Computing the joint distribution of the total tree length across loci in populations with variable size

Alexey Miroshnikov[1,3] and Matthias Steinrücken[2,3]

[1]University of California, Los Angeles, Department of Mathematics
[2]University of Chicago, Department of Ecology and Evolution
[3]University of Massachusetts Amherst, Department of Biostatistics and Epidemiology

Abstract

In recent years, a number of methods have been developed to infer complex demographic histories, especially historical population size changes, from genomic sequence data. Coalescent Hidden Markov Models have proven to be particularly useful for this type of inference. Due to the Markovian structure of these models, an essential building block is the joint distribution of local genealogical trees, or statistics of these genealogies, at two neighboring loci in populations of variable size. Here, we present a novel method to compute the marginal and the joint distribution of the total length of the genealogical trees at two loci separated by at most one recombination event for samples of arbitrary size. To our knowledge, no method to compute these distributions has been presented in the literature to date. We show that they can be obtained from the solution of certain hyperbolic systems of partial differential equations. We present a numerical algorithm, based on the method of characteristics, that can be used to efficiently and accurately solve these systems and compute the marginal and the joint distributions. We demonstrate its utility to study properties of the joint distribution. Our flexible method can be straightforwardly extended to handle an arbitrary fixed number of recombination events, to include the distributions of other statistics of the genealogies as well, and can also be applied in structured populations.

Keywords: coalescent theory, variable population size, hyperbolic systems of PDEs

AMS subject classification: 92D10, 60J27, 60J28, 35L40

1 Introduction

Unravelling the complex demographic histories of humans or other species and understanding their effects on contemporary genetic variation is a central goal of population genetics. In addition to advancing our knowledge of the evolutionary processes that shape genomic variation, demographic inference is also an important step towards understanding disease related genetic variation. Recent rapid population growth, for example, severely affects the distribution of rare genetic variants (Keinan and Clark, 2012), which have been linked to complex genetic diseases. Moreover, ancient and contemporary population structure can lead to the accumulation of private genetic variation in certain sub-populations.

Methods to study genetic variation, or perform inference, in populations with varying size or more complex demographic histories have been developed based on the Wright-Fisher diffusion, describing the evolution of population allele frequencies forward in time (Griffiths, 2003; Živković et al., 2015; Gutenkunst et al., 2009; Excoffier et al., 2013), or the Coalescent process, a model for the genealogical relationship in a sample of individuals (Griffiths and Tavaré, 1994; Griffiths and Marjoram, 1996; Griffiths and Tavaré, 1998; Živković and Wiehe, 2008; Bhaskar et al., 2015; Kamm et al., 2017). A powerful representation of genetic

variation data that has been used in this context is the Site-Frequency-Spectrum. In this representation, however, any linkage information present in the genetic data is ignored. With the increasing availability of full-genomic sequence data, linkage information is more readily available, and approaches based on Coalescent Hidden Markov Models (HMM) that use this linkage information have proven to be particularly successful for demographic inference and other population genetic applications.

In a population-sample of genomic sequences, the genealogical relationships vary along the genome, due to intra-chromosomal recombination. The Coalescent-HMMs approximate the intricate correlation structure between these local genealogical trees by a Markov chain, the Sequentially Markovian Coalescent (Wiuf and Hein, 1999; McVean and Cardin, 2005). Due to the Markovian structure of the SMC-approximation, an essential building block is thus the transition or joint distribution of these local genealogies at two neighboring loci. In a sample of size two, the local genealogies are simple trees with two leaves, that is, one-dimensional objects at each locus. The transitions can be readily computed, and Li and Durbin (2011) employed this framework to developed a powerful approach to infer population size history. Moreover, Dutheil et al. (2009) used Coalescent-HMMs to explore the divergence patterns between humans and great apes, using up to 4 genomic sequences, one for each species. However, due to the increase in complexity of the local genealogies with increasing sample size, these approaches cannot be generalized efficiently to larger sample sizes.

For large sample sizes, approaches that use Monte-Carlo Markov Chain techniques (Rasmussen et al., 2014), suitable composite likelihood frameworks (Sheehan et al., 2013; Steinrücken et al., 2015), or representations of the local genealogical trees by lower-dimensional summaries (Schiffels and Durbin, 2014; Terhorst et al., 2017) have been developed. In the latter, the choice on how to represent the local genealogical trees affects the performance of the inference procedure. Li and Durbin (2011) observed that using the coalescence time between two lineages lacks information in the more recent past, whereas using the first coalescence time in a large sample is less accurate for ancient times (Schiffels and Durbin, 2014). A promising low-dimensional representation is the total branch length of the genealogical tree at each locus. In expectation, this quantity grows without bound as the sample size increases, thus retaining not only information about ancient events, but also about the more recent dynamics. However, to implement a Coalescent-HMM inference framework using the tree length, it is crucial to efficiently compute the joint distribution of the total tree length at two neighboring loci.

