_t)$ that were derived by Griffiths (1984), or by using an alternative expression for $\mathbb{P}(n_t)$ (Griffiths, 2006). However, as we will discuss, approximations to coalescent formulas obtained by this approach may have similar computational complexities to the exact formulas, and can therefore be computationally slow or intractable on large datasets. Therefore, it is of interest to devise general procedures for deriving approximate coalescent formulas without requiring conditional sums over all possible values of n_t.

One alternative to summing over n_t is to use an approximation in which n_t is assumed to be equal to its expected value $E[n_t]$ with probability one. This approximation was used by Slatkin (2000) to

* Corresponding author.
E-mail addresses: emjewett@stanford.edu (E.M. Jewett), noahr@stanford.edu (N.A. Rosenberg).

0040-5809/$ – see front matter © 2014 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.tpb.2013.12.007

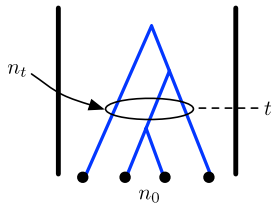

Fig. 1. The number n_t of coalescent lineages at time t in the past that are ancestral to a set of n_0 lineages sampled at time $t = 0$ in the present. In this example, $n_0 = 4$ and $n_t = 3$ at the given time t.

address the problem of round-off error in the distribution $\mathbb{P}(n_t)$ and by Volz et al. (2009) to obtain approximate distributions of coalescent waiting times. The approximation can greatly reduce the complexity of computing coalescent formulas by reducing the number of different values of n_t over which conditional summations must be computed (Jewett and Rosenberg, 2012).

The surprising fact is that approximations of this kind are often very accurate because n_t changes almost deterministically over time and is well approximated by its expected value (Watterson, 1975; Slatkin, 2000; Maruvka et al., 2011). In fact, Maruvka et al. (2011) demonstrated that the deterministic nature of n_t is apparent even when the number n_t of ancestral lineages is not large. From Fig. 2, it can be seen that the variance in n_t increases as the number of ancestral lineages decreases, with n_t deviating most from $E[n_t]$ when $n_t \lesssim 30$ in the example shown. However, n_t is well approximated by its mean when t is small. $E[n_t]$ is also a good approximation of n_t as $t \to \infty$ and both n_t and $E[n_t]$ approach unity. The approximation $n_t \approx E[n_t]$ can be used to obtain approximations of coalescent distributions that are computationally fast, numerically stable, and accurate for a broad range of sample sizes n_0.

In addition to deriving fast and numerically stable approximations to coalescent formulas, the approximation $n_t \approx E[n_t]$ can be combined with simple approximate formulas for $E[n_t]$ (Slatkin and Rannala, 1997; Slatkin, 2000; Rauch and Bar-Yam, 2005; Volz et al., 2009; Frost and Volz, 2010; Maruvka et al., 2011) to derive functionally simple approximate expressions for coalescent quantities (Slatkin, 2000; Volz et al., 2009; Jewett and Rosenberg, 2012).

Despite the utility of the approximation $n_t \approx E[n_t]$, it is not widely known and general procedures for applying it to obtain approximate coalescent formulas have not been developed. Moreover, the theoretical accuracy of the approximate formulas is not well understood. Here, we discuss general approaches by which the approximation $n_t \approx E[n_t]$ can be applied to obtain functionally simple, computationally efficient, and numerically stable approximations of coalescent distributions. We show that the resulting approximate formulas converge to their true values under simple assumptions, and we derive approximate expressions for the error. We also discuss methods for approximating $E[n_t]$ under demographic models that include multiple populations of time-varying size with migration among them. Our results facilitate the use of the approximation $n_t \approx E[n_t]$ for obtaining computationally fast and numerically stable formulas that can be applied to enhance coalescent computations on large genomic datasets with complicated demographic histories.

2. Approximating formulas that condition on n_t

2.1. Difficulties of computing coalescent formulas

We first consider applications of the approximation $n_t \approx E[n_t]$ to the problem of reducing the computational complexity and numerical instability of coalescent formulas that are derived by

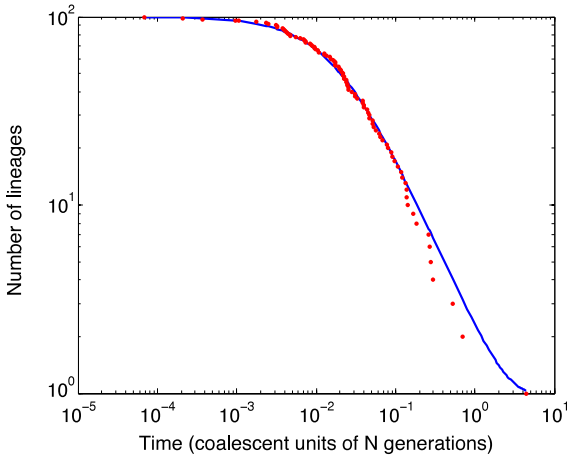

Fig. 2. The deterministic nature of the number of ancestral lineages n_t at time t in the past. Red dots indicate the number of lineages remaining at each coalescent event in a single genealogy of $n_0 = 100$ lineages sampled from a population of constant size under the coalescent model. The expectation $E[n_t]$ computed using Eq. (13) is shown in blue. It can be seen that n_t is well-approximated by its expected value.

conditioning on n_t at a particular time t in the past. In particular, we consider functions of the form

$$f(x) = \sum_{\mathbf{n}_t} f(x|\mathbf{n}_t)\mathbb{P}(\mathbf{n}_t), \tag{1}$$

where $\mathbf{n}_t = (n_{1,t}, \ldots, n_{k,t})$ is a vector describing the number of ancestors of each of k different sets of sampled alleles with initial sample sizes $\{n_{i,0}\}_{i=1}^k$. The sets of lineages of sizes $\{n_{i,0}\}_{i=1}^k$ can be drawn from different populations, but they can also come from the same population. Here, $f(x)$ is a quantity of interest that we wish to compute, such as an expectation parameterized by a variable x or a probability distribution function for a random variable X. The sum is carried out over k variables, one for each entry in $\mathbf{n}_t$, and the ith sum proceeds from 1 to $n_{i,0}$.

Two primary difficulties arise when evaluating functions of the form in Eq. (1). First, summing over all values of $\mathbf{n}_t$ can be computationally expensive, making conditional formulas computationally intractable when many lineages are sampled. Second, for any given number of sampled alleles, i, the distribution $\mathbb{P}(n_{i,t})$ of the number of ancestors is given by a complicated expression

$$\mathbb{P}(n_{i,t}) = \sum_{j=n_{i,t}}^{n_{i,0}} \frac{(-1)^{j-n_{i,t}}(2j-1)(n_{i,t})_{(j-1)}(n_{i,0})_{[j]}}{n_{i,t}!(i-n_{i,t})!(n_{i,0})_{(j)}} e^{-\binom{j}{2}t}, \tag{2}$$

where $n_{[j]} = n!/(n-j)!$ and $n_{(j)} = (n+j-1)!/(n-1)!$ and where time, t, is in coalescent units of N generations (Tavaré, 1984). Due to terms of alternating sign in Eq. (2), this distribution is subject to round-off error when $n_0 \gtrsim 50$ and $t \lesssim 10^{-2}$, making calculations inaccurate. Therefore, because of difficulties with computational complexity and numerical instability, it is of interest to find other means of evaluating formulas of the form given in Eq. (1).

2.1.1. The Griffiths approximation

One approach for eliminating round-off error in coalescent formulas of the form given in Eq. (1) is to use a set of asymptotic approximations derived by Griffiths (1984). Griffiths showed that as $n_0 \to \infty$ and $t \to 0$, n_t has an asymptotically normal distribution. He derived expressions for the asymptotic mean μ_t and variance σ_t^2 of this distribution. Griffiths' asymptotic formulas can be used to obtain numerically stable approximations to formulas of the form given in Eq. (1) by replacing the distribution $\mathbb{P}(n_{i,t})$ $(i = 1, \ldots, k)$ with the corresponding asymptotic normal distribution (Chen and Chen, 2013). Using Griffiths' asymptotic formulas,

the approximation of Eq. (1) is

$$f(x) = \sum_{\mathbf{n}_t} f(x|\mathbf{n}_t) \prod_{i=1}^{k} \frac{1}{\sqrt{2\pi}\sigma_{i,t}} e^{-(n_{i,t}-\mu_{i,t})^2/(2\sigma_{i,t}^2)}, \tag{3}$$

where $\mu_{i,t}$ and $\sigma_{i,t}$ are the mean and variance of Griffiths' normal approximation to the distribution $\mathbb{P}(n_{i,t})$, and where the summation is taken over $n_{i,t} = 1, \ldots, n_{i,0}$ for $i = 1, \ldots, k$. Throughout this manuscript, we refer to an approximation of the form in Eq. (3) to an exact coalescent formula of the form given in Eq. (1) as *Griffiths' approximation* of the formula.

The asymptotic approximations derived by Griffiths are useful for eliminating round-off error when evaluating the distribution of n_t. However, although Griffiths' normal approximations are very fast to compute, the complexity of Eq. (3) is similar to that of Eq. (1) because the same number of terms of approximately the same complexity must be computed in both formulas. Thus, it is of interest to identify alternatives to Griffiths' asymptotic formulas that can be used to evaluate coalescent expressions in a computationally efficient way when the sample size is large. The key challenge is to eliminate the multiple summation over $\prod_{i=1}^{k} n_{i,0}$ terms.

2.1.2. The deterministic approximation

We consider an alternative to Griffiths' asymptotic formulas that is useful for reducing the computational complexity of equations of the form given in Eq. (1) when the number n_0 of sampled lineages is large. The alternative is to assume that the number n_t of lineages ancestral to a given sample of n_0 alleles is equal to its expected value $E[n_t]$ with probability 1. The result of this approximation is that the summation in Eq. (1) collapses to a single term

$$f(x) = \sum_{\mathbf{n}_t} f(x|\mathbf{n}_t)\mathbb{P}(\mathbf{n}_t) \approx f(x|E[\mathbf{n}_t]), \tag{4}$$

which is fast to evaluate. Throughout this manuscript we refer to an approximation of the form in Eq. (4) to an exact coalescent formula of the form given in Eq. (1) as the *deterministic approximation* of the formula.

To our knowledge, the deterministic approximation was first used by Slatkin (2000) to treat problems with round-off error in the distribution $\mathbb{P}(n_{i,t})$. We demonstrate here that this approximation can often be used as an alternative to Griffiths' approximation, to reduce the computational complexity of coalescent formulas that contain terms of the form in Eq. (1).

2.2. Approximating distributions that condition on the path of n_t

A more general version of the approximation in Eq. (4) applies to formulas that can be obtained by conditioning on the path of the stochastic process $\mathbf{n}_t$ over a range of time values $[r, s]$, rather than on the instantaneous value of the process $\mathbf{n}_t$ at the single time point t. In particular, consider the stochastic process $\mathbf{n}_t$ $(0 \le t \le \infty)$, where the value at $t = \infty$ refers to the $t \to \infty$ limit, and let $\mathbf{n}_{[r,s]}$ denote a sample path of the process on the time interval $[r, s]$. We consider approximations to coalescent quantities $f(x)$ that can be expressed using formulas of the form

$$f(x) = \int_{\mathcal{A}_{[r,s]}} f(x|\mathbf{n}_{[r,s]})p(\mathbf{n}_{[r,s]})d\mathcal{A}_{[r,s]}, \tag{5}$$

where $f(x|\mathbf{n}_{[r,s]})$ is the conditional expression for $f(x)$ given a particular sample path $\mathbf{n}_{[r,s]}$ on the interval $[r, s]$, $\mathcal{A}_{[r,s]}$ is the sample space of all paths of the stochastic process $\mathbf{n}_t$ on the time interval $[r, s]$, and $p(\mathbf{n}_{[r,s]})$ is the probability density function of these paths.

Such conditional formulas represent a wide variety of coalescent quantities. For example, consider a single set of sampled alleles ($k = 1$ and $\mathbf{n}_t = n_t$) on the time interval $[r, s] = [0, \infty)$. If we define the conditional function

$$f(x|n_{[0,\infty)}) = \begin{cases} 1 & \text{if } n_x = 1 \\ 0 & \text{otherwise,} \end{cases} \tag{6}$$

then Eq. (6) is an indicator random variable that takes on the value 1 if the n_0 sampled alleles find their most recent common ancestor before time x. In this case, Eq. (5) is the cumulative distribution function of the time to the most recent common ancestor (TMRCA).

Alternatively, we could consider the time interval $[r, s]$ and define the conditional function $f(x|n_{[r,s]})$ to be

$$f(x|n_{[r,s]}) = \int_{z=r}^{s} n_z \, dz.$$

This quantity is the total sum of branch lengths of the sample path on the time interval $[r, s]$. In this case, $f(x)$ in Eq. (5) is the expected branch length of the genealogy on the time interval $[r, s]$.

2.2.1. Approximating Eq. (5)

By analogy with Eq. (4), quantities of the form given in Eq. (5) can be approximated as

$$f(x) = \int_{\mathcal{A}_{[r,s]}} f(x|\mathbf{n}_{[r,s]})p(\mathbf{n}_{[r,s]})d\mathcal{A}_{[r,s]} \approx f(x|E[\mathbf{n}_{[r,s]}]), \tag{7}$$

where $E[\mathbf{n}_{[r,s]}]$ is the expected sample path of the stochastic process $\mathbf{n}_t$ over the time interval $[r, s]$. Such approximations not only reduce the complexity of computing coalescent quantities by eliminating the integral over all possible paths, they also facilitate the derivation of approximate coalescent formulas that would otherwise be difficult to derive analytically.

2.2.2. An application of Eq. (7)

For a single sample of n_0 alleles, specifying the term $f(x|n_{[r,s]})$ in Eq. (7) by $f(x|n_{[r,s]}) = \int_{z=r}^{s} n_z dz$ is particularly useful for computing quantities that depend on the expected number of segregating sites in all or in part of a genealogy. In particular, under the infinitely-many-sites model, the expected number of mutations S on a genealogy at a locus of length b bases is proportional to the expected total branch length L of the genealogy:

$$E[S] = E[E[S|L]] = E[\theta bL/4] = \frac{\theta b}{4}E[L], \tag{8}$$

where $\theta = 4N\mu$ is the population-scaled mutation rate per-site per-generation, N is a specified haploid effective population size, μ is the per-site per-generation mutation rate, and L is given in units of N generations. If $L_{[r,s]}$ is the total length of a genealogy over the time interval $[r, s]$, then the expected number of segregating sites $S_{[r,s]}$ in the interval is

$$E[S_{[r,s]}] = \frac{\theta b}{4}E[L_{[r,s]}]. \tag{9}$$

The expectation on the right-hand side of Eq. (9) can be computed using the following theorem:

Theorem 2.1. *Let $L_{[r,s]}$ be the total sum of branch lengths of the genealogy of n_0 sampled alleles in the time interval $[r, s]$ with $0 \le r \le s \le \infty$. Then the expectation $E[L_{[r,s]}]$ is given by*

$$E[L_{[r,s]}] = \int_{z=r}^{s} E[n_z]dz. \tag{10}$$

The proof of Theorem 2.1 is given in Appendix A. As we demonstrate in Section 5, Eq. (10) can be used to compute quantities such as the number of mutations that are private to a given population or sample and terms in the joint allele frequency spectrum among a pair of populations. A result similar to Theorem 2.1 that considers the full genealogy up until the time to the most recent common ancestor was proved by Chen and Chen (2013).