Thus, in this paper, we present a novel efficient and accurate method to numerically compute the joint distribution of the total branch length of the genealogical trees at two neighboring loci for a sample of arbitrary size n in populations of varying size, as well as the single-locus marginal distribution. To our knowledge, no method to compute these distributions has been presented in the literature to date that can be applied to arbitrary sample sizes. Moreover, even computing the marginal distribution of the total tree length at a single locus has only received limited attention (Pfaffelhuber et al., 2011; Wiuf and Hein, 1999). We present analytical details and numerical results for the case of at most one recombination event separating the two loci, but our methodology can be readily extended to handle an arbitrary, but fixed, maximal number of recombination events, by suitably augmenting the underlying process.

The inter-coalescent times $T_k^{(n)}$, that is the time period during which k lineages persist in the genealogical tree for a sample of size n can be used to compute the total branch length at a single locus as

$$\mathcal{L} = \sum_{k=2}^{n} k T_k^{(n)}, \tag{1.1}$$

since in the period $T_k^{(n)}$, k lineages contribute towards the total length. In the case of a panmictic population of constant size, formulas for the first two moments of the total tree length can be readily obtained using standard arguments for sums of the independently exponentially distributed random variables $T_k^{(n)}$. Furthermore, $\mathcal{L}$ is distributed like the maximum of $k-1$ exponential variables with intensity $\frac{1}{2}$ (Wiuf and Hein, 1999, p. 255). However, non-constant population size histories introduce intricate dependencies among the inter-coalescent times, and thus it is not straightforward to generalize this approach. Polanski et al. (2003) introduced a method to compute the expected inter-coalescence times under variable population size. However, the coalescence rates of ancestral lineages in the genealogical process depend on past population

2

sizes, whereas the rate for ancestral recombination is constant along each ancestral lineage. The approach of Polanski et al. (2003) depends on the fact that all rates of the process are rescaled uniformly with the same factor, and thus it cannot be extended to the case when ancestral recombination between two linked loci is taken into account.

Ferretti et al. (2013) used another approach to investigate the correlation between the times to the most recent common ancestor at two neighboring loci. The authors approached the problem using coalescent arguments to quantify the changes recombination induces on the local trees, but it is unclear how to generalize their approach efficiently to the total length of the genealogical trees. Furthermore, Li and Durbin (2011) presented analytic formulas for the joint distribution of the local genealogies for a sample of size two under variable population size, but these cannot readily be extended to an arbitrary sample size n. Eriksson et al. (2009) presented similar analytic formulas for a population of constant size and explored more complex demographic scenarios using simulations. Introducing suitable Markov chains, Hobolth and Jensen (2014) investigated the transitional distribution of the local genealogies for samples of size 4, and discussed approximations for larger sample sizes. These Markov chains are closely related to our methodology, but our focus is on exact computations for large sample sizes.

Although we focus on the total tree length under variable population size in a single panmictic population in this paper, our approach can be extended to compute the transition densities for the coalescence time in a sample of size two (Li and Durbin, 2011), the coalescence time of two distinguished lineages (Terhorst et al., 2017), and the time of the first coalescent event amongst the sampled sequences (Schiffels and Durbin, 2014). Furthermore, our method can be generalized to multiple sub-populations related by a complex demographic history (see discussion in Section 5).

This article is structured as follows. In Section 2, we introduce the requisite notation and the stochastic processes that are involved in computing the marginal and joint distributions. We further introduce a hyperbolic system of partial differential equations (PDEs) in Section 3 that can be solved to compute the distributions of interest. We provide a proof of the main proposition used to derive these equations in Appendix A. In Section 3, we also provide details of our novel numerical algorithm based on the method of characteristics that can be used to efficiently compute the solutions to these PDEs. We demonstrate the accuracy of the method, and study properties of the joint distribution function in Section 4. Finally, we discuss future applications and extensions of this method in Section 5.

2   Background and Notation

In this section, we will introduce the necessary background and notation for the stochastic processes that we employ to compute the marginal and joint distribution of the length of the genealogical trees. We will also provide some details about computing the distribution of these processes, since our main result extends upon the underlying ideas.

2.1   Ancestral Process at a Single Locus

The genealogical relationship of a sample of n haploid individuals in a panmictic population of constant size is commonly modeled using Kingman's coalescent (Kingman, 1982; Wakeley, 2008), and this process and its extensions have found widespread applications. It is a Markov process that describes the dynamics of the ancestral lineages of the sample backwards in time. Here we focus on the ancestral process $A(t)$ (Tavaré and Zeitouni, 2004, Chapter 4.1). This coarser process records only the number of ancestral lineages in the coalescent process at time t before present, which is sufficient to compute the total branch length of the coalescent tree. The initial number of lineages is equal to the sample size n. Furthermore, at time t, each pair of lineages coalesces at rate one, thus if there are $A(t) = k$ lineages at time t, then coalescence of any two lineages happens at rate $\binom{k}{2}$. This dynamics is followed until all lineages coalesced into a single lineage, and this time is denoted by T_{MRCA}, the time to the most recent common ancestor.

Variable population size is commonly modeled by a positive, real-valued function $\lambda(t)$, which provides the coalescent rate for each pair of ancestral lineages at time t in the past (Tavaré and Zeitouni, 2004,

Chapter 4.1). If the size of the population changes at different points in the past, the rate of coalescence at a given time is inversely proportional to the relative population size at that time. Intuitively, for two lineages to coalesce, they have to find a common ancestor. If the population consists of a large number of individuals, this happens at a lower rate, whereas in small populations, the ancestral lineages coalesce more quickly. In the remainder of this paper, we assume that $\lambda(t)$ is continuous. If $\lambda(t)$ is piece-wise continuous, we can obtain the same results by considering each continuous piece separately. For convenience, we further introduce the cumulative coalescent rate at time t as

$$\Lambda(t) = \int_0^t \lambda(s)ds.$$

These considerations yield the following definition.

Definition 2.1 (Ancestral Process with variable population size)**.** *The* ancestral process with variable population size $\{A(t)\}_{t \in \mathbb{R}_+}$ *is a time-inhomogeneous Markov chain on* $\{1, \ldots, n\}$ *with initial state* $A(0) = n$, *and the transition rates at time* t *are given by the infinitesimal generator matrix*

$$Q(t) = \lambda(t)Q,$$

with

$$Q_{k,j} := \begin{cases} -\binom{k}{2}, & \textit{if } j = k, \\ \binom{k}{2}, & \textit{if } j = k - 1, \\ 0, & \textit{otherwise.} \end{cases} \tag{2.1}$$

Remark 2.2. *Note that we do require* $A(0) = n$, *and thus this definition of the ancestral process does depend on the sample size* n. *However, for different sample sizes* n', *the rates of the process are given by equation* (2.1) *as well, only the initial state changes. The dynamics of the process is essentially the same, independent of the sample size, and we therefore do not include the dependence on the sample size explicitly in the notation for the remainder of this article.*

The ancestral process can be used to formally define the time to the most recent common ancestor as

$$T_{\text{MRCA}} := \inf\{t \in \mathbb{R}_+ : A(t) \leq 1\},$$

the time when the number of lineages reaches one. Furthermore, with

$$p_k(t) := \mathbb{P}\{A(t) = k\},$$

for $k \in \{1, \ldots, n\}$, the distribution of the ancestral process can be obtained by solving the Kolmogorov-forward-equation (Stroock, 2008, Chapter 5), a system of ordinary differential equations (ODEs) given by

$$\frac{d}{dt}\big(p_1(t), \ldots, p_n(t)\big) = \big(p_1(t), \ldots, p_n(t)\big)Q(t). \tag{2.2}$$

Equivalently, perhaps more familiar to the reader, this system can be expressed as

$$\frac{d}{dt}p_k(t) = \lambda(t)\binom{k+1}{2}p_{k+1}(t) - \lambda(t)\binom{k}{2}p_k(t), \tag{2.3}$$

for all $k \in \{1, \ldots, n\}$. The latter version is more explicit about the influence of the number of ancestral lineages and the coalescent-speed function on the dynamics of the ODEs. The relevant solution is given by

$$\big(p_1(0), \ldots, p_n(0)\big) = (0, \ldots, 0, 1)$$

and

$$\big(p_1(t), \ldots, p_n(t)\big) = \big(e^{\Lambda(t) \cdot Q}\big)_{n, \cdot}. \tag{2.4}$$

for $t \in \mathbb{R}_+$, where $(\cdot)_{n,\cdot}$ refers to the n-th row of the matrix. In Tavaré and Zeitouni (2004), the authors provide an analytic expression for these probabilities using the spectral decomposition of the rate matrix $Q(t)$. However, the resulting formulas are numerically unstable, so for practical purposes it can be more efficient to solve the system of ODEs numerically using step-wise solution schemes. Furthermore, note that

$$\mathbb{P}\{T_{\mathrm{MRCA}} \leq t^*\} = \left(e^{\Lambda(t^*)\cdot Q}\right)_{n,1} \tag{2.5}$$

holds for $t^* \in \mathbb{R}_+$, thus equation (2.4) can also be used to compute the cumulative distribution function of the time to the most recent common ancestor.