3. The theoretical accuracy of the approximate formula

In this section we consider the accuracy of the approximate coalescent formula obtained using Eq. (4). In comparison with Griffiths' approximation (Eq. (3)), which was shown to converge to the correct value in the double limit as n_0 increases to infinity and t decreases to zero (Griffiths, 1984), we show that the deterministic approximation (Eq. (4)) of a coalescent formula converges to the true value as $t \to 0$ and as $t \to \infty$ with the value of n_0 fixed. As we will see, these less stringent criteria for convergence often allow the deterministic approximation to be more accurate than Griffiths' approximation when the sample size n_0 is small. The accuracy of the deterministic approximation is formalized in the following theorem.

Theorem 3.1. *Suppose that a function $f(x)$ can be expressed as $f(x) = \sum_{\mathbf{n}_t} f(x|\mathbf{n}_t)\mathbb{P}(\mathbf{n}_t)$, where $f(x|\mathbf{n}_t)$ is defined for all x in some domain $D \subseteq \mathbb{R}$ and for $n_{i,t} \in \mathcal{N}_i = [1, n_{i,0}]$ $(i = 1, \ldots, k)$. Suppose that the second-order partial derivatives $\frac{\partial^2}{\partial n_{i,t}\partial n_{j,t}} f(x|\mathbf{n}_t)$ exist and are continuous and bounded in $n_{i,t}$ $(i = 1, \ldots, k)$ for all $x \in D$ and for $\mathbf{n}_t \in \mathcal{N} = \mathcal{N}_1 \times \cdots \times \mathcal{N}_k$. Then for a fixed value of $\mathbf{n}_0$, $f(x|E[\mathbf{n}_t])$ converges uniformly to $f(x)$ on D as $t \to 0$ and as $t \to \infty$.*

The proof of Theorem 3.1 follows from a lemma proved in Appendix B and is given in Appendix C. We also obtain an approximate expression for the error in the deterministic approximation as $t \to 0$ and as $t \to \infty$. In particular, we show that the error $|f(x) - f(x|E[\mathbf{n}_t])|$ is given approximately by

$$|f(x) - f(x|E[\mathbf{n}_t])| \approx \frac{1}{2}\left|\sum_{i=1}^{k}\sum_{j=1}^{k}\mathrm{Cov}(n_{i,t}, n_{j,t})\frac{\partial^2}{\partial n_{i,t}\partial n_{j,t}} f(x|E[\mathbf{n}_t])\right| \tag{11}$$

as $t \to 0$ and as $t \to \infty$ (Appendix D). In the commonly-occurring scenario in which the numbers of ancestors $n_{i,t}$ $(i = 1, \ldots, k)$ are independent of one another, Eq. (11) reduces to

$$|f(x) - f(x|E[\mathbf{n}_t])| \approx \frac{1}{2}\left|\sum_{i=1}^{k}\mathrm{Var}(n_{i,t})\frac{\partial^2}{\partial n_{i,t}^2} f(x|E[\mathbf{n}_t])\right|. \tag{12}$$

Eq. (12) can be evaluated for any given quantity $f(x)$ either by evaluating Tavaré's expression for $\mathrm{Var}(n_t)$ (Eq. (B.10)), or by using one of the asymptotic expressions for $\mathrm{Var}(n_t)$ given in Theorem 2 of Griffiths (1984).

4. Approximating $E[n_t]$

In order to apply the approximation $n_t \approx E[n_t]$, it is necessary to compute $E[n_t]$. Chen and Chen (2013) noted that the expected value $E[n_t]$ can be computed for a population of variable size $N(t)$ at time t in the past using the formula derived by Tavaré (1984)

$$E[n_t|n_0] = \sum_{i=1}^{n_0}(2i-1)\frac{(n_0)_{[i]}}{(n_0)_{(i)}}e^{-\binom{i}{2}\tau(t)}, \tag{13}$$

where $n_{[i]} = n!/(n-i)!$ and $n_{(i)} = (n-1+i)!/(n-1)!$, where time t is in units of generations, and where $\tau(t) = \int_{z=0}^{t} 1/N(z)dz$ is a rescaling of time (see Section 4.1). In a population of constant size $N(t) = N$, $\tau(t)$ simplifies to $\tau(t) = t/N$. Although Eq. (13) has a functionally simple form (a polynomial in $e^{-\tau(t)}$), it can be slow to compute when the sample size n_0 is large, and it does not hold for complicated demographic models with migration. Because there is currently no closed-form expression for $E[n_t]$ in the case of migration, it is of interest to obtain accurate approximations of $E[n_t]$ in this more complicated scenario. Note that the problem of approximating $E[n_t]$ is distinct from the problem of approximating n_t by $E[n_t]$.

Several studies derived simple deterministic approximations of $E[n_t]$ in a single panmictic population (Griffiths, 1984; Slatkin and Rannala, 1997; Rauch and Bar-Yam, 2005; Volz et al., 2009; Frost and Volz, 2010; Maruvka et al., 2011). With the exception of the approximations derived by Griffiths (1984), these studies all used a differential equation approach to obtain approximations of $E[n_t]$, all employing slight variations on the same differential equation. Here, we show that this differential equation can be extended to obtain an approximation of $E[n_t]$ under models with migration among populations.

As background for our derivation, we begin with a brief overview of approximations of $E[n_t]$ in a single population. We also take the opportunity to compare these approximations of $E[n_t]$ to one another in terms of their relative accuracy, and we theoretically validate these approximations by showing that they are in fact asymptotically equal to $E[n_t]$ in certain limits.

4.1. Approximating $E[n_t]$ in a single population

Slatkin and Rannala (1997) derived a differential equation for $E[n_t]$ in a single population:

$$\frac{dE[n_t]}{dt} = -\binom{E[n_t]}{2}\frac{1}{N(t)} - \frac{\mathrm{Var}(n_t)}{2N(t)}, \tag{14}$$

where $N(t)$ is the size of the population at time t in the past. The approximate formulas for $E[n_t]$ derived by Slatkin and Rannala (1997), Volz et al. (2009), Frost and Volz (2010), and Maruvka et al. (2011) can each be derived by making various simplifying approximations of Eq. (14). In each approximation, $\mathrm{Var}(n_t)$ is assumed to be much smaller than $\binom{E[n_t]}{2}$, so that the term $\mathrm{Var}(n_t)/(2N(t))$ can be neglected. Slatkin and Rannala (1997) and Volz et al. (2009) further assumed that $E[n_t] \gg 1$ in order to obtain the approximation

$$\frac{dE[n_t]}{dt} \approx -\frac{E[n_t]^2}{2N(t)}. \tag{15}$$

Frost and Volz (2010) and Maruvka et al. (2011) retained the term $-E[n_t]/(2N(t))$, obtaining the approximation

$$\frac{dE[n_t]}{dt} \approx -\binom{E[n_t]}{2}\frac{1}{N(t)}. \tag{16}$$

Eqs. (15) and (16) can both be simplified further by using a trick implemented by Slatkin and Rannala (1997). In particular, Griffiths and Tavaré (1994) showed that the distribution of the number of ancestral lineages at time t generations in a population of time-varying size $N(t)$ is the same as the distribution of the number of ancestral lineages in a constant population of size $N = 1$ at time $\tau(t) = \int_{z=0}^{t} 1/N(z)dz$. Thus, Slatkin and Rannala (1997) noted, it is sufficient to solve Eqs. (15) and (16) for the case of $N = 1$ and then evaluate the solution at time $\tau(t)$. This approach yields the solution

$$E[n_t] \approx \frac{n_0}{1 + n_0\tau(t)/2} \tag{17}$$

for Eq. (15) and the solution

$$E[n_t] \approx \frac{n_0}{n_0 + (1-n_0)e^{-\tau(t)/2}} \tag{18}$$

for Eq. (16). These approximations of $E[n_t]$ are summarized in Table 1.

Eqs. (17) and (18) are well-motivated by the approximations used to obtain Eqs. (15) and (16) from Eq. (14). However, these approximations do not guarantee that Eqs. (17) and (18) will be accurate, nor do they shed light on the ranges of parameter values over which we can expect the approximate expressions for $E[n_t]$ to hold. By comparing Eqs. (17) and (18) to asymptotic formulas for $E[n_t]$ derived by Griffiths (1984), for which theoretical results on accuracy exist, a characterization of their accuracy can be obtained.

Table 1
Approximations of $E[n_t]$, with $\tau(t) = \int_{z=0}^{t} 1/N(z)dz$.

Authors	Assumptions	Equation	Solution
Slatkin and Rannala (1997), Volz et al. (2009)	$\mathrm{Var}(n_t) \ll E[n_t]$, $E[n_t] \gg 1$	$\frac{d}{dt}E[n_t] \approx -\frac{E[n_t]^2}{2}$	$E[n_t] \approx \frac{n_0}{1+n_0\tau(t)/2}$
Frost and Volz (2010), Maruvka et al. (2011)	$\mathrm{Var}(n_t) \ll E[n_t]$	$\frac{d}{dt}E[n_t] \approx -\binom{E[n_t]}{2}$	$E[n_t] \approx \frac{n_0}{n_0+(1-n_0)e^{-\tau(t)/2}}$
This paper	$\mathrm{Var}(n_t) \ll E[n_t]$	$\frac{d}{dt}E[n_{\ell t}] \approx -\binom{E[n_{\ell t}]}{2} + \sum_{\substack{i=1 \\ i\neq\ell}}^{k} (E[n_{it}]m_{i\ell} - E[n_{\ell t}]m_{\ell i})$	Numerical solution
Griffiths (1984)[a]	$n_0 \to \infty$, $t \to 0$, $n_0 t < \infty$	No equation. Derived using a limit theorem approach.	$E[n_t] = \frac{n_0}{n_0+(1-n_0)e^{-\tau(t)/2}}$

[a] The equation for $E[n_t]$ presented in Griffiths (1984) is given in terms of variables that are functions of n_0 and t, and is expressed for the case of a population of constant size. For purposes of comparison, we have expressed the formula from Griffiths in terms of n_0 and t, and we have modified it to include the transformation $\tau(t)$ to account for the variability in population size.

4.1.1. Accuracy of approximations of $E[n_t]$ in the double limit as $t \to 0$ and $n_0 \to \infty$

Griffiths (1984) proved that as $n_0 \to \infty$ and as $t \to 0$, $E[n_t]$ is asymptotically given by the simple expression

$$E[n_t] \approx \frac{n_0}{n_0 + (1-n_0)e^{-\tau(t)/2}}, \tag{19}$$

which is exactly equal to the expression of Frost and Volz (2010) and Maruvka et al. (2011) (Eq. (18)). Thus, Eq. (18) is asymptotically equal to $E[n_t]$ in the double limit as $n_0 \to \infty$ and $t \to 0$. Furthermore, because $\tau(t) \to 0$ as $t \to 0$, it follows that $e^{-\tau(t)/2} \approx 1-\tau(t)/2$ as $t \to 0$. Thus, in the double limits $n_0 \to \infty$ and $t \to 0$, we have

$$\begin{aligned} \frac{n_0}{n_0 + (1-n_0)e^{-\tau(t)/2}} &\approx \frac{n_0}{n_0 + (1-n_0)(1-\tau(t)/2)} \\ &\approx \frac{n_0}{1+n_0\tau(t)/2}. \end{aligned} \tag{20}$$

Eq. (20) implies that the approximation of Slatkin and Rannala (1997) and Volz et al. (2009) (Eq. (17)) is asymptotic to $E[n_t]$ in the double limit $n_0 \to \infty$ and $t \to 0$.

4.1.2. Accuracy of approximations of $E[n_t]$ in the single limit as $t \to 0$ for fixed n_0

Comparing Eqs. (17) and (18) with Tavaré's (1984) formula for $E[n_t]$ (Eq. (13)) allows us to establish that Eqs. (17) and (18) are asymptotically equal to $E[n_t]$ as $t \to 0$ for fixed values of n_0. In particular, from Eq. (B.8), we have

$$E[n_t] = n_0 - \tau(t)\binom{n_0}{2} + \mathcal{O}(\tau(t)^2). \tag{21}$$

In comparison to Eq. (21), expanding Eq. (17) around $\tau(t) = 0$ gives $n_0 - \tau(t)n^2/2 + \mathcal{O}(\tau(t)^2)$, and expanding Eq. (18) around $\tau(t) = 0$ gives $n_0 - \tau(t)n_0(n_0-1)/2 + \mathcal{O}(\tau(t)^2)$. Thus, Eqs. (17) and (18) are both asymptotic to $E[n_t]$ as $t \to 0$, with Eq. (18) holding more accurately when n_0 is small.

4.1.3. Accuracy of approximations of $E[n_t]$ in the single limit as $t \to \infty$ for fixed n_0

Although both Eqs. (17) and (18) are asymptotically equal to $E[n_t]$ as $t \to 0$, only Eq. (18) is asymptotic to $E[n_t]$ as $t \to \infty$. This result follows from the fact that as $t \to \infty$, Eq. (18) approaches unity, which is the limiting value of $E[n_t]$ as $t \to \infty$, whereas Eq. (17) approaches zero.

The asymptotic behavior of approximations (17) and (18) is shown in Fig. 3 for the case of $n_0 = 10$ sampled alleles in a population of constant size. It can be seen that both formulas (17) and (18) converge to the true mean $E[n_t]$ as $t \to 0$ with n_0 fixed, with Eq. (18) converging more quickly. Although the sample size n_0 is small, Eqs. (17) and (18) are still very good approximations of $E[n_t]$ as $t \to 0$. Furthermore, although Eq. (17) is inaccurate for large times t, it has comparable accuracy to Eq. (18) at small t and has a functionally simpler form. Thus, the simpler Eq. (17) can be useful for deriving simple approximate formulas when accuracy is needed only at small t.

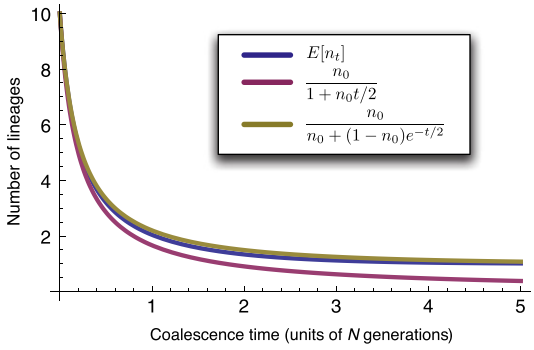

Fig. 3. Comparison of simple approximations of $E[n_t]$ in one population with $n_0 = 10$ sampled alleles. The exact mean $E[n_t]$ (Eq. (13), blue) is compared to the approximation of Slatkin and Rannala (1997) and Volz et al. (2009) (Eq. (17), purple) and to the approximation of Frost and Volz (2010) and Maruvka et al. (2011) (Eq. (18), green).

4.2. Approximating $E[n_t]$ under migration

In this section, we extend the derivation of Slatkin and Rannala (1997) to the case of k populations, each of variable size $N_i(z)$ ($i = 1, \ldots, k$) at time $z \geq 0$ in the past, with migration among them. In the model we consider, lineages in population i migrate to population j at rate m_{ij} as time moves backward, where the m_{ij} represent backwards migration rates.