We can employ the ancestral process to compute the total tree length as follows. If at a given time t there are k ancestral lineages or branches in the coalescent tree, each branch extends further into the past. Thus, we can say that the total sum of branch lengths in the coalescent tree grows at a rate of k. Once all lineages have coalesced into a single common ancestral lineage, the most recent common ancestor is reached, and the coalescent tree stops growing. This motivates the following definition.

Definition 2.3. *The* accumulated tree length $L(t) \in \mathbb{R}_+$ *by time* $t \in \mathbb{R}_+$ *is given by*

$$L(t) := \int_0^t \mathbb{1}_{\{A(s)>1\}} A(s) ds. \tag{2.6}$$

With this definition, the *total tree length* or the *total sum of the branch lengths* at a single locus is given by

$$\mathcal{L} := L\big(T_{\mathrm{MRCA}}\big). \tag{2.7}$$

Note that

$$\mathcal{L} = \sum_{k=2}^n k T_k^{(n)}$$

holds, which is equal to equation (1.1). Here $T_k^{(n)}$ is the period of time for which k lineages persist in the ancestral process, the inter-coalescent time. The main goal of this paper is to study the distribution of $\mathcal{L}$ for populations with arbitrary coalescent-rate function $\lambda(t)$ marginally at a single locus and jointly at two loci, which can be computed using a system of hyperbolic PDEs that is closely related to the ODE (2.3). For the two-locus case, we will now introduce the joint ancestral process at two linked loci.

2.2 Ancestral Process with Recombination

The joint genealogy of the ancestral lineages for two loci, locus a and b, separated by a recombination distance ρ is commonly modeled by the coalescent with recombination (Hudson, 1990). The initial state in the coalescent with recombination for a sample of size n is comprised of n lineages, each ancestral to both loci of one sampled haplotype. As in the single-locus coalescent with variable population size, at time t, each pair of lineages can coalesce at rate $\lambda(t)$. In addition, ancestral recombination events happen at rate $\rho/2$ along each active lineage. At a recombination event, the lineage splits into two new lineages, each ancestral to the respective haplotype of the original lineage at only one of the two loci. Note that recombination happens along each lineage at a constant rate and, unlike the coalescent rate, is not affected by the population size, and thus it does not scale with $\lambda(t)$.

Again, we do not focus on the exact genealogical relationships, but only on the number of lineages at time t that are ancestral to a certain locus, given by the *ancestral process with recombination* $A^\rho(t)$. The process A^ρ for a sample of size two under constant population size is described in detail by Simonsen and Churchill (1997). Here we use an extension of this process to samples of arbitrary size n and variable population size. A similar model has also been introduced by Hobolth and Jensen (2014).

Definition 2.4 (Ancestral Process with Recombination). *For a sample of size* $n \in \mathbb{N}$ *and* $t \in \mathbb{R}_+$, *the* ancestral process with recombination *in a population of variable size*

$$A^\rho(t) = \big(K_{ab}(t), K_a(t), K_b(t)\big)$$

5

is a time-inhomogeneous Markov chain with state space

$$\mathcal{S}^\rho := \left\{ s \in \mathbb{N}_0^3 \big| s_1 + \max\{s_2, s_3\} \leq n \right\} \backslash \{(0,0,0), (0,1,0), (0,0,1)\}.$$

The component $K_{ab}(t)$ gives the number of lineages that are ancestral to both loci, $K_a(t)$ is the number ancestral to locus a only, and $K_b(t)$ is the number ancestral to locus b only. The initial state is

$$A^\rho(0) = (n, 0, 0),$$

all n lineages ancestral to both loci. The transition rates are given by the infinitesimal generator matrix

$$\tilde{Q}(t) = \lambda(t) Q^c + Q^\rho,$$

where all off-diagonal entries of Q^c (coalescence) are zero, except

$$Q^c_{(k_{ab},k_a,k_b),(k_{ab}-1,k_a,k_b)} = \binom{k_{ab}}{2},$$

$$Q^c_{(k_{ab},k_a,k_b),(k_{ab},k_a-1,k_b)} = \binom{k_a}{2} + k_{ab}k_a,$$

$$Q^c_{(k_{ab},k_a,k_b),(k_{ab},k_a,k_b-1)} = \binom{k_b}{2} + k_{ab}k_b,$$

$$and$$

$$Q^c_{(k_{ab},k_a,k_b),(k_{ab}+1,k_a-1,k_b-1)} = k_a k_b, \tag{2.8}$$

and all off-diagonal entries of Q^ρ (recombination) are zero, except

$$Q^\rho_{(k_{ab},k_a,k_b),(k_{ab}-1,k_a+1,k_b+1)} = \frac{\rho}{2} k_{ab}.$$

The state $(1,0,0)$ is defined to be the absorbing state, so all rates leaving this state are set to zero. Furthermore, the diagonal entries of both matrices are set to minus the sum of the off-diagonal entries in the corresponding row.

Remark 2.5. *i) Two versions of the coalescent with recombination are commonly used in the literature, one version for the infinitely-many-sites (IMS) model (Hudson, 1990; Griffiths and Marjoram, 1997), and another version for the finitely-many-sites (FMS) model (Paul et al., 2011; Steinrücken et al., 2015). In the IMS version, the chromosome is modeled as the interval $[0, 1]$, and whenever recombination occurs, it occurs at a uniformly chosen point in this interval. As a result, recombination always occurs at a novel site, and two neighboring local genealogies are separated by at most one recombination event. In the FMS version, multiple recombination events can occur between two loci. It can be obtained from the IMS version by considering the local genealogies at two fixed loci along the continuous chromosome that are separated by a certain fixed recombination distance. Our definition of the ancestral process with recombination is in line with the FMS version for two loci.*
ii) The ancestral process with recombination can be defined for an arbitrary number of loci. However, in the remainder of the paper, we will only use the process for two loci.
iii) In the literature, some authors use the 'full' coalescent with recombination and others the 'reduced' coalescent with recombination. The difference between the two is that the 'full' version always keeps track of both ancestral lineages that branch off at a recombination event, whereas in the 'reduced' version, lineages that don't leave any descendant ancestral material in the contemporary sample are not traced. Our definition of the ancestral process with recombination is compatible with the 'reduced' version. Thus, the number of ancestral lineages is bounded, which is not the case in the 'full' version.
iv) Following the ideas of Wiuf and Hein (1999), the correlation structure between all local genealogies along a chromosome can be approximated using the Sequentially Markovian Coalescent (SMC) (McVean and Cardin,

2005), or the modified version SMC' (Marjoram and Wall, 2006). In the SMC, if a lineage has been hit by a recombination event and branches into two, subsequently, the two resulting branches are not allowed to coalesce with each other, whereas such events are permitted under the SMC'. Thus, under the SMC', the rates for coalescence of lineages with no overlapping ancestral material (equation (2.8)) are as given in Definition 2.4, whereas under the SMC, these rates have to be set to zero.

Again, the Kolmogorov-forward-equation can be used to compute the distribution of the ancestral process $A^\rho(t)$ as the solution of

$$\frac{d}{dt}\mathbf{p}(t) = \mathbf{p}(t)\tilde{Q}(t), \tag{2.9}$$

where the row-vector $\mathbf{p}(t)$ is defined by

$$\mathbf{p}(t) := \left(\mathbb{P}\{A^\rho(t) = s\}\right)_{s\in\mathcal{S}^\rho}.$$

Note that the rate matrix $Q(t)$ in the ODE (2.2) for the ancestral process at a single locus is triangular for all t. This simplifies approaches to compute solutions substantially, as the solutions can be obtained sequentially for each state of the corresponding Markov chain. In the ancestral process with recombination for two loci on the other hand, with a positive probability, the underlying Markov chain can transition back to a state it already visited before. Consequently, the rate matrix $\tilde{Q}(t)$ in the ODE (2.9) is not triangular, and it is also not possible to transform it into a triangular matrix by permuting the rows and columns.

Since a triangular rate matrix simplifies analytical and numerical approaches significantly, we introduce an approximation to the full ancestral process with recombination that exhibits this property and compute the distributions of the tree lengths under this approximation. To achieve this, we explicitly account for the number of recombination events that have occurred up to a certain time t. For ease of exposition, we further limit the maximal number of recombination events to one. Since in most organisms the per generation recombination probability is very small between loci that are physically close, this approximation is justified. Furthermore, numerical experiments supporting this approximation are provided in Section 4. Note that this limiting the number of recombination events to one yields effectively a first-order approximation to the full ancestral process.

Definition 2.6 (Ancestral Process with Limited Recombination). *For a sample of size $n \in \mathbb{N}$ and $t \in \mathbb{R}_+$, the* ancestral process with limited recombination

$$\bar{A}^\rho(t) = \left(\bar{K}_{ab}(t), \bar{K}_a(t), \bar{K}_b(t), \bar{R}(t)\right)$$

is a time-inhomogeneous Markov chain with state space

$$\begin{aligned}\bar{\mathcal{S}}^\rho :=&\Big(\{1,\ldots,n\} \times \{(0,0,0)\}\\&\cup \{1,\ldots,n\} \times \{0,1\} \times \{0,1\} \times \{1\}\Big)\\&\setminus\big\{(n,1,1,1),(n,1,0,1),(n,0,1,1)\big\}.\end{aligned} \tag{2.10}$$