Let $\mathbf{n}_t = (n_{1,t}, n_{2,t}, \ldots, n_{k,t})$ record the number of ancestral lineages in all populations at time t in the past. If the lineages follow a coalescent process in each population, then $\mathbf{n}_t$ satisfies a time-inhomogeneous Markov process with instantaneous transition probabilities given by

$$\begin{aligned} &\mathbb{P}(\mathbf{n}_{t+\delta} = \varphi' | \mathbf{n}_t = \varphi) \\ &= \begin{cases} 1 - \sum_{i=1}^{k} \binom{\varphi_i}{2} \frac{1}{N_i(t)}\delta & \\ \quad - \sum_{i=1}^{k} \sum_{\substack{j=1 \\ j\neq i}}^{k} \varphi_i m_{ij}\delta + o(\delta) & \text{if } \varphi = \varphi' \\ \varphi_i m_{ij}\delta + o(\delta) & \text{if } \varphi = \varphi' + \mathbf{e}_i - \mathbf{e}_j \\ \binom{\varphi_i}{2} \frac{1}{N_i(t)}\delta + o(\delta) & \text{if } \varphi = \varphi' + \mathbf{e}_i \\ 0 & \text{otherwise,} \end{cases} \end{aligned} \tag{22}$$

where $\mathbf{e}_i$ is the ith standard basis vector in which element i is equal to one and all other elements are equal to zero. In Eq. (22), the term $\binom{\varphi_i}{2}\frac{1}{N_i(t)}$ is the instantaneous rate at which a coalescent

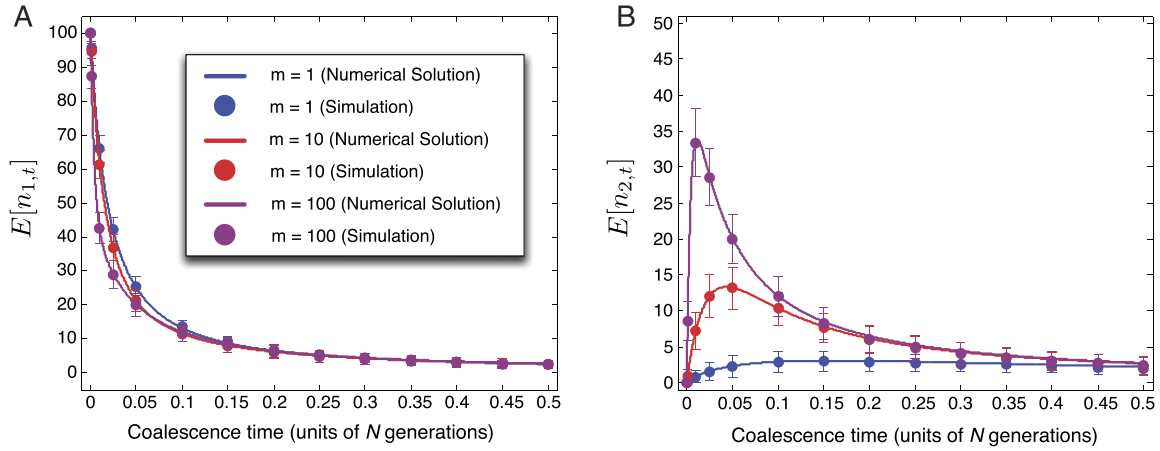

Fig. 4. The accuracy of the approximation to $E[n_t]$ under migration (Eq. (26)) for two populations of time-varying sizes $N_1(t)$ and $N_2(t)$. The two populations have the same size $N_1(t) = N_2(t) = N(t)$, which grows faster-than-exponentially over time according to the formula $dN(t)/dt = -\alpha N(t)^\beta$, where $\alpha = 10$, $\beta = 5$, and $N(0) = 1$. The migration rates satisfy $m_{12} = m_{21} = m$. **(A)** Curves show the approximation of $E[n_{1,t}]$ (the expected number of lineages in population 1 at time t) obtained by numerically solving Eq. (26). Dots show the estimates of $E[n_{1,t}]$ at the times $t = 0.001, 0.01, 0.025, 0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45$ and 0.5 obtained by simulating 10^3 genealogies under the coalescent model according to the transition probabilities in Eq. (22) and computing the number of lineages over this grid of times. Different colored lines correspond to different values of m. The total length of each error bar is equal to two standard deviations of $n_{1,t}$ or $n_{2,t}$, estimated from the 10^3 replicate simulations (10^3 sampled genealogies). **(B)** The corresponding plot for $E[n_{2,t}]$, the expected number of lineages in population 2 at time t. For each value of m, $n_{1,0} = 100$ and $n_{2,0} = 0$ lineages were sampled from populations 1 and 2, respectively.

event occurs in population i, and $\varphi_i m_{ij}$ is the instantaneous rate at which a lineage migrates from population i to population j, when φ_i lineages remain in population i at time t. The notation $\varphi' = \varphi - \mathbf{e}_i$ indicates that a coalescent event occurred in population i between the state φ at time t and the state φ' at time $t + \delta$. Eq. (22) is the generalization of the transition probabilities used in the derivation of Volz et al. (2009, p.1880).

Using the transition probabilities in Eq. (22) and conditioning on the state at time t, we obtain the following conditional expression for $\mathbb{P}(\mathbf{n}_{t+\delta} = \varphi)$, which we denote by $p_\varphi(t + \delta)$:

$$\begin{aligned} p_\varphi(t+\delta) = &\left[1 - \sum_{i=1}^{k}\binom{\varphi_i}{2}\frac{1}{N_i(t)}\delta - \sum_{i=1}^{k}\sum_{\substack{j=1\\ j\neq i}}^{k}\varphi_i m_{ij}\delta\right] p_\varphi(t) \\ &+ \sum_{i=1}^{k}\sum_{\substack{j=1\\ j\neq i}}^{k}(\varphi_i+1)m_{ij}\delta p_{\varphi+\mathbf{e}_i-\mathbf{e}_j}(t) \\ &+ \sum_{i=1}^{k}\binom{\varphi_i+1}{2}\frac{1}{N_i(t)}\delta p_{\varphi+\mathbf{e}_i}(t) + o(\delta). \end{aligned} \tag{23}$$

Subtracting the term $p_\varphi(t)$ from both sides, dividing by δ, and letting $\delta \to 0$ gives the differential equation

$$\begin{aligned} \frac{dp_\varphi(t)}{dt} = &-\sum_{i=1}^{k}\binom{\varphi_i}{2}\frac{1}{N_i(t)}p_\varphi(t) - \sum_{i=1}^{k}\sum_{\substack{j=1\\ j\neq i}}^{k}\varphi_i m_{ij}p_\varphi(t) \\ &+ \sum_{i=1}^{k}\sum_{\substack{j=1\\ j\neq i}}^{k}(\varphi_i+1)m_{ij}p_{\varphi+\mathbf{e}_i-\mathbf{e}_j}(t) \\ &+ \sum_{i=1}^{k}\binom{\varphi_i+1}{2}\frac{1}{N_i(t)}p_{\varphi+\mathbf{e}_i}(t). \end{aligned} \tag{24}$$

To obtain the differential equation for $E[n_{\ell t}]$ ($\ell = 1, \ldots, k$), we can multiply both sides of Eq. (24) by φ_ℓ and sum over φ_ℓ (Appendix E) to obtain

$$\begin{aligned} \frac{dE[n_{\ell t}]}{dt} = &-\binom{E[n_{\ell t}]}{2}\frac{1}{N_\ell(t)} - \frac{\mathrm{Var}(n_{\ell t})}{2N_\ell(t)} \\ &+ \sum_{\substack{i=1\\ i\neq\ell}}^{k}(m_{i\ell}E[n_{it}] - m_{\ell i}E[n_{\ell t}]). \end{aligned} \tag{25}$$

If we assume that $\mathrm{Var}(n_{\ell t}) = 0$, then we obtain the system of k approximate differential equations

$$\frac{dE[n_{\ell t}]}{dt} \approx -\binom{E[n_{\ell t}]}{2}\frac{1}{N_\ell(t)} + \sum_{\substack{i=1\\ i\neq\ell}}^{k}(m_{i\ell}E[n_{it}] - m_{\ell i}E[n_{\ell t}]) \tag{26}$$

for $\ell = 1, \ldots, k$, which can be solved numerically to obtain approximations of $E[n_{\ell t}]$.

The accuracy of the approximation obtained by solving the system of equations in Eq. (26) is shown in Fig. 4 for the case of two populations with migration among them. The populations have equal and exponentially growing sizes given by $N_1(t) = N_2(t) = N(t)$, where $N(t)$ satisfies the differential equation

$$N'(t) = -\alpha N(t)^\beta. \tag{27}$$

This equation represents the model of super-exponential growth proposed by Reppell et al. (2012). When $\beta = 1$, the population size changes exponentially over time according to $N(t) = N(0)e^{-\alpha t}$. In the example in Fig. 4, we have constrained the migration rates to be equal, and we consider the case in which $n_{1,0} = 100$ lineages are sampled from the first population and $n_{2,0} = 0$ lineages are sampled from the second population. From Fig. 4, it can be seen that the approximation obtained by solving Eq. (26) is accurate across a range of migration rates.

5. Applications

In this section, we apply the approximations in Eqs. (4), (7) and (10) to a set of example problems that demonstrate their utility for approximating coalescent formulas. We explore the accuracy of the resulting approximations using Theorems 2.1 and 3.1. We also demonstrate how approximations of $E[n_t]$ for the case of multiple populations with migration (Eq. (26)) can be used to obtain approximate coalescent formulas under complicated demographic scenarios.

5.1. The expected joint allele frequency spectrum

We first consider the problem of approximating Wakeley and Hey's (1997) formula for the expected joint allele frequency spectrum between a pair of populations without migration. In Wakeley

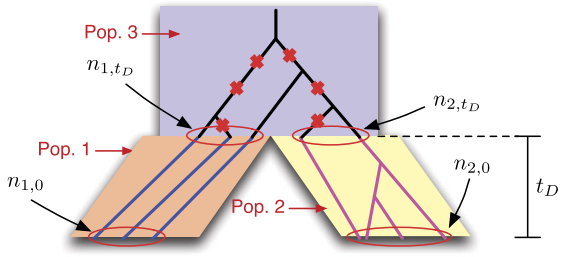

Fig. 5. Wakeley and Hey's model for computing the expected joint allele frequency spectrum between a pair of populations. Two daughter populations, 1 and 2, diverge at time t_D in the past from an ancestral population (population 3). At the present time $t = 0$, $n_{1,0}$ and $n_{2,0}$ lineages are sampled from populations 1 and 2, respectively. Wakeley and Hey's formula for the expected joint allele frequency spectrum computes the expected number z_{ij} of segregating sites at which the derived allele appears in i copies in the sample from population 1 and in j copies in the sample from population 2, where $i \in \{1, \ldots, n_{1,0}\}$ and $j \in \{1, \ldots, n_{2,0}\}$. The model considers only mutations that arose in the ancestral population (red crosses).

and Hey's model, two populations diverge at time t_D from an ancestral population (Fig. 5). A sample of $n_{1,0}$ alleles is taken from the first population and a sample of $n_{2,0}$ alleles is taken from the second population. Let z_{ij} be the random variable recording the number of polymorphic sites for which the derived allele appears in i copies in the sample from the first population and in j copies in the sample from the second population. The expected joint allele frequency spectrum (JAFS) for the two populations is the collection of expectations $E[z_{ij}]$ for $i = 1, \ldots, n_{1,0} - 1$ and $j = 1, \ldots, n_{2,0} - 1$.

The expected JAFS is useful for performing inference on demographic parameters such as divergence times and ancestral population sizes (Wakeley and Hey, 1997; Gutenkunst et al., 2009; Nielsen et al., 2009). Wakeley and Hey's formula for the expected JAFS is of the form

$$E[z_{ij}] = \sum_{n_{1,t_D}=2}^{n_{1,0}} \sum_{n_{2,t_D}=2}^{n_{2,0}} C_{ij}(n_{1,t_D}, n_{2,t_D}) \mathbb{P}(n_{1,t_D}) \mathbb{P}(n_{2,t_D}). \tag{28}$$

Here, t_D is the divergence time between the two populations, and

$$C_{ij}(n_{1,t_D}, n_{2,t_D}) = \sum_{k_1=1}^{n_{1,t_D}-1} \sum_{k_2=1}^{n_{2,t_D}-1} P(k_1 \to i | n_{1,0}, n_{1,t_D}) \times P(k_2 \to j | n_{2,0}, n_{2,t_D}) \frac{\binom{n_{1,t_D}}{k_1}\binom{n_{2,t_D}}{k_2}}{\binom{n_{1,t_D}+n_{2,t_D}}{k_1+k_2}} \frac{\theta_3}{k_1+k_2}, \tag{29}$$

where $P(k \to i | n, n') = \binom{n-n'}{i-k} k_{(i-k)}(n'-k)_{(n-n'-i+k)}/n'_{(n-n')}$, and where $\theta_3 = 4N_3\mu b$ is the population-scaled mutation rate in the sequence of length b bases in the ancestral population of size N_3.

The term $C_{ij}(n_{1,t_D}, n_{2,t_D})$ is time-consuming to evaluate, and the formula in Eq. (28) quickly becomes computationally burdensome as $n_{1,0}$ and $n_{2,0}$ increase in size (Fig. 6A). Dependence on the distribution $\mathbb{P}(n_{t_D})$ also leads to round-off error when $n_{1,0}$ or $n_{2,0}$ is large and t_D is small. This round-off error is visible in Fig. 6B as points that deviate from the smooth curve for sample sizes greater than $n_{1,0} = n_{2,0} \approx 60$.

5.1.1. Approximating the JAFS

Although Griffiths' approximation (Eq. (3)) can eliminate the round-off error in evaluating Eq. (28), the time needed to compute the formula using Griffiths' approximation is nearly the same as the time needed to compute the exact formula (Fig. 6A). In addition, the approximation deviates from the true value when the sample size is small (Fig. 7A).

Instead of using Griffiths' approximation, we can approximate Eq. (28) using the deterministic approximation (Eq. (4)). In particular, we can approximate Eq. (28) as

$$E[z_{ij}] \approx C_{ij}(E[n_{1,t_D}], E[n_{2,t_D}]). \tag{30}$$

The expectations $E[n_{1,t_D}]$ and $E[n_{2,t_D}]$ in Eq. (30) can be computed using Eq. (13), or they can be approximated using Eq. (17) or Eq. (18). Because $E[n_{1,t_D}]$ and $E[n_{2,t_D}]$ are not generally integer-valued, the factorials and binomial coefficients in Eq. (29) can be computed by reformulating them in terms of gamma functions using the definitions $n! = \Gamma(n+1)$ and $\binom{n}{k} = n!/[k!(n-k)!] = \Gamma(n+1)/[\Gamma(k+1)\Gamma(n-k+1)]$. The result of the approximation is a considerable reduction in computation time (Fig. 5A) and a considerable improvement in accuracy both for small and for large sample sizes (Fig. 5B).

5.1.2. The accuracy and computational complexity of the approximation in Eq. (30)

Theorem 3.1 tells us that when the second partial derivatives $\partial^2_{n_{1,t_D}} C_{ij}(n_{1,t_D}, n_{2,t_D})$ and $\partial^2_{n_{2,t_D}} C_{ij}(n_{1,t_D}, n_{2,t_D})$ exist and are continuous and bounded in n_{1,t_D} and n_{2,t_D}, then the approximation in Eq. (30) converges to the true distribution in Eq. (28) as $t \to 0$ and as $t \to \infty$. Because Eq. (29) is a finite sum of fractions of gamma functions in n_{1,t_D} and n_{2,t_D}, which are smooth, bounded, and nonzero for $(n_{1,t_D}, n_{2,t_D}) \in [1, n_{1,0}] \times [1, n_{2,0}]$, the second partial derivatives of Eq. (29) are smooth and bounded on $[1, n_{1,0}] \times [1, n_{2,0}]$. Therefore, for fixed values of $n_{1,0}$ and $n_{2,0}$, the error in the approximation in Eq. (30) decreases to zero as $t \to 0$ and as $t \to \infty$.

We can also estimate the magnitude of the error in the deterministic approximation using the result in Appendix D. In particular, because the lineages in populations 1 and 2 coalesce independently of one another, we can estimate the error using Eq. (12), which applies when $n_{1,t}$ and $n_{2,t}$ are independent. In Eq. (12), the variances $\mathrm{Var}(n_{1,t_D})$ and $\mathrm{Var}(n_{2,t_D})$ can be computed using Tavaré's formula given in Eq. (B.3). Because the second partial derivatives $\partial^2_{n_{1,t_D}} C_{ij}(n_{1,t_D}, n_{2,t_D})$ and $\partial^2_{n_{2,t_D}} C_{ij}(n_{1,t_D}, n_{2,t_D})$ are difficult to compute analytically, we can evaluate them using finite-difference approximations; in this example, we used the second-order forward finite-difference approximation.

The asymptotic accuracy of the approximation in Eq. (30) can be seen in Fig. 7A for the term $E[z_{11}]$. In particular, the blue curve, which corresponds to the error in the deterministic approximation, approaches zero as $t \to 0$ and as $t \to \infty$. From Fig. 7A it can also be seen that the estimated error in the approximation to the term $E[z_{11}]$ closely matches the true error, and that it is approximately equal to the true error in the limits $t \to 0$ and $t \to \infty$. The error is also small for the other terms in the JAFS. For example, for the fixed value $t_D = 0.01$ and for $n_{1,0} = n_{2,0} = 30$, the fit of the approximation in Eq. (30) is very accurate for all values of i and j (Fig. 7B).

In contrast with the deterministic approximation, the error in Griffiths' approximation (the green curve in Fig. 7A) does not converge to zero as $t \to 0$. Although Griffiths' approximation is less accurate than the deterministic approximation for the particular choice of parameter values considered here, Griffiths' approximation is guaranteed to converge to the exact value as $t \to 0$ and as $n_{1,0}$ and $n_{2,0}$ increase to infinity. Thus, the accuracy of Griffiths' approximation will improve for larger sample sizes.

5.2. Expected numbers of segregating sites under migration

In this section, we demonstrate how approximate expected numbers of segregating sites can be computed under complicated demographic scenarios involving variable population sizes and migration. In particular, we combine Eq. (10) with approximations

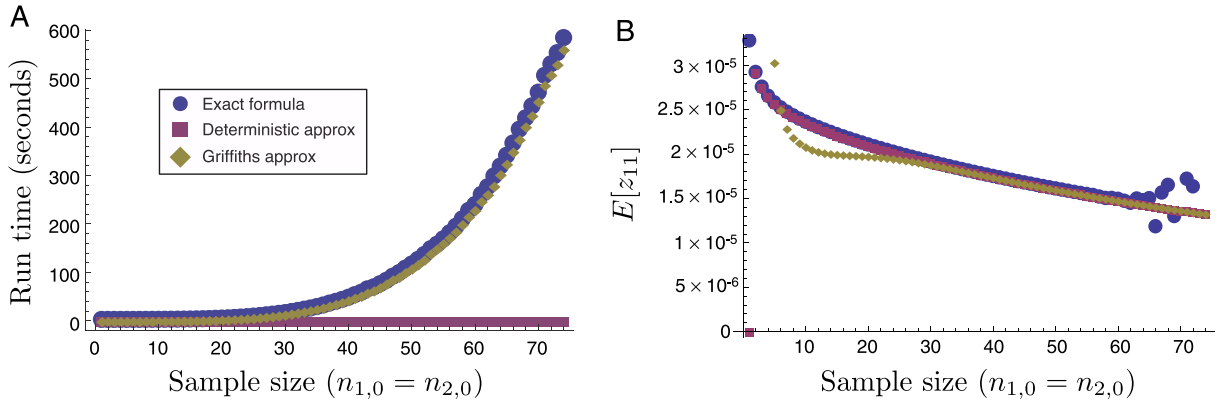

Fig. 6. Three different approaches for computing the first term $E[z_{11}]$ in the expected joint allele frequency spectrum between two populations for different numbers $n_{1,0}$ and $n_{2,0}$ of sampled alleles, with $n_{1,0} = n_{2,0}$: Wakeley and Hey's (1997) exact formula (Eq. (28), blue), the deterministic approximation computed using Eq. (4) (magenta), and Griffiths' approximation computed using Eq. (3) (green). (A) The run time is shown as a function of the sample sizes $n_{1,0}$ and $n_{2,0}$ in the two populations. (B) The value from each of three methods for computing the term $E[z_{11}]$.

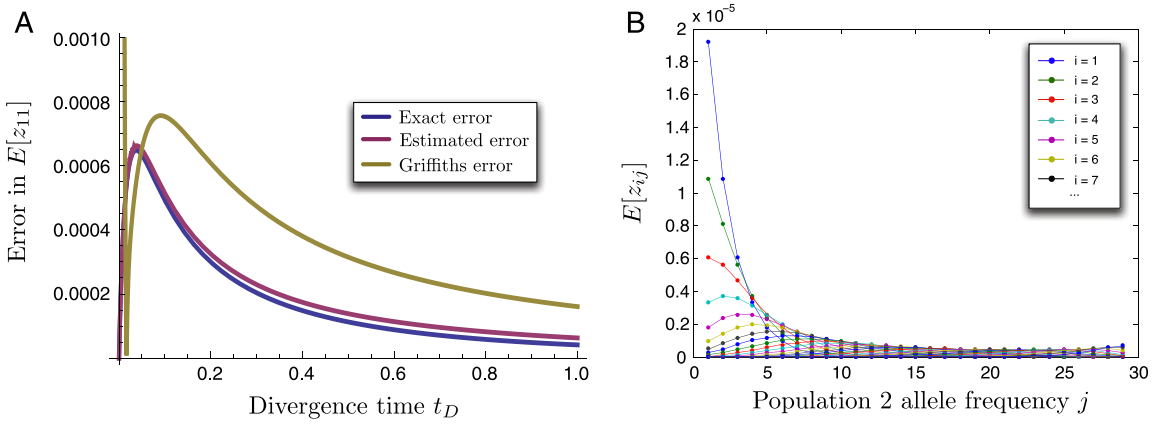

Fig. 7. The error in approximating the expected joint allele frequency spectrum (JAFS). (A) The error in approximating the first term $E[z_{11}]$ in the expected JAFS between two populations for the case of $n_{1,0} = n_{2,0} = 30$ sampled alleles for a range of divergence times t_D. The absolute value of the exact error in the deterministic approximation of Wakeley and Hey's (1997) formula for the expected JAFS (Eq. (30), blue), the estimated error of the approximation in Eq. (30) obtained using Eq. (12) (magenta), and the error in Griffiths' approximation obtained using Eq. (3) (green) are shown. (B) Comparison of Wakeley and Hey's exact formula (Eq. (28)) with the deterministic approximation (Eq. (30)) for all possible combinations of allele counts i and j for the case of constant and equal-sized populations ($N_1 = N_2 = N_3 = N$), for sample sizes $n_{1,0} = n_{2,0} = 30$, and for a divergence time of $t_D = 0.01$ coalescent units of N generations. The value of Wakeley and Hey's formula is shown with a solid line, and the deterministic approximation is shown with dots.

of $E[n_t]$ obtained using Eq. (26) to compute the expected number of private alleles in a sample from a population. Private alleles are useful for studying the historical relationships among populations (Tishkoff and Kidd, 2004; Szpiech et al., 2008), and the number of private alleles is a commonly-used measure of distinctiveness in conservation studies (e.g., Kalinowski, 2004; Wilson et al., 2012; Ariani et al., 2013).

In this example, we again consider two populations, 1 and 2, that diverged at time t_D in the past and that have continued to share migrants since their divergence (Fig. 8A). Let $N_1(t)$ and $N_2(t)$ be the sizes of populations 1 and 2 at time t in the past. We consider the case in which each population has grown faster-than-exponentially over time (Eq. (27)) according to $N_i'(t) = \alpha N_i(t)^\beta$ ($i = 1, 2$), where α and β are the same for both populations. We assume that $n_{1,0}$ and $n_{2,0}$ alleles were sampled from populations 1 and 2, respectively (Fig. 8).

5.2.1. *Approximating the expected number of private segregating sites in a sample*

Let S_1 be the number of mutations that are observed in a region of length b bases in a sample of $n_{1,0}$ lineages from population 1 and not in a sample of $n_{2,0}$ lineages from population 2. The expectation $E[S_1]$ can be obtained by computing the total sum of lengths L_1 of genealogy branches that are ancestral only to the sample from population 1 (Fig. 8B). Using Eqs. (9) and (10), $E[S_1]$ can be computed as

$$E[S_1] = \frac{\theta b}{4} E[L_1] = \frac{\theta b}{4} \int_{z=0}^{\infty} E[\tilde{n}_{1,z}]dz, \tag{31}$$

where $\tilde{n}_{1,t}$ is the number of lineages that are ancestral only to the sample from population 1 and that are not ancestral to the sample from population 2.

To compute $E[\tilde{n}_{1,t}]$, we can solve Eq. (26) for two populations with initial conditions in which $n_{1,0}$ and $n_{2,0}$ alleles are initially sampled from populations 1 and 2, respectively. The solution gives us $E[n_{1,t}]$ and $E[n_{2,t}]$, the numbers of ancestral lineages remaining in populations 1 and 2, respectively, at time t in the past. Solving the system again with initial conditions in which no alleles are sampled from population 1 and in which $n_{2,0}$ alleles are sampled from population 2 yields the solutions $E[y_{1,t}]$ and $E[y_{2,t}]$, which are the numbers of lineages in populations 1 and 2, respectively, that are ancestral at time t to the $n_{2,0}$ lineages sampled from population 2.

The number of lineages $E[\tilde{x}_{1,t}]$ in population 1 at time t that are ancestral only to the sample from population 1, and not to the sample from population 2, is then given by $E[\tilde{x}_{1,t}] = E[n_{1,t}] - E[y_{1,t}]$. Similarly, the number of lineages $E[\tilde{x}_{2,t}]$ in population 2 at time t that are ancestral only to the sample from population 1, and

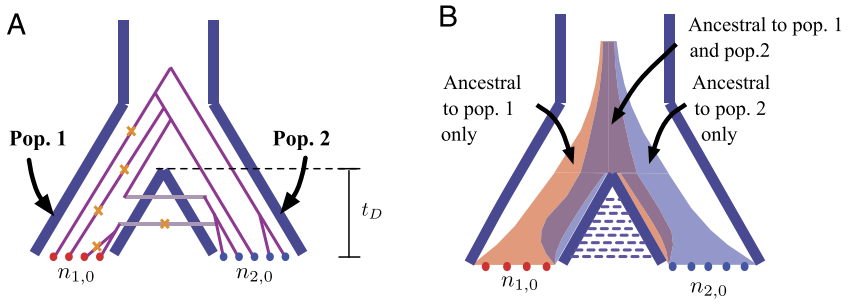

Fig. 8. Comparison of stochastic and deterministic coalescent models for computing the expected number of mutations that are private to a sample of alleles from a population. In each model, two populations, 1 and 2, diverge at time t_D in the past. Samples of sizes $n_{1,0}$ and $n_{2,0}$ are taken from populations 1 and 2, respectively. (A) The classical stochastic coalescent model. Orange crosses indicate mutations that occur on lineages that are ancestral only to the sample from population 1. (B) The deterministic coalescent model. The red region indicates lineages ancestral only to the sample from population 1, the blue region indicates lineages ancestral only to the sample from population 2, and the purple region indicates lineages ancestral to both samples. The width of the shaded region of each color in each population at a fixed time t is the expected number of lineages of the given type in the given population at that time. The total sum of branch lengths on which a mutation ancestral only to the sample from population 1 can occur is the area of the region shaded in red.

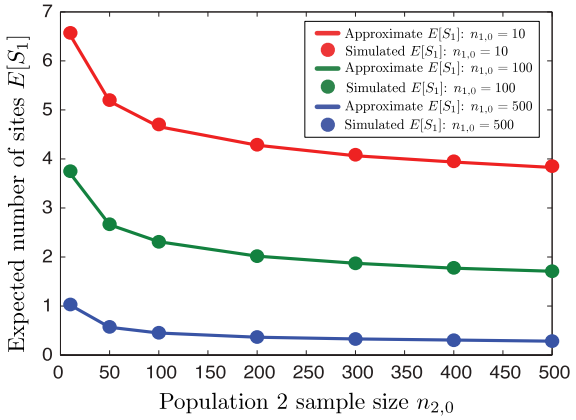

Fig. 9. Comparison with simulations of analytical approximations of $E[S_1]$ obtained using Eq. (31) with simulations.

not to the sample from population 2, is given by $E[\tilde{x}_{2,t}] = E[n_{2,t}] - E[y_{2,t}]$. The expected total number of lineages ancestral only to the sample from population 1 is given by $E[\tilde{n}_{1,t}] = E[\tilde{x}_{1,t}] + E[\tilde{x}_{2,t}]$. The expectation $E[S_1]$ is then obtained by plugging the value of $E[\tilde{n}_{1,t}]$ into Eq. (31) for a given choice of θ and b.

Theorem 2.1 implies that Eq. (31) is exact if $E[\tilde{n}_{1,t}]$ is exact. However, because the differential equation in Eq. (26) is approximate, there will be a small amount of error in our computation of $E[S_1]$. We examine this error empirically in Section 5.2.2.

5.2.2. The accuracy of the approximation in Eq. (31)

To examine the error in Eq. (31) that arises from the approximation in Eq. (26), we compared the analytical results obtained using Eqs. (26) and (31) to simulations. Simulations were performed by sampling genealogies from the Markov chain with transition probabilities given by Eq. (22) using an approach similar to that described by Jewett et al. (2012). We discuss the simulation procedure in more detail in Appendix F.

Approximations of $E[S_1]$ appear in Fig. 9 for various sample sizes $n_{1,0}$ and $n_{2,0}$, along with simulated values for comparison. In our computations and simulations, we have taken $N_1(0) = N_2(0) = 1$, and we have set $N_3(t) = N_1(t_D) + N_2(t_D)$ at the divergence time t_D. The other parameters were chosen in order to model moderate levels of faster-than-exponential growth and migration: $\alpha = 5$, $\beta = 10$, and $m_{12} = m_{21} = 10$. Because the parameters b and θ in our model only affect the computed values of $E[S_1]$ by a constant scaling factor, we set each of these values to unity for simplicity ($b = 1$ and $\theta = 1$). From Fig. 9, it can be seen that the approximation is very accurate over the range of parameter values, even when the sample sizes are small.

5.3. The time to the first inter-sample coalescent event

In the examples in Sections 5.1 and 5.2, we have used the approximation $n_t \approx E[n_t]$ to compute expected values. However, the approximation can also be used to derive approximate probability distributions. For example, Volz et al. (2009) used a version of the approximation in Eq. (4) to compute the joint distribution of coalescent waiting times among a set of sampled lineages in a single population of variable size (Volz et al., 2009, Eq. (12)). Here, we consider the related problem of computing the distribution of the time until the first coalescent event between two different sets of sampled alleles in a model with two populations of variable size with migration among them (Fig. 10).

We again consider a model in which two populations diverge at time t_D from a common ancestral population (Fig. 10). Consider a sample of $n_{1,0}$ alleles from one or both of the populations, and denote these as "type-1" alleles. Suppose that a second sample of $n_{2,0}$ alleles is taken from one or both populations and denote these as "type-2" alleles. We refer to lineages ancestral to type-1 alleles as "type-1" lineages, and we refer to lineages ancestral to type-2 alleles as "type-2" lineages. We are interested in computing the distribution of the random time V until the first coalescent event occurs between a type-1 lineage and a type-2 lineage when the migration rates between the populations are nonzero. We refer to a coalescent event between a type-1 lineage and a type-2 lineage as an *inter-sample* coalescent event.