The components $\bar{K}_{ab}(t)$, $\bar{K}_a(t)$, $\bar{K}_b(t)$ have the same interpretation as before, and $\bar{R}(t)$ is the number of recombination events that have happened by time t. The first line in equation (2.10) corresponds to the states that can be reached without recombination, and the second line to those that require one recombination event. The initial state is

$$\bar{A}^\rho(0) = (n,0,0,0),$$

and the transition rates are given by the infinitesimal generator matrix

$$\bar{Q}(t) = \lambda(t)\bar{Q}^c + \bar{Q}^\rho, \tag{2.11}$$

where the entries of $\bar{Q}^c$ (coalescence) are given by

$$\bar{Q}^c_{(k_{ab},k_a,k_b,r),(k_{ab},k_a,k_b,r)} = Q^c_{(k_{ab},k_a,k_b),(k_{ab},k_a,k_b)},$$

7

and all off-diagonal entries of Q^ρ (recombination) are zero, except

$$\bar{Q}^\rho_{(k_{ab},k_a,k_b,0),(k_{ab}-1,k_a+1,k_b+1,1)} = \frac{\rho}{2}k_{ab},$$

allowing at most one recombination event. The diagonal entries are set to minus the sum of the off-diagonal entries in the corresponding row. The states $(1,0,0,0)$ and $(1,0,0,1)$ are absorbing states, so all rates leaving these states are set to zero.

For later convenience, define the relation $\prec$ on $\bar{\mathcal{S}}^\rho$ as

$$s \prec s' :\Leftrightarrow \bar{Q}_{s',s}(t) > 0, \tag{2.12}$$

that is, $s \prec s'$ holds if s can be reached from s' in one step. Note that embedded into the ancestral process with recombination (limited or not) is a single-locus ancestral process for locus a and for locus b. Thus, we can define the branch length of the genealogical tree at locus a and b similar to the one-locus case as follows, and study their joint distribution.

Definition 2.7. *For a given time $t \in \mathbb{R}_+$, the* accumulated tree lengths *$L^a(t) \in \mathbb{R}^+$ at locus a and $L^b(t) \in \mathbb{R}_+$ at locus b are given by*

$$L^a(t) := \int_0^t \mathbb{1}_{\{\bar{K}_{ab}(s)+\bar{K}_a(s)>1\}}\big(\bar{K}_{ab}(s) + \bar{K}_a(s)\big)ds,$$

and

$$L^b(t) := \int_0^t \mathbb{1}_{\{\bar{K}_{ab}(s)+\bar{K}_b(s)>1\}}\big(\bar{K}_{ab}(s) + \bar{K}_b(s)\big)ds.$$

Remark 2.8. *This definition of the accumulated tree length can be applied to $\bar{A}^\rho$, as well as A^ρ. We will not distinguish these cases in our notation, since in the remainder of the paper, we will use $\bar{A}^\rho$.*

The *total tree length* at locus a is thus given by

$$\mathcal{L}^a := L^a\big(T^a_{\mathrm{MRCA}}\big), \tag{2.13}$$

and at locus b by

$$\mathcal{L}^b := L^b\big(T^b_{\mathrm{MRCA}}\big). \tag{2.14}$$

Here, T^a_{MRCA} is the time to the most recent common ancestor at locus a

$$T^a_{\mathrm{MRCA}} := \inf\big\{t \in \mathbb{R}_+ : \bar{K}_{ab}(t) + \bar{K}_a(t) \le 1\big\},$$

and thus its distribution is given by

$$\mathbb{P}\{T^a_{\mathrm{MRCA}} \le t^*\} = \mathbb{P}\big\{\bar{K}_{ab}(t^*) + \bar{K}_a(t^*) \le 1\big\}$$

for $t^* \in \mathbb{R}_+$. Similar relations hold for locus b. We will now study the joint distribution of $\mathcal{L}^a$ and $\mathcal{L}^b$, and also the marginal $\mathcal{L}$. Note that these quantities are computed under the ancestral process with limited recombination, but we will demonstrate in Section 4 that they give an accurate approximation to the respective quantities under the true ancestral process.

3 Marginal and Joint Distribution of the Total Tree Length

The main goal of this paper is to present a method to compute the marginal and joint cumulative distribution function (CDF) of the total tree length at two linked loci. Thus, we aim at computing

$$\mathbb{P}\{\mathcal{L} \le x\} \tag{3.1}$$

and

$$\mathbb{P}\{\mathcal{L}^a \leq x, \mathcal{L}^b \leq y\} \tag{3.2}$$

for $x, y \in \mathbb{R}_+$.

Note that equation (2.5) can be used to compute the distribution of the time until the ancestral process reaches the absorbing state, which yields the marginal distribution of the T_{MRCA}. The latter is equal to the sum of the inter-coalescence times, and in a population of constant size, equation (2.5) can also be obtained by convolving the densities of independent exponential variables. The total branch length is a more general linear combination of the inter-coalescence times, but Wiuf and Hein (1999) used a similar convolution approach to derive its marginal density. In a population with variable population size, the inter-coalescence times are not mutually independent. However, Polanski et al. (2003) derived formulas for the density of T_{MRCA} using a uniform rescaling of time by the coalescent-rate function $\lambda(t)$.