Inter-sample coalescence times have a number of applications. For example, when the type-1 and type-2 alleles are sampled from two different populations, the time to the first inter-sample coalescent event can be used to estimate the divergence time of the two populations (Takahata and Nei, 1985; Mossel and Roch, 2010; Liu et al., 2010; Jewett and Rosenberg, 2012). When $n_{1,0} = 1$, the distribution of the time to the first inter-sample coalescent event can be used to compute the probability of observing a new haplotype, conditional on an observed set of $n_{2,0}$ haplotypes (Paul and Song, 2010), or to predict the accuracy of imputing genotypes on a haplotype using a reference panel of existing haplotypes (Jewett et al., 2012; Huang et al., 2013). The expected time of the first inter-sample coalescent event was computed in a migration model using simulations by Takahata and Slatkin (1990). Here, we show how a simple approximate analytical distribution can be derived using Eq. (26).

5.3.1. Approximating the distribution of the inter-sample coalescence time

At time t in the past, suppose that $x_{1,t}$ type-1 lineages and $y_{1,t}$ type-2 lineages remain in population 1 and suppose that $x_{2,t}$ type-1 lineages and $y_{2,t}$ type-2 lineages remain in population 2. Under

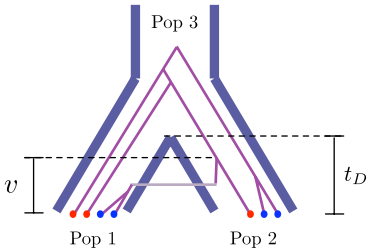

Fig. 10. The time v until the first coalescent event occurs between an ancestor of one of $n_{1,0}$ type-1 alleles (red) and an ancestor of one of $n_{2,0}$ type-2 alleles (blue). The alleles are sampled from two populations, 1 and 2, of sizes $N_1(t)$ and $N_2(t)$ that diverged at time t_D from an ancestral population (population 3) of size $N_3(t)$.

the classical stochastic coalescent model, the instantaneous rate of coalescence between type-1 and type-2 lineages in population 1 is $x_{1,t}y_{1,t}/N_1(t)$ and the instantaneous rate of coalescence among type-1 and type-2 lineages in population 2 is $x_{2,t}y_{2,t}/N_2(t)$. Therefore, because lineages can only coalesce within the same population, the instantaneous rate of coalescence among type-1 and type-2 lineages overall is $x_{1,t}y_{1,t}/N_1(t) + x_{2,t}y_{2,t}/N_2(t)$.

Let $x_{1,[0,\infty]}$, $x_{2,[0,\infty]}$, $y_{1,[0,\infty]}$, and $y_{2,[0,\infty]}$ denote sample paths of the stochastic processes describing the numbers of ancestors of each type in the time interval $[0, \infty]$, and denote the collection of these paths by $\mathbf{x}_{[0,\infty]}$. Conditional on the sample paths $\mathbf{x}_{[0,\infty]}$ and on the event that no inter-sample coalescent event has occurred by time t, it follows that in the small time interval $[t, t+\delta]$, the probability that no inter-sample coalescent event occurs is

$$\begin{aligned}\mathbb{P}(\mathcal{1}_{[t,t+\delta]}|\mathcal{1}_{[0,t]}, \mathbf{x}_{[0,\infty]}) &\approx 1 - (x_{1,t}y_{1,t}/N_1(t) + x_{2,t}y_{2,t}/N_2(t))\delta \\ &\approx \exp\{-(x_{1,t}y_{1,t}/N_1(t) + x_{2,t}y_{2,t}/N_2(t))\delta\}, \end{aligned} \tag{32}$$

where $\mathcal{1}_{[r,s]}$ is the event that no inter-sample coalescence occurs in the time interval $[r, s]$. Thus, conditional on the sample paths $\mathbf{x}_{[0,\infty]}$, the probability that no inter-sample coalescent event occurs in any of v/δ small time intervals of length δ between time 0 and time v is given approximately by

$$\begin{aligned}&\mathbb{P}(\mathcal{1}_{[0,\delta]}, \mathcal{1}_{[\delta,2\delta]}, \ldots, \mathcal{1}_{[v-\delta,v]}|\mathbf{x}_{[0,\infty]}) \\ &\quad\approx \prod_{i=1}^{v/\delta} e^{-(x_{1,i\delta}y_{1,i\delta}/N_1(i\delta)+x_{2,i\delta}y_{2,i\delta}/N_2(i\delta))\delta} \\ &\quad= e^{-\sum_{i=1}^{v/\delta}(x_{1,i\delta}y_{1,i\delta}/N_1(i\delta)+x_{2,i\delta}y_{2,i\delta}/N_2(i\delta))\delta}. \end{aligned} \tag{33}$$

A similar result was obtained for the case of a single population by Jewett and Rosenberg (2012).

By letting $\delta \to 0$ in Eq. (33), we obtain an approximation of the survival function $S_{V|\mathbf{x}}(v)$ of the time until the first inter-sample coalescent event, conditional on the sample paths $\mathbf{x}_{[0,\infty]}$:

$$\begin{aligned}S_{V|\mathbf{x}}(v) &= \mathbb{P}(\mathcal{1}_{[0,v]}|\mathbf{x}_{[0,\infty]}) \\ &\approx e^{-\int_{z=0}^{v}(x_{1,z}y_{1,z}/N_1(z)+x_{2,z}y_{2,z}/N_2(z))dz}. \end{aligned} \tag{34}$$

The unconditional survival function $S(v)$ can be obtained by integrating over all sample paths as follows:

$$\begin{aligned}S(v) &= \int_{\mathcal{A}_{[0,\infty]}} S_{V|\mathbf{x}}(v)p(\mathbf{x}_{[0,\infty]})d\mathcal{A}_{[0,\infty]} \\ &= \int_{\mathcal{A}_{[0,\infty]}} e^{-\int_{z=0}^{v}(x_{1,z}y_{1,z}/N_1(z)+x_{2,z}y_{2,z}/N_2(z))dz} \\ &\quad\times p(\mathbf{x}_{[0,\infty]})d\mathcal{A}_{[0,\infty]}, \end{aligned} \tag{35}$$

where $p(\mathbf{x}_{[0,\infty]})$ is the probability density function of the sample paths $\mathbf{x}_{[0,\infty]}$. Eq. (35) is of the form given in Eq. (5), which is time-consuming to compute due to the integral over all sample paths $\mathbf{x}_{[0,\infty]}$. However, using an approximation of the form given in Eq. (7), we can approximate $S(v)$ by

$$S(v) \approx e^{-\int_{z=0}^{v}(E[x_{1,z}]E[y_{1,z}]/N_1(z)+E[x_{2,z}]E[y_{2,z}]/N_2(z))dz}. \tag{36}$$

Compared with Eq. (35), Eq. (36) is considerably faster to compute and it has a simple functional form.

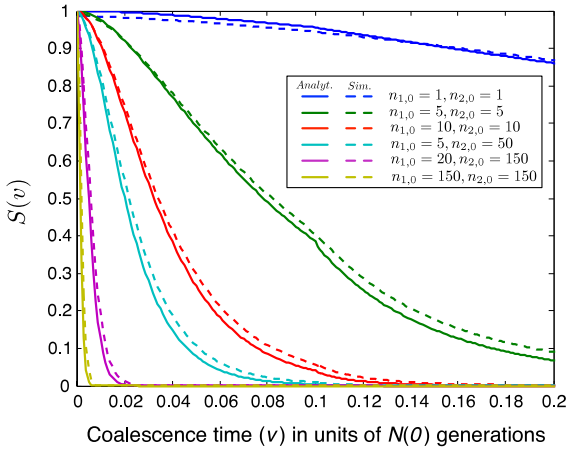

Fig. 11. Kernel density estimates (dashed lines) and analytical approximations (solid lines) of the survival function $S(v)$ of the time V to the first inter-sample coalescent event between two samples of lineages taken from two separate populations. Analytical approximations were computed using Eq. (36). These values were generated using a model in which the type-1 and type-2 lineages were sampled from different populations (Fig. 10) that diverged at time $t_D = 0.1$ and that had equal and faster-than-exponentially growing sizes given by $N_1(t) = N_2(t) = N(t)$. The migration rates between the populations in the time interval $[0, t_D]$ are $m_{12} = m_{21} = 10$. The population sizes $N(t)$ satisfied Eq. (27) with $N(0) = 1$, $\beta = 10$, and $\alpha = 5$. The ancestral population was of constant size $N_3(t) = 1$ for $t \geq t_D$. The sharp change in the slope of the curves at time $v = 0.1$ is due to the instantaneous transition from two populations to a single population at the divergence time t_D.

5.3.2. *The accuracy of the approximation in Eq. (36)*

We compared the approximate distribution $S(v)$ given in Eq. (36) with kernel density estimates of $S(v)$ from simulations (Appendix F). In our example, we considered a scenario in which the type-1 and type-2 lineages were sampled from different populations that diverged at time $t_D = 0.1$ and which had equal and faster-than-exponentially growing sizes given by $N_1(t) = N_2(t)$. The population sizes $N_i(t)$ $(i = 1, 2)$ satisfied Eq. (27) with $N_i(0) = 1$, $\beta = 10$, and $\alpha = 5$. The ancestral population was of constant size $N_3(t) = 1$ for $t \geq t_D$.

To obtain kernel density estimates of $S(v)$, we simulated genealogies from a coalescent model with transition probabilities given by Eq. (22) as described in Appendix F. Fig. 11 shows comparisons of $S(v)$ computed using Eq. (36) with kernel densities computed from 10^5 replicates for a variety of different sample sizes $n_{1,0}$ and $n_{2,0}$. From the density plots, it can be seen that the approximation is very accurate, even when the sample sizes are small.

6. Discussion

In this paper, we have considered the accuracy and applications of the deterministic approximation $n_t \approx E[n_t]$ for deriving approximate coalescent distributions that are fast and numerically stable to compute. In particular, we identified ways in which the approximation $n_t \approx E[n_t]$ can be applied procedurally to reduce the computational complexity and numerical instability of coalescent formulas that involve conditional summations over all possible values of n_t, or that involve integrals over all possible sample paths $n_{[r,s]}$ of the coalescent process describing the number of ancestral lineages in a given time interval $[r, s]$.

We have considered two different kinds of approximation. In Sections 2 and 3, we considered the approximation of n_t by its expected value $E[n_t]$. In Section 4, we considered a second kind of approximation: approximate formulas for $E[n_t]$. The first approximation, of n_t by $E[n_t]$, holds whenever the behavior of n_t is nearly deterministic. As we showed in Lemma B.1, this deterministic behavior occurs in the limit as $t \to 0$ and as $t \to \infty$. By contrast, the range of values over which any given approximation of $E[n_t]$ is valid depends on the approximation that is used. For instance, in Fig. 3, we saw that the approximate function in Eq. (18) is sensible in the limit as $t \to 0$ and as $t \to \infty$, whereas the simpler approximation in Eq. (17) is sensible only in the limit as $t \to 0$.

To facilitate the application of these approximations in practice, we showed that approximate coalescent formulas of the form given in Eq. (4) converge to their true values as $t \to 0$ and as $t \to \infty$ under simple assumptions. We also derived an approximate expression for the error in these deterministic approximations (Eq. (11)). This approximate expression for the error can be used in practice to evaluate when any given approximate formula of the form given in Eq. (4) is accurate.

We obtained approximate formulas for $E[n_t]$ in the case of multiple populations with time-varying sizes and migration among them (Eq. (26)). These approximations were produced by extending differential equations for $E[n_t]$ derived for the case of a single panmictic population by Slatkin and Rannala (1997), Volz et al. (2009), and Maruvka et al. (2011). The approximations of $E[n_t]$ that we obtained facilitate the derivation of approximate coalescent formulas under complicated demographic scenarios. For example, we showed how approximations of $E[n_t]$ under migration could be used to approximate the expected number of mutations occurring along the branches of a genealogy (Section 5.2) or to compute an approximate distribution of coalescent waiting times (Section 5.3) in demographic models involving multiple populations with migration. Such applications of the approximation $n_t \approx E[n_t]$ are useful because deriving exact formulas for coalescent quantities under models with both migration and population size changes can be difficult.

We have described a number of problems to which the approximation $n_t \approx E[n_t]$ can be applied. However, we have focused on quantities that can be derived conditional on knowledge of the total number of ancestral lineages remaining at a given time t or over a given time interval $[r, s]$. Quantities that require knowledge of the topology of the coalescent tree relating the ancestral lineages, or of the number of lineages of a particular type, may be more difficult to derive. It is likely that the approximation $n_t \approx E[n_t]$ can be used to derive a variety of approximate distributions beyond those discussed here; however, the approximation $n_t \approx E[n_t]$ must be applied in a new way for each new class of problem, and the theoretical accuracy of these applications must be evaluated anew.

One common use of the approximation $n_t \approx E[n_t]$ that we did not consider in this paper is the inference of the size of a population at each time in the past by fitting the observed values of n_t obtained from a reconstructed genealogy of a set of sampled alleles to the expected values $E[n_t]$ $(t \geq 0)$ under a given demographic history (Frost and Volz, 2010; Maruvka et al., 2011). The theoretical accuracy of such fitting approaches is difficult to determine analytically and remains a subject for further work.

The importance of coalescent approximations has been a subject of much recent interest, as it has become increasingly recognized that exact formulas or algorithms can be intractable in practical scenarios. Many recent studies have made use of a variety of simplifying assumptions and approximations to the coalescent, and to coalescent-like problems (Li and Stephens, 2003; McVean and Cardin, 2005; Marjoram and Wall, 2006; Davison et al., 2009; Paul and Song, 2010; RoyChoudhury, 2011; Li and Durbin, 2011; Sheehan et al., 2013). Our results on the approximation $n_t \approx E[n_t]$ contribute to this growing toolbox of coalescent-based approximations that can be used to derive functionally simple, computationally efficient, and numerically stable approximations of coalescent formulas under a variety of coalescent models. These, and similar kinds of approximations, will become increasingly important for making population-genetic computations tractable as the sizes of genomic data sets continue to grow.

Acknowledgments

We are grateful to Monty Slatkin and Michael DeGiorgio for helpful comments and discussions. This work was supported by NSF grant DBI-1146722, NIH grant HG005855, and by the Burroughs Wellcome Fund.

Appendix A. Proof of Theorem 2.1

Proof. Let $([r, s], \mathcal{L}, \lambda)$ denote the measure space defined on the interval $[r, s]$ with the Lebesgue σ-algebra on $[r, s]$ and Lebesgue measure λ. Let $\mathcal{A}_{[r,s]}$ denote the space of sample paths $n_{[r,s]}$ of the stochastic process n_t over the time interval $[r, s]$, and define the measure space $(\mathcal{A}_{[r,s]}, S, p)$, where S is the σ-algebra generated by the process n_t and p is the probability distribution of sample paths on $\mathcal{A}_{[r,s]}$. We assume that $(\mathcal{A}_{[r,s]}, S, p)$ is complete, or if not, we assume that it is equal to its completion, which exists by the Completion Theorem (Rudin, 1975, p. 29). We have

$$E[L_{[r,s]}] = E\left[\int_{z=r}^{s} n_z dz\right] = \int_{\mathcal{A}_{[r,s]}} \int_{z=r}^{s} n_z p(n_{[r,s]}) dz \, d\mathcal{A}_{[r,s]}. \quad \text{(A.1)}$$

Tonelli's theorem (DiBenedetto, 2002, Theorem 14.2, p. 148) states that the integrals on the right-hand side of Eq. (A.1) can be exchanged if $([r, s], \mathcal{L}, \lambda)$ and $(\mathcal{A}_{[r,s]}, S, p)$ are complete σ-finite measure spaces and if $n_z p(n_{[r,s]})$ is a nonnegative measurable function on $[r, s] \times \mathcal{A}_{[r,s]}$. The function n_z is a positive step function on $[r, s]$ and it is therefore measurable because a measurable function can be defined as a limit of step functions (Atkinson and Han, 2009, p. 17). The density function $p(n_{[r,s]})$ is also positive and measurable because probability density functions are positive and measurable by definition (Tao, 2011, p. 193). Therefore, the product $n_z p(n_{[r,s]})$ is positive and it is measurable because the product of measurable functions is measurable (Franks, 2009, Page 48, Exercise 3.1.11). The space $([r, s], \mathcal{L}, \lambda)$ is complete because the Lebesgue σ-algebra combined with the Lebesgue measure on a subset of the real numbers forms a complete measure space (Mas-Colell, 1989, p. 23), and $(\mathcal{A}_{[r,s]}, S, p)$ is complete by assumption. Because $\lambda([r, s]) = s - r < \infty$ and $p(\mathcal{A}_{[r,s]}) = 1 < \infty$, the measure spaces $([r, s], \mathcal{L}, \lambda)$ and $(\mathcal{A}_{[r,s]}, S, p)$ are both sets of finite measure, and are therefore σ-finite by definition (DiBenedetto, 2002, p.71). Therefore, it follows that the integrals in Eq. (A.1) can be exchanged by Tonelli's Theorem, yielding

$$\begin{aligned} E[L_{[r,s]}] &= \int_{\mathcal{A}_{[r,s]}} \int_{z=r}^{s} n_z p(n_{[r,s]}) dz \, d\mathcal{A}_{[r,s]} \\ &= \int_{z=r}^{s} \int_{\mathcal{A}_{[r,s]}} n_z p(n_{[r,s]}) d\mathcal{A}_{[r,s]} \, dz \\ &= \int_{z=r}^{s} E[n_z] dz, \end{aligned} \quad \text{(A.2)}$$

which completes the proof. □

Appendix B. A lemma for proving Theorem 3.1

In this section we present a lemma that is necessary for proving Theorem 3.1. The lemma states that the number of lineages n_t

that are ancestral to a set of n_0 sampled lineages approaches its expected value $E[n_t]$ as $t \to 0$ and as $t \to \infty$. Specifically, we show that the random variable $n_t - E[n_t]$ converges in probability to 0 as $t \to 0$ and as $t \to \infty$. We first show that $\mathrm{Var}(n_t) \to 0$ as $t \to 0$ and as $t \to \infty$ for fixed n_0 in a population of arbitrary size $N(t)$.

Lemma B.1. *Consider a panmictic population of variable size $N(t)$ such that $\lim_{t\to 0} \int_{z=0}^{t} \frac{1}{N(z)} dz = 0$ and $\lim_{t\to\infty} \int_{z=0}^{t} \frac{1}{N(z)} dz = \infty$. For a fixed number, n_0, of lineages sampled at time $t = 0$ from this population, $Var(n_t) \to 0$ as $t \to 0$ and as $t \to \infty$.*

Proof. Tavaré (1984, p. 131) showed that the moments of n_t in a panmictic population of constant effective size N can be obtained using the function

$$E[(n_t)_{[k]}] = \sum_{i=k}^{n_0} (2i-1) \binom{i-1}{k-1} \frac{i_{(k-1)}(n_0)_{[i]}}{(n_0)_{(i)}} e^{-i(i-1)t/2}, \tag{B.1}$$

where $E[(n_t)_{[k]}|n_0]$ is the kth factorial moment of n_t, $n_{[i]} = n!/(n-i)!$ and $n_{(i)} = (n-1+i)!/(n-1)!$, and where time t is in coalescent units of N generations.

Chen and Chen (2013) noted that this formula can be extended to the case of a population of variable size $N(t)$ using a result from Griffiths and Tavaré (1994). Specifically, Griffiths and Tavaré showed that in a population of variable size $N(t)$, n_t has the same distribution as the number $n_{\tau(t)}$ of ancestral lineages at time $\tau(t) = \int_{z=0}^{t} \frac{1}{N(z)} dz$ in a population of constant size one. Thus, in a population of variable size $N(t)$, Eq. (B.1) becomes

$$E[(n_t)_{[k]}] = \sum_{i=k}^{n_0} (2i-1) \binom{i-1}{k-1} \frac{i_{(k-1)}(n_0)_{[i]}}{(n_0)_{(i)}} e^{-i(i-1)\tau(t)/2}, \tag{B.2}$$

where $\tau(t) = \int_{z=0}^{t} \frac{1}{N(z)} dz$, and where t is in units of generations.

Using the definitions $(n_t)_{[2]} = n_t^2 - n_t$ and $(n_t)_{[1]} = n_t$, we can write

$$\begin{aligned} \mathrm{Var}(n_t) &= E[n_t^2] - E[n_t]^2 \\ &= E[(n_t)_{[2]}] + E[(n_t)_{[1]}] - E[(n_t)_{[1]}]^2, \end{aligned} \tag{B.3}$$

where, from Eq. (B.2), we have

$$E[(n_t)_{[2]}] = \sum_{i=2}^{n_0} (2i-1)(i-1) \frac{i(n_0)_{[i]}}{(n_0)_{(i)}} e^{-\binom{i}{2}\tau(t)} \tag{B.4}$$

and

$$E[(n_t)_{[1]}] = \sum_{i=1}^{n_0} (2i-1) \frac{(n_0)_{[i]}}{(n_0)_{(i)}} e^{-\binom{i}{2}\tau(t)}. \tag{B.5}$$

By assumption, we have $\tau(t) \to \infty$ as $t \to \infty$. Since $e^{-\binom{i}{2}\tau(t)} \to 0$ as $\tau(t) \to \infty$ for $i \geq 2$, it follows from Eq. (B.4) that $E[(n_t)_{[2]}] \to 0$ as $t \to \infty$. Similarly, since $n_{[1]} = n_{(1)}$, Eq. (B.5) yields

$$\begin{aligned} E[(n_t)_{[1]}] &= 1 + \sum_{i=2}^{n_0} (2i-1) \frac{(n_0)_{[i]}}{(n_0)_{(i)}} e^{-\binom{i}{2}\tau(t)} \\ &= 1 + \mathcal{O}(e^{-\tau(t)}), \end{aligned} \tag{B.6}$$

from which it follows that $E[(n_t)_{[1]}|n_0] \to 1$ as $t \to \infty$. Thus, $\mathrm{Var}(n_t) \to 0$ as $t \to \infty$ by plugging the limiting values of Eqs. (B.4) and (B.5) into the right-hand side of Eq. (B.3).

To obtain the limiting behavior of $\mathrm{Var}(n_t)$ as $t \to 0$, we can use the fact that $e^{-\binom{i}{2}\tau(t)} = 1 - \binom{i}{2}\tau(t) + \mathcal{O}(\tau(t)^2)$. Thus, from Eq. (B.4), we have

$$\begin{aligned} &E[(n_t)_{[2]}] \\ &= \sum_{i=2}^{n_0} (2i-1)(i-1) \frac{i(n_0)_{[i]}}{(n_0)_{(i)}} \left[1 - \binom{i}{2}\tau(t) + \mathcal{O}(\tau(t)^2)\right] \\ &= \sum_{i=2}^{n_0} (2i-1)(i-1) \frac{i(n_0)_{[i]}}{(n_0)_{(i)}} \\ &\quad - \tau(t) \sum_{i=2}^{n_0} (2i-1)(i-1) \frac{i(n_0)_{[i]}}{(n_0)_{(i)}} \binom{i}{2} + \mathcal{O}(\tau(t)^2) \\ &= {n_0}^2 - n_0 - \tau(t) \sum_{i=2}^{n_0} (2i-1) i(i-1) \\ &\quad \times \frac{(n_0)_{[i]}}{(n_0)_{(i)}} \binom{i}{2} + \mathcal{O}(\tau(t)^2), \end{aligned} \tag{B.7}$$

where the three terms in the second equality correspond to the three terms in brackets in the first equality. The first term, $n_0^2 - n_0$, in the third equality is obtained by noting that the first term in the second equality is equal to $E[(n_0)_{[2]}] = n_0^2 - n_0$ (Eq. (B.4)).

Similarly, from Eq. (B.5) we have

$$\begin{aligned} E[(n_t)_{[1]}] &= \sum_{i=1}^{n_0} (2i-1) \frac{(n_0)_{[i]}}{(n_0)_{(i)}} \left[1 - \binom{i}{2}\tau(t) + \mathcal{O}(\tau(t)^2)\right] \\ &= n_0 - \tau(t) \sum_{i=1}^{n_0} (2i-1) \frac{(n_0)_{[i]}}{(n_0)_{(i)}} \binom{i}{2} + \mathcal{O}(\tau(t)^2) \\ &= n_0 - \tau(t) \binom{n_0}{2} + \mathcal{O}(\tau(t)^2), \end{aligned} \tag{B.8}$$

where the third equality is obtained by noting that the second term in the second equality is equal to half the expression for $E[(n_t)_{[2]}]$ evaluated at time $t = 0$; it is therefore equal to $\binom{n_0}{2}$. Squaring Eq. (B.8) gives

$$E[(n_t)_{[1]}]^2 = n_0^2 - 2n_0\tau(t) \binom{n_0}{2} + \mathcal{O}(\tau(t)^2). \tag{B.9}$$

Thus, by plugging Eqs. (B.7)–(B.9) into Eq. (B.3), we obtain

$$\begin{aligned} \mathrm{Var}(n_t) &= n_0^2 - n_0 + \mathcal{O}(\tau(t)) + n_0 + \mathcal{O}(\tau(t)) - n_0^2 + \mathcal{O}(\tau(t)) \\ &= \mathcal{O}(\tau(t)). \end{aligned} \tag{B.10}$$

Here, we have used the fact that $\tau(t)^2 = \mathcal{O}(\tau(t))$. The right-hand side of Eq. (B.10) follows from the linearity of order notation (Miller, 2006, p. 21). Thus, it follows from our assumption that $N(t)$ varies in such a way that $\tau(t) \to 0$ as $t \to 0$ that $\mathrm{Var}(n_t) \to 0$ as $t \to 0$ for fixed values of n_0. □

We now show that $n_t - E[n_t]$ converges in probability to 0 as $t \to 0$ and as $t \to \infty$.

Lemma B.2. *Consider a panmictic population of variable size $N(t)$ at time t, such that $\lim_{t\to 0} \int_{z=0}^{t} \frac{1}{N(z)} dz = 0$ and $\lim_{t\to\infty} \int_{z=0}^{t} \frac{1}{N(z)} dz = \infty$. Suppose that n_0 lineages are sampled from this population and consider the number of ancestral lineages n_t at time t in the past. Under the coalescent model, the random variable $n_t - E[n_t]$ converges in probability to 0 as $t \to 0$ and as $t \to \infty$.*

Proof. The quantity n_t is bounded above by n_0 and below by unity. Thus, n_t has finite mean and variance and therefore satisfies Chebyshev's inequality (Ross, 2007, p. 77). In particular, for any $\epsilon > 0$, direct application of Chebyshev's inequality gives

$$\Pr(|n_t - E[n_t]| > \epsilon) \leq \frac{\mathrm{Var}(n_t)}{\epsilon^2}. \tag{B.11}$$

In Lemma B.1 we showed that for fixed n_0, $\mathrm{Var}(n_t) \to 0$ as $t \to 0$ and as $t \to \infty$. By the sandwich theorem applied to Eq. (B.11), it follows that $\Pr(|n_t - E[n_t]| > \epsilon) \to 0$ as $t \to 0$ and as $t \to \infty$. Thus, by the definition of convergence in probability (Casella and Berger, 2002, p. 232), $n_t - E[n_t]$ converges in probability to 0. □

Appendix C. Proof of Theorem 3.1

Here, we prove that the deterministic approximation (Eq. (4)) is accurate as $t \to 0$ and as $t \to \infty$ for fixed $\mathbf{n}_0$.

Proof. To prove Theorem 3.1, we can expand $f(x|\mathbf{n}_t)$ around the point $E[\mathbf{n}_t]$. The first term in this expansion is simply our approximation $f(x|E[\mathbf{n}_t])$, and we can show that the higher-order terms in the expansion converge to zero as $t \to 0$ and as $t \to \infty$.

By the second-order mean value theorem (Hendrix and Tóth, 2010, p. 41), we have

$$f(x|\mathbf{n}_t) = f(x|E[\mathbf{n}_t]) + \nabla_{\mathbf{n}_t} f(x|E[\mathbf{n}_t])(\mathbf{n}_t - E[\mathbf{n}_t]) + (\mathbf{n}_t - E[\mathbf{n}_t])^T \frac{1}{2} H_{\mathbf{n}_t}[f(x|\mathbf{c}_t)](\mathbf{n}_t - E[\mathbf{n}_t]), \tag{C.1}$$

where $H_{\mathbf{n}_t}[f(x|\mathbf{c}_t)]$ is the Hessian of $f(x|\mathbf{n}_t)$ with respect to $\mathbf{n}_t$ evaluated at a point $\mathbf{c}_t$ given by $\mathbf{c}_t = E[\mathbf{n}_t] + q(\mathbf{n}_t - E[\mathbf{n}_t])$ for some $q \in [0, 1]$. Taking the expectation of both sides with respect to $\mathbf{n}_t$ and noting that $f(x) = \sum_{\mathbf{n}_t} f(x|\mathbf{n}_t) \Pr(\mathbf{n}_t) = E[f(x|\mathbf{n}_t)]$, we obtain

$$\begin{aligned} f(x) &= E[f(x|\mathbf{n}_t)] \\ &= f(x|E[\mathbf{n}_t]) + \frac{1}{2} E\left[(\mathbf{n}_t - E[\mathbf{n}_t])^T H_{\mathbf{n}_t}[f(x|\mathbf{c}_t)] \right. \\ &\quad \left. \times (\mathbf{n}_t - E[\mathbf{n}_t])\right], \end{aligned} \tag{C.2}$$

where the expectation of the second term in Eq. (C.1) is equal to zero because $E[\mathbf{n}_t - E[\mathbf{n}_t]] = 0$. Rearranging Eq. (C.2) and taking absolute values gives

$$\begin{aligned} &|f(x) - f(x|E[\mathbf{n}_t])| \\ &= \left| E\left[\frac{1}{2}(\mathbf{n}_t - E[\mathbf{n}_t])^T H_{\mathbf{n}_t}[f(x|\mathbf{c}_t)](\mathbf{n}_t - E[\mathbf{n}_t])\right]\right| \\ &= \left| E\left[\frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{k}(n_{i,t} - E[n_{i,t}])(n_{j,t} - E[n_{j,t}]) \right.\right. \\ &\quad \left.\left. \times \frac{\partial^2}{\partial n_{i,t}\partial n_{j,t}} f(x|\mathbf{c}_t)\right]\right|. \end{aligned} \tag{C.3}$$

To prove that $f(x|E[\mathbf{n}_t])$ converges uniformly to $f(x)$ on D as $t \to 0$ and as $t \to \infty$, we can bound the right-hand side of Eq. (C.3) and show that this bounded quantity goes to zero as $t \to 0$ and as $t \to \infty$ for all $x \in D$. From Eq. (C.3), we have

$$\begin{aligned} &|f(x) - f(x|E[\mathbf{n}_t])| \\ &\le \frac{1}{2} E\left[\sum_{i=1}^{k}\sum_{j=1}^{k}\left|(n_{i,t} - E[n_{i,t}])(n_{j,t} - E[n_{j,t}])\right| \right. \\ &\quad \left. \times \left|\frac{\partial^2}{\partial n_{i,t}\partial n_{j,t}} f(x|\mathbf{c}_t)\right|\right] \\ &\le M \sum_{i=1}^{k}\sum_{j=1}^{k} E\left[\left|(n_{i,t} - E[n_{i,t}])(n_{j,t} - E[n_{j,t}])\right|\right]. \end{aligned} \tag{C.4}$$

Here, $M = \max_{i,j\in\{1,\ldots,k\}} \sup_{c\in\mathcal{N}} \frac{1}{2}\left|\frac{\partial^2}{\partial n_{i,t}\partial n_{j,t}} f(x|c)\right|$ exists on $D \times \mathcal{N}$ because we have assumed that the second-order partial derivatives $\frac{\partial^2}{\partial n_{i,t}\partial n_{j,t}} f(x|\mathbf{n}_t)$ are bounded. Considering the summand on the right-hand side of Eq. (C.4), we have

$$\begin{aligned} &E[|(n_{i,t} - E[n_{i,t}])\,(n_{j,t} - E[n_{j,t}])|] \\ &\le E[|n_{i,t} - E[n_{i,t}]||n_{j,t} - E[n_{j,t}]|] \\ &\le n_{i,0} E[|n_{j,t} - E[n_{j,t}]|], \end{aligned} \tag{C.5}$$

because $|n_{i,t} - E[n_{i,t}]| \le n_{i,0}$. Now, to show that the term on the right-hand side in Eq. (C.5) converges to 0 as $t \to 0$ and as $t \to \infty$, we can use a convergence theorem from Van der Vaart (2000, Theorem 2.20). This theorem states that if a sequence W_n of random variables converges in probability to W in the limit as $n \to \infty$, then $E[W_n] \to E[W]$ as $n \to \infty$, whenever W_n is asymptotically uniformly integrable. Thus, in Eq. (C.5), $E[|n_{j,t} - E[n_{j,t}]|] \to E[0] = 0$ if $|n_{i,t} - E[n_{i,t}]|$ is asymptotically uniformly integrable.