The two main difficulties in extending these considerations to the total tree length in a two-locus model with variable population size are as follows: Firstly, in a model that includes recombination, only the coalescence rates scale with $\lambda(t)$, while the recombination rate is constant along each lineage. The approach of Polanski et al. (2003), however, relies on a uniform rescaling of all rates, and therefore it cannot be applied. Secondly, note that, similar to equation (2.6), we can define

$$T(t) := \int_0^t \mathbb{1}_{\{A(s)>1\}} ds = \begin{cases} t, & \text{if } t < T_{\mathrm{MRCA}}, \\ T_{\mathrm{MRCA}}, & \text{otherwise.} \end{cases}$$

With this definition, the quantity t is not only the time elapsed in the ancestral process, but it can also be interpreted as the amount accumulated towards T_{MRCA}. The absorption time of the ancestral process can thus be used in equation (2.5) to compute the distribution of T_{MRCA}. However, when the accumulated tree length $L(t)$ defined in equation (2.6) is considered, the quantity t only gives the elapsed time, and it cannot be used as the amount accumulated towards $\mathcal{L}$. Thus, our approach to compute the distribution of $\mathcal{L}$, and the joint distribution of $\mathcal{L}^a$ and $\mathcal{L}^b$ has to explicitly account for both, the time that has elapsed in the ancestral process, as well as the amount accumulated towards the total tree length.

To this end, with $t \in \mathbb{R}_+$, we introduce the time-dependent cumulative distribution functions

$$F_k(t, x) := \mathbb{P}\{A(t) = k, L(t) \leq x\} \tag{3.3}$$

for $k \in \{1, \dots, n\}$ and

$$F_s(t, x, y) := \mathbb{P}\{\bar{A}^\rho(t) = s, L^a(t) \leq x, L^b(t) \leq y\} \tag{3.4}$$

for $s \in \bar{\mathcal{S}}^\rho$.

We will show that the CDFs (3.1) and (3.2) can be computed from the time-dependent CDFs (3.3) and (3.4). Furthermore, we will present numerical schemes, to efficiently and accurately compute the time-dependent CDFs (3.3) and (3.4).

3.1 Distribution of the Total Tree length at a Single Locus

The following lemma shows that the CDF (3.1) can be computed from the time-dependent CDF (3.3).

Lemma 3.1. *With definition* (3.3)*, the relation*

$$\mathbb{P}\{\mathcal{L} \leq x\} = \mathbb{P}\{A(\bar{t}) = 1, L(\bar{t}) \leq x\} = F_1(\bar{t}, x)$$

holds for $x \in \mathbb{R}_+$ and $\bar{t} \geq x/2$.

Proof. First, observe that

$$2T_{\mathrm{MRCA}} \leq \int_0^{T_{\mathrm{MRCA}}} \mathbb{1}_{\{A(s)>1\}} A(s) \, ds = \mathcal{L},$$

9

since $A(s) \geq 2$ holds for $s < T_{\mathrm{MRCA}}$. Thus, on the event $\{\mathcal{L} \leq x\}$, the relation $T_{\mathrm{MRCA}} \leq x/2 \leq \bar{t}$ holds, which implies $A(\bar{t}) = 1$, and therefore

$$\{\mathcal{L} \leq x\} = \{A(\bar{t}) = 1, \mathcal{L} \leq x\}.$$

On the event $\{A(\bar{t}) = 1\}$, $\bar{t} \geq T_{\mathrm{MRCA}}$ and $L(\bar{t}) = \mathcal{L}$ hold, and thus

$$\{A(\bar{t}) = 1, \mathcal{L} \leq x\} = \{A(\bar{t}) = 1, L(\bar{t}) \leq x\},$$

which proves the statement of the lemma. $\qquad\square$

Lemma 3.1 shows that the CDF of $\mathcal{L}$ can be computed from the time-dependent CDF $F_1(t,x)$. Due to the structure of the underlying Markov chain, it is necessary to compute the time-dependent CDFs for all states in order to compute it for the absorbing state. Thus, in the remainder of this section, we focus on computing the time-dependent CDFs for all $k \in \{1, \ldots, n\}$. Proposition A.14 derived in Appendix A can be applied to show that the time-dependent CDFs solve a certain system of linear hyperbolic PDEs. This yields the following corollary.