A sequence of random variables W_n is asymptotically uniformly integrable (Van der Vaart, 2000, p. 17) if

$$\lim_{M\to\infty} \limsup_{n\to\infty} E[|W_n| \mathbf{1}_{\{|W_n|>M\}}] = 0, \tag{C.6}$$

where $\mathbf{1}_{\{|W_n|>M\}}$ is the indicator random variable with $\mathbf{1}_{\{|W_n|>M\}} = 1$ if $|W_n| > M$ and $\mathbf{1}_{\{|W_n|>M\}} = 0$, otherwise. From this definition, it can be seen that $|n_{j,t} - E[n_{j,t}]|$ is asymptotically uniformly integrable because $E[|n_{j,t} - E[n_{j,t}]|\mathbf{1}_{|n_{j,t}-E[n_{j,t}]|>M}] = 0$ whenever $M > \sup|n_{j,t} - E[n_{j,t}]| = n_{j,0}$. Therefore, the right-hand side of Eq. (C.5) converges to zero as $t \to 0$ and as $t \to \infty$ for all $x \in D$ and for fixed $\mathbf{n}_0 \in \mathcal{N}$. By the sandwich theorem, it follows that $E[|(n_{i,t} - E[n_{i,t}])(n_{j,t} - E[n_{j,t}])|] \to 0$ as $t \to 0$ and as $t \to \infty$. From a second application of the sandwich theorem, it follows that the left-hand-side of Eq. (C.4), $|f(x) - f(x|E[\mathbf{n}_t])|$, converges uniformly to 0 for all x in D as $t \to 0$ and as $t \to \infty$. □

Appendix D. Approximate error in the deterministic approximation

Eq. (C.3) in the proof of Theorem 3.1 allows us to obtain an estimate of the error $|f(x) - f(x|E[\mathbf{n}_t])|$ in the deterministic approximation $f(x) \approx f(x|E[\mathbf{n}_t])$. From Eq. (C.3), we have

$$|f(x) - f(x|E[\mathbf{n}_t])| = \left|\frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{k} E\left[(n_{i,t} - E[n_{i,t}])(n_{j,t} - E[n_{j,t}]) \times \frac{\partial^2}{\partial n_{i,t}\partial n_{j,t}} f(x|\mathbf{c}_t)\right]\right|. \tag{D.1}$$

Now, we showed in Lemma B.2 that $n_{i,t} - E[n_{i,t}]$ converges in probability to 0 as $t \to 0$ and as $t \to \infty$. It follows that $\mathbb{P}(\|\mathbf{n}_t - E[\mathbf{n}_t]\| > \epsilon) \to 0$ for any $\epsilon > 0$ as $t \to 0$ and as $t \to \infty$. Thus, recalling that $\mathbf{c}_t = E[\mathbf{n}_t] + q(\mathbf{n}_t - E[\mathbf{n}_t])$, we have $\mathbb{P}(\|[\mathbf{c}_t - E[\mathbf{n}_t]]\| > \epsilon) = \mathbb{P}(\|[\mathbf{n}_t - E[\mathbf{n}_t]]\| > \epsilon/q) \to 0$ as $t \to 0$ and as $t \to \infty$. Therefore, as $t \to 0$ and as $t \to \infty$, we can make the approximation $\mathbf{c}_t \approx E[\mathbf{n}_t]$. Using the approximation $\mathbf{c}_t \approx E[\mathbf{n}_t]$ as $t \to 0$ and as $t \to \infty$, and approximating the expectation of a product by the product of the expectations, we obtain

$$\begin{aligned} |f(x) - f(x|E[\mathbf{n}_t])| &\approx \left|\frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{k} E[(n_{i,t} - E[n_{i,t}])(n_{j,t} - E[n_{j,t}])] \right. \\ &\quad \left. \times E\left[\frac{\partial^2}{\partial n_{i,t}\partial n_{j,t}} f(x|E[\mathbf{n}_t])\right]\right| \\ &= \left|\frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{k} \mathrm{Cov}(n_{i,t}, n_{j,t}) \frac{\partial^2}{\partial n_{i,t}\partial n_{j,t}} f(x|E[\mathbf{n}_t])\right|. \end{aligned} \tag{D.2}$$

Appendix E. Details of the derivation of Eq. (25)

To obtain Eq. (25), we multiply both sides of Eq. (24) by φ_ℓ and sum over all $\prod_{i=1}^{k} n_{i,0}$ possible values of φ.

$$\begin{aligned}\frac{dE[n_{\ell t}]}{dt} &= \frac{d}{dt}\sum_{\varphi}\varphi_\ell p_\varphi(t)\\ &= -\sum_{i=1}^{k}\sum_{\varphi}\varphi_\ell\binom{\varphi_i}{2}\frac{1}{N_i(t)}p_\varphi(t)\\ &\quad-\sum_{i=1}^{k}\sum_{\substack{j=1\\ j\neq i}}^{k}\sum_{\varphi}\varphi_\ell\varphi_i m_{ij}p_\varphi(t)\\ &\quad+\sum_{i=1}^{k}\sum_{\substack{j=1\\ j\neq i}}^{k}\sum_{\varphi}\varphi_\ell(\varphi_i+1)m_{ij}p_{\varphi+\mathbf{e}_i-\mathbf{e}_j}(t)\\ &\quad+\sum_{i=1}^{k}\sum_{\varphi}\varphi_\ell\binom{\varphi_i+1}{2}\frac{1}{N_i(t)}p_{\varphi+\mathbf{e}_i}(t).\end{aligned}\tag{E.1}$$

Each term in Eq. (E.1) can be separated into cases: cases in which $\ell \neq i, j$, or $\ell = i$ and $\ell \neq j$, or $\ell = j$ and $\ell \neq i$. The first and last terms on the right-hand side of Eq. (E.1) separate into two terms each (corresponding to the cases $\ell = i$ and $\ell \neq i$), and the middle terms on the right-hand side of Eq. (E.1) separate into three terms each (corresponding to the cases $\ell = i$, $\ell = j$, and $\ell \neq i, j$). Each of these terms can be further simplified by noting that summations over indices φ_h ($h \neq \ell, i$) are summations over the marginal densities and result in factors of one. Thus, we obtain

$$\begin{aligned}\frac{dE[n_{\ell t}]}{dt} &= -\sum_{\substack{i=1\\ i\neq\ell}}^{k}\sum_{\varphi_\ell}\sum_{\varphi_i}\varphi_\ell\binom{\varphi_i}{2}\frac{1}{N_i(t)}p_{\varphi_\ell,\varphi_i}(t)\\ &\quad-\sum_{\varphi_\ell}\varphi_\ell\binom{\varphi_\ell}{2}\frac{1}{N_i(t)}p_{\varphi_\ell}(t)\\ &\quad-\sum_{\substack{i=1\\ i\neq\ell}}^{k}\sum_{\substack{j=1\\ j\neq i,\ell}}^{k}\sum_{\varphi_\ell}\sum_{\varphi_i}\varphi_\ell\varphi_i m_{ij}p_{\varphi_\ell,\varphi_i}(t)\\ &\quad-\sum_{\substack{j=1\\ j\neq\ell}}^{k}\sum_{\varphi_\ell}\varphi_\ell^2 m_{\ell j}p_{\varphi_\ell}(t)\\ &\quad-\sum_{\substack{i=1\\ i\neq\ell}}^{k}\sum_{\varphi_\ell}\sum_{\varphi_i}\varphi_\ell\varphi_i m_{i\ell}p_{\varphi_\ell,\varphi_i}(t)\\ &\quad+\sum_{\substack{i=1\\ i\neq\ell}}^{k}\sum_{\substack{j=1\\ j\neq i,\ell}}^{k}\sum_{\varphi_\ell}\sum_{\varphi_i}\varphi_\ell(\varphi_i+1)m_{ij}p_{\varphi_\ell,\varphi_i+1}(t)\\ &\quad+\sum_{\substack{j=1\\ j\neq\ell}}^{k}\sum_{\varphi_\ell}\varphi_\ell(\varphi_\ell+1)m_{\ell j}p_{\varphi_\ell+1}(t)\\ &\quad+\sum_{\substack{i=1\\ i\neq\ell}}^{k}\sum_{\varphi_\ell}\sum_{\varphi_i}\varphi_\ell(\varphi_i+1)m_{i\ell}p_{\varphi_\ell-1,\varphi_i+1}(t)\\ &\quad+\sum_{\substack{i=1\\ i\neq\ell}}^{k}\sum_{\varphi_\ell}\sum_{\varphi_i}\varphi_\ell\binom{\varphi_i+1}{2}\frac{1}{N_i(t)}p_{\varphi_\ell,\varphi_i+1}(t)\\ &\quad+\sum_{\varphi_\ell}\varphi_\ell\binom{\varphi_\ell+1}{2}\frac{1}{N_\ell(t)}p_{\varphi_\ell+1}(t),\end{aligned}\tag{E.2}$$

where $p_{\varphi_h,\varphi_m}(t)$ is the probability that $n_{h,t}$ and $n_{m,t}$ lineages remain at time t from the sampled sets of alleles h and m, respectively.

Numbering the terms in Eq. (E.2) from 1 to 10, terms 1 and 9 cancel because they differ only by a shifted index ($\varphi_i + 1$ in term 9, compared with φ_i in term 1). Similarly, terms 3 and 6 cancel. In contrast, terms 2 and 10 do not cancel because the index is shifted only in the binomial coefficient in term 10. For the same reason, terms 4 and 7, and terms 5 and 8 do not cancel. Therefore, canceling terms in Eq. (E.2) and reordering them in the order 2, 10, 4, 7, 5, 8, we obtain

$$\begin{aligned}\frac{dE[n_{\ell t}]}{dt} &= -\sum_{\varphi_\ell}\varphi_\ell\binom{\varphi_\ell}{2}\frac{1}{N_\ell(t)}p_{\varphi_\ell}(t)\\ &\quad+\sum_{\varphi_\ell}\varphi_\ell\binom{\varphi_\ell+1}{2}\frac{1}{N_\ell(t)}p_{\varphi_\ell+1}(t)\\ &\quad-\sum_{\substack{j=1\\ j\neq\ell}}^{k}\sum_{\varphi_\ell}\varphi_\ell^2 m_{\ell j}p_{\varphi_\ell}(t)\\ &\quad+\sum_{\substack{j=1\\ j\neq\ell}}^{k}\sum_{\varphi_\ell}\varphi_\ell(\varphi_\ell+1)m_{\ell j}p_{\varphi_\ell+1}(t)\\ &\quad-\sum_{\substack{i=1\\ i\neq\ell}}^{k}\sum_{\varphi_\ell}\sum_{\varphi_i}\varphi_\ell\varphi_i m_{i\ell}p_{\varphi_\ell,\varphi_i}(t)\\ &\quad+\sum_{\substack{i=1\\ i\neq\ell}}^{k}\sum_{\varphi_\ell}\sum_{\varphi_i}\varphi_\ell(\varphi_i+1)m_{i\ell}p_{\varphi_\ell-1,\varphi_i+1}(t).\end{aligned}\tag{E.3}$$

Each pair of consecutive terms in Eq. (E.3) can be simplified by adding and subtracting an additional term to each pair to facilitate the matching of indices as follows:

$$\begin{aligned}\frac{dE[n_{\ell t}]}{dt} &= -\sum_{\varphi_\ell}\varphi_\ell\binom{\varphi_\ell}{2}\frac{1}{N_\ell(t)}p_{\varphi_\ell}(t)+\sum_{\varphi_\ell}(\varphi_\ell+1)\binom{\varphi_\ell+1}{2}\\ &\quad\times\frac{1}{N_\ell(t)}p_{\varphi_\ell+1}(t)-\sum_{\varphi_\ell}\binom{\varphi_\ell+1}{2}\frac{1}{N_\ell(t)}p_{\varphi_\ell+1}(t)\\ &\quad-\sum_{\substack{j=1\\ j\neq\ell}}^{k}\sum_{\varphi_\ell}\varphi_\ell^2 m_{\ell j}p_{\varphi_\ell}(t)+\sum_{\substack{j=1\\ j\neq\ell}}^{k}\sum_{\varphi_\ell}(\varphi_\ell+1)^2 m_{\ell j}p_{\varphi_\ell+1}(t)\\ &\quad-\sum_{\substack{j=1\\ j\neq\ell}}^{k}\sum_{\varphi_\ell}(\varphi_\ell+1)m_{\ell j}p_{\varphi_\ell+1}(t)\\ &\quad-\sum_{\substack{i=1\\ i\neq\ell}}^{k}\sum_{\varphi_\ell}\sum_{\varphi_i}\varphi_\ell\varphi_i m_{i\ell}p_{\varphi_\ell,\varphi_i}(t)\\ &\quad+\sum_{\substack{i=1\\ i\neq\ell}}^{k}\sum_{\varphi_\ell}\sum_{\varphi_i}(\varphi_\ell-1)(\varphi_i+1)m_{i\ell}p_{\varphi_\ell-1,\varphi+1}(t)\\ &\quad+\sum_{\substack{i=1\\ i\neq\ell}}^{k}\sum_{\varphi_\ell}\sum_{\varphi_i}(\varphi_i+1)m_{i\ell}p_{\varphi_\ell-1,\varphi_i+1}(t).\end{aligned}\tag{E.4}$$

Numbering the terms in Eq. (E.4) from 1 to 9, the adjacent terms 1 and 2, 4 and 5, and 7 and 8 cancel because they differ only by a shifted index. Thus, we obtain

$$\begin{aligned}\frac{dE[n_{\ell t}]}{dt} &= -\sum_{\varphi_\ell}\binom{\varphi_\ell+1}{2}\frac{1}{N_\ell(t)}p_{\varphi_\ell+1}(t)\\ &\quad-\sum_{\substack{j=1\\ j\neq\ell}}^{k}\sum_{\varphi_\ell}(\varphi_\ell+1)m_{\ell j}p_{\varphi_\ell+1}(t)\end{aligned}$$

$$
\begin{aligned}
&+\sum_{\substack{i=1\\ i\neq\ell}}^{k}\sum_{\varphi_\ell}\sum_{\varphi_i}(\varphi_i+1)m_{i\ell}p_{\varphi_\ell-1,\varphi_i+1}(t)\\
&=-\frac{1}{N_\ell(t)}E\left[\binom{n_{\ell t}}{2}\right]-\sum_{\substack{j=1\\ j\neq\ell}}^{k}E[n_{\ell t}]m_{\ell j}+\sum_{\substack{i=1\\ i\neq\ell}}^{k}E[n_{it}]m_{i\ell}\\
&=-\frac{1}{2N_\ell(t)}\left[E[n_{\ell t}^2]-E[n_{\ell t}]\right]\\
&\quad+\sum_{\substack{i=1\\ i\neq\ell}}^{k}(E[n_{it}]m_{i\ell}-E[n_{\ell t}]m_{\ell i})\\
&=-\frac{1}{2N_\ell(t)}\left[E[n_{\ell t}^2]-E[n_{\ell t}]^2+E[n_{\ell t}]^2-E[n_{\ell t}]\right]\\
&\quad+\sum_{\substack{i=1\\ i\neq\ell}}^{k}(E[n_{it}]m_{i\ell}-E[n_{\ell t}]m_{\ell i})\\
&=-\frac{\mathrm{Var}(n_{\ell t})}{2N_\ell(t)}-\frac{1}{N_\ell(t)}\binom{E[n_{\ell t}]}{2}\\
&\quad+\sum_{\substack{i=1\\ i\neq\ell}}^{k}(E[n_{it}]m_{i\ell}-E[n_{\ell t}]m_{\ell i}).
\end{aligned}
\tag{E.5}
$$

This completes the derivation of Eq. (25) from Eq. (24).

Appendix F. Simulation procedure

The accuracy of the approximate expressions in Eqs. (31) and (36) was evaluated by comparing each approximation with estimates of the exact values obtained using simulations. The simulation procedure that was used to validate each approximation was similar to that described elsewhere (Jewett et al., 2012); however, we provide a brief description of the procedure here.

All simulations were performed under a model in which two populations of sizes $N_1(t)$ and $N_2(t)$, respectively, diverged at time t_D in the past from an ancestral population of size $N_3(t)$. Under this model, if a alleles remain at time t in population i, then the additional time t_a until a coalescent event occurs among these a lineages can be simulated by first sampling the time t_a to coalescence in a population of constant size 1, and then rescaling this time according to the formula $\tau_a(t)=\int_{z=t}^{t_a}1/N_i(z)dz$ (see the discussion of time scaling in Section 4.1). In a population of constant size 1, the time t_a until a alleles coalesce is exponentially distributed with mean $1/\binom{a}{2}$ generations.

In contrast to coalescence times, waiting times between migration events can be sampled without rescaling time. If a lineages remain at time t in population i, then the time until one of these a lineages migrates to the other population j is exponentially distributed with mean $1/(am_{ij})$, where m_{ij} is the backward rate of migration from population i to population j.

The simulation proceeds as follows. Suppose that $n_{1,0}$ and $n_{2,0}$ lineages are initially sampled from populations 1 and 2, respectively. The time until the first event of any kind (coalescence or migration) is sampled by sampling the time t_{1C} until the first coalescence in population 1, the time t_{2C} until the first coalescence occurs in population 2, the time t_{1M} until the first migration from population 1 to population 2, and the time t_{2M} until the first migration event from population 2 to population 1. The minimum of these times, $\min\{t_{1C},t_{1M},t_{2C},t_{2M}\}$, is then identified. If t_{iC} ($i=1$ or 2) is the minimum time, then two lineages from population i are randomly chosen and combined. If t_{iM} ($i=1$ or 2) is the minimum time, then a lineage in population i is randomly chosen and moved to population $j\neq i$. The current time is set to $t=\min\{t_{1C},t_{1M},t_{2C},t_{2M}\}$ and the time until the next event (coalescence or migration) is sampled using the same procedure. This procedure is repeated until the time $t+\min\{t_{1C},t_{1M},t_{2C},t_{2M}\}$ exceeds the divergence time t_D. Once $t+\min\{t_{1C},t_{1M},t_{2C},t_{2M}\}$ exceeds t_D, all remaining lineages are merged into the ancestral population of size $N_3(t)$ and, starting from time t_D, coalescence times are sampled until a single lineage remains.

F.1. Simulating the number of private alleles under migration

To obtain a Monte Carlo estimate of the number of private alleles in a sample of $n_{1,0}$ alleles from population 1, we sampled genealogies using the above procedure. For each sampled genealogy, the total sum of lengths L_1 of branches ancestral only to the sample of $n_{1,0}$ alleles from population 1 was computed. $E[S_1]$ was obtained by multiplying each sampled value of L_1 by $\theta b/4$ and averaging the resulting values across all replicates. For each combination of parameter values we tested, $E[S_1]$ was computed using 10^4 sampled genealogies.

F.2. Simulating the time until the first inter-sample coalescent event

To obtain a Monte Carlo estimate the distribution of the time until the first inter-sample coalescent event occurs between $n_{1,0}$ type-1 lineages and $n_{2,0}$ type-2 lineages sampled from two populations, we sampled genealogies using the above procedure. For each pair of sample sizes $n_{1,0}$ and $n_{2,0}$ that we considered, we simulated 10^5 genealogies. For each genealogy, we recorded the time V of the first coalescent event between a type-1 and a type-2 lineage. We then computed kernel density estimates on the 10^5 sampled values of V using Matlab's *ksdensity* function with default parameters and with the option *'function', 'survivor'*.

References

Ariani, C.V., Pickles, R.S.A., Jordan, W.C., Lobo-Hajdu, G., Rocha, C.F.D., 2013. Mitochondrial DNA and microsatellite loci data supporting a management plan for a critically endangered lizard from Brazil. Conserv. Genet. 14, 943–951.

Atkinson, K.E., Han, W., 2009. Theoretical Numerical Analysis: a Functional Analysis Framework. Springer-Verlag, New York.

Bryant, D., Bouckaert, R., Felsenstein, J., Rosenberg, N.A., RoyChoudhury, A., 2012. Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. Mol. Biol. Evol. 29, 1917–1932.

Casella, G., Berger, R.L., 2002. Statistical Inference, second ed. Duxbury Press, Pacific Grove, CA.

Chen, H., Chen, K., 2013. Asymptotic distributions of coalescence times and ancestral lineage numbers for populations with temporally varying size. Genetics 194, 721–736.

Davison, D., Pritchard, J.K., Coop, G., 2009. An approximate likelihood for genetic data under a model with recombination and population splitting. Theor. Popul. Biol. 75, 331–345.

Degnan, J.H., 2010. Probabilities of gene trees with intraspecific sampling given a species tree. In: Knowles, L.L., Kubatko, L.S. (Eds.), Estimating Species Trees: Practical and Theoretical Aspects. Wiley-Blackwell, Hoboken, NJ, pp. 53–78.

Degnan, J.H., Salter, L.A., 2005. Gene tree distributions under the coalescent process. Evolution 59, 24–37.

DiBenedetto, E., 2002. Real Analysis. Birkhauser, Boston.

Donnelly, P., 1984. The transient behaviour of the Moran model in population genetics. Math. Proc. Cambridge Philos. Soc. 95, 349–358.

Efromovich, S., Kubatko, L., 2008. Coalescent time distributions in trees of arbitrary size. Stat. Appl. Genet. Mol. Biol. 7, Article 2.

Franks, J.M., 2009. A (Terse) Introduction to Lebesgue Integration. American Mathematical Society, Providence, RI.

Frost, S.D.W., Volz, E.M., 2010. Viral phylodynamics and the search for an effective number of infections. Philos. Trans. R. Soc. Lond. B 365, 1879–1890.

Griffiths, R.C., 1980. Lines of descent in the diffusion approximation of neutral Wright–Fisher models. Theor. Popul. Biol. 17, 37–50.

Griffiths, R.C., 1984. Asymptotic line-of-descent distributions. J. Math. Biol. 21, 67–75.

Griffiths, R.C., 2006. Coalescent lineage distributions. Adv. Appl. Probab. 38, 405–429.

Griffiths, R.C., Tavaré, S., 1994. Sampling theory for neutral alleles in a varying environment. Philos. Trans. R. Soc. Lond. B 29, 403–410.

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

Gutenkunst, R.N., Hernandez, R.D., Williamson, S.H., Bustamante, C.D., 2009. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. PLoS Genet. 5, e1000695.
Helmkamp, L.J., Jewett, E.M., Rosenberg, N.A., 2012. Improvements to a class of distance matrix methods for inferring species trees from gene trees. J. Comput. Biol. 19, 632–649.
Hendrix, E.M.T., Tóth, B.G., 2010. Introduction to Nonlinear and Global Optimization. Springer, New York.
Huang, L., Buzbas, E.O., Rosenberg, N.A., 2013. Genotype imputation in a coalescent model with infinitely-many-sites mutation. Theor. Popul. Biol. 87, 62–74.
Hudson, R.R., Coyne, J.A., 2002. Mathematical consequences of the genealogical species concept. Evolution 56, 1557–1565.
Jewett, E.M., Rosenberg, N.A., 2012. iGLASS: an improvement to the GLASS method for estimating species trees from gene trees. J. Comput. Biol. 19, 293–315.
Jewett, E.M., Zawistowski, M., Rosenberg, N.A., Zöllner, S., 2012. A coalescent model for genotype imputation. Genetics 191, 1239–1255.
Kalinowski, S.T., 2004. Counting alleles with rarefaction: private alleles and hierarchical sampling designs. Conserv. Genet. 5, 539–543.
Li, H., Durbin, R., 2011. Inference of human population history from individual whole-genome sequences. Nature 475, 493–496.
Li, N., Stephens, M., 2003. Modeling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data. Genetics 165, 2213–2233.
Liu, L., Yu, L., Pearl, D.K., 2010. Maximum tree: a consistent estimator of the species tree. J. Math. Biol. 60, 95–106.
Marjoram, P., Wall, J.D., 2006. Fast ''coalescent'' simulation. BMC Genet. 7, 16.
Maruvka, Y.E., Shnerb, N.M., Bar-Yam, Y., Wakeley, J., 2011. Recovering population parameters from a single gene genealogy: an unbiased estimator of the growth rate. Mol. Biol. Evol. 28, 1617–1631.
Mas-Colell, A., 1989. The Theory of General Economic Equilibrium: a Differentiable Approach. Cambridge University Press, New York.
McVean, G.A.T., Cardin, N.J., 2005. Approximating the coalescent with recombination. Philos. Trans. R. Soc. Lond. B 360, 1387–1393.
Miller, P.D., 2006. Applied Asymptotic Analysis. American Mathematical Society, Providence, RI.
Mossel, E., Roch, S., 2010. Incomplete lineage sorting: consistent phylogeny estimation from multiple loci. IEEE/ACM Trans. Comput. Biol. Bioinform. 7, 166–171.
Nielsen, R., Hubisz, M.J., Hellmann, I., Torgerson, D., Andrés, A.M., Albrechtsen, A., Gutenkunst, R., Adams, M.D., Cargill, M., Boyko, A., Indap, A., Bustamante, C.D., Clark, A.G., 2009. Darwinian and demographic forces affecting human protein coding genes. Genome Res. 19, 838–849.
Paul, J.S., Song, Y.S., 2010. A principled approach to deriving approximate conditional sampling distributions in population genetics models with recombination. Genetics 186, 321–338.
Rauch, E.M., Bar-Yam, Y., 2005. Estimating the total genetic diversity of a spatial field population from a sample and implications of its dependence on habitat area. Proc. Natl. Acad. Sci. USA 102, 9826–9829.
Reppell, M., Boehnke, M., Zöllner, S., 2012. FTEC: a coalescent simulator for modeling faster than exponential growth. Bioinformatics 28, 1282–1283.
Rosenberg, N.A., 2002. The probability of topological concordance of gene trees and species trees. Theor. Popul. Biol. 61, 225–247.
Rosenberg, N.A., 2003. The shapes of neutral gene genealogies in two species: probabilities of monophyly, paraphyly, and polyphyly in a coalescent model. Evolution 57, 1465–1477.
Rosenberg, N.A., Feldman, M.W., 2002. The relationship between coalescence times and population divergence times. In: Slatkin, M., Veuille, D. (Eds.), Modern Developments in Theoretical Population Genetics. Oxford University Press, Oxford, pp. 130–164.
Ross, S., 2007. Introduction to Probability Models, ninth ed. Academic Press, New York.
RoyChoudhury, A., 2011. Composite likelihood-based inferences on genetic data from dependent loci. J. Math. Biol. 62, 65–80.
Rudin, W., 1975. Real and Complex Analysis. McGraw Hill, New York.
Sheehan, S., Harris, K., Song, Y.S., 2013. Estimating variable effective population sizes from multiple genomes: a sequentially Markov conditional sampling distribution approach. Genetics 194, 647–662.
Slatkin, M., 2000. Allele age and a test for selection on rare alleles. Philos. Trans. R. Soc. Lond. B 355, 1663–1668.
Slatkin, M., Rannala, B., 1997. Estimating the age of alleles by use of intraallelic variability. Am. J. Hum. Genet. 60, 447–458.
Szpiech, Z.A., Jakobsson, M., Rosenberg, N.A., 2008. ADZE: a rarefaction approach for counting alleles private to combinations of populations. Bioinformatics 24, 2498–2504.
Takahata, N., 1989. Gene genealogy in three related populations: consistency probability between gene and population trees. Genetics 122, 957–966.
Takahata, N., Nei, M., 1985. Gene genealogy and variance of interpopulational nucleotide differences. Genetics 110, 325–344.
Takahata, N., Slatkin, M., 1990. Genealogy of neutral genes in two partially isolated populations. Theor. Popul. Biol. 38, 331–350.
Tao, T., 2011. An Introduction to Measure Theory. Amerian Mathematical Society, Providence, RI.
Tavaré, S., 1984. Line-of-descent and genealogical processes, and their applications in population genetics models. Theor. Popul. Biol. 26, 119–164.
Tishkoff, S.A., Kidd, K.K., 2004. Implications of biogeography of human populations for 'race' and medicine. Nat. Genet. 36, S21–S27.
Van der Vaart, A.W., 2000. Asymptotic Statistics. Cambridge University Press.
Volz, E.M., Kosakovsky Pond, S.L., Ward, M.J., Leigh Brown, A.J., Frost, S.D.W., 2009. Phylodynamics of infectious disease epidemics. Genetics 183, 1421–1430.
Wakeley, J., Hey, J., 1997. Estimating ancestral population parameters. Genetics 145, 847–855.
Watterson, G.A., 1975. On the number of segregating sites in genetical models without recombination. Theor. Popul. Biol. 7, 256–276.
Wilson, A.S., Marra, P.P., Fleischer, R.C., 2012. Temporal patterns of genetic diversity in Kirtlands warblers *(Dendroica kirtlandii)*, the rarest songbird in North America. BMC Ecol. 12, 8.
Wu, Y., 2012. Coalescent-based species tree inference from gene tree topologies under incomplete lineage sorting by maximum likelihood. Evolution 66, 763–775.