Corollary 3.2. *The row-vector*

$$\mathbf{F}(t,x) := \big(F_1(t,x), \ldots, F_n(t,x)\big)$$

can be obtained for all points in $\mathcal{U} = \Big\{(x,t) : 0 < x < nt,\, t > 0\Big\}$ as the strong solution of

$$\partial_t \mathbf{F}(t,x) + \partial_x \mathbf{F}(t,x) V = \mathbf{F}(t,x) Q(t), \tag{3.5}$$

with

$$V = \boldsymbol{diag}(0, 2, 3, \ldots, n),$$

boundary conditions

$$\begin{aligned}
\mathbf{F}(t,x) &= \big(\mathbb{P}\{A(t) = 1\}, \ldots, \mathbb{P}\{A(t) = n-1\}, 0\big), \quad x = nt \\
\mathbf{F}(t,0) &= \big(0, 0, \ldots, 0\big), \quad t > 0,
\end{aligned} \tag{3.6}$$

*and matrix $Q(t)$ as defined in equation (2.1).*

Proof. Define the function

$$v(k) := k \cdot \mathbb{1}_{\{k > 1\}}$$

on the state space $\mathcal{S} = \{1, \ldots, n\}$ of the ancestral process. This function and the generator $Q(t)$ satisfy the requirements of Proposition A.14, and thus, the statement of the corollary follows from Proposition A.14 and Remark A.16. $\qquad\square$

Remark 3.3. *The n-th component of the boundary condition (3.6) is equal to 0 and not $\mathbb{P}\{A(t) = n\}$. This holds for technical reasons that will be detailed in the proof of Proposition A.14.*

Note that the process $\big(A(t), L(t)\big)_{t \in \mathbb{R}_+}$ is a piecewise-deterministic Markov process (see Remark A.17). The right-hand side of equation (3.5) is essentially equal to the right-hand side of equation (2.3), because the only stochastic element in the underlying dynamics is the ancestral process $A(t)$. Given a certain number of lineages $\{A(t) = k\}$, the accumulation towards the total tree length happens deterministically at rate k, and is captured by the term $V \partial_x \mathbf{F}(t,x)$.

To derive a numerical scheme for the efficient and accurate computation of the time-dependent CDF $\mathbf{F}(t,x)$, note that the system of PDEs introduced in Corollary 3.2 can be solved using the method of characteristics (Renardy and Rogers, 2004, Chapter 3). Due to the triangular structure of the matrix $Q(t)$, for a given component with $k \in \{1, \ldots, n\}$, the right-side of equation (3.5) does only depend on F_ℓ
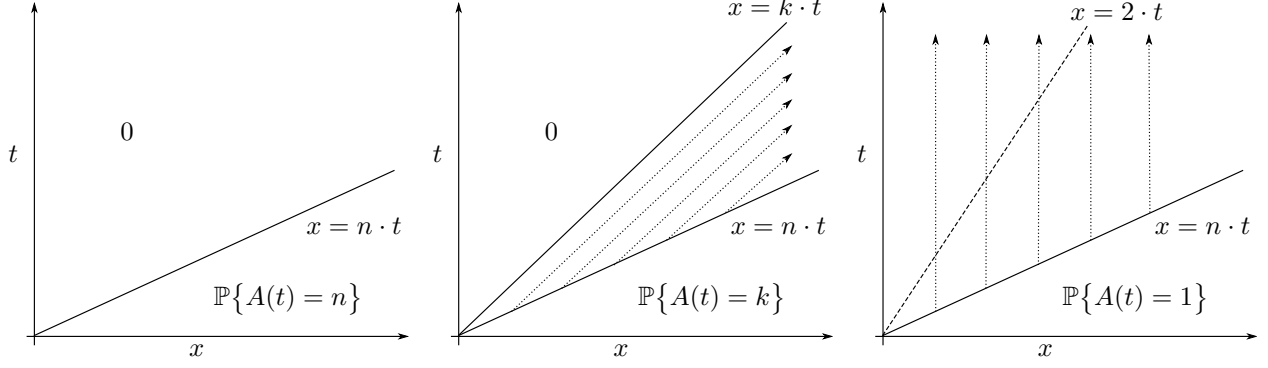
(a) The two only regions for the initial state $k = n$. The function has a discontinuity at $x = nt$.

(b) The three regions and the characteristics in the interior for an intermediate state with $1 < k < n$.

(c) Regions and characteristics for the absorbing state $k = 1$. The characteristics are parallel to the t-axis.

Figure 1: The different regions and characteristics of $F_k(t, x)$ (defined in equation (3.3)) for different values of k. In (c), according to Lemma 3.1, $F_k(t, x)$ does not depend on t beyond the dashed line $x = 2t$.

whit $\ell \geq k$. Thus, the system of PDEs (3.1) can be solved separately for each k, starting at $k = n$, and decreasing it step-by-step.

Furthermore, note that for $k \in \{1, \ldots, n\}$,

$$\mathbb{P}\{A(t) = k, L(t) \leq x\} = 0, \qquad \text{if } x < v(k)t, \tag{3.7}$$

since if the ancestral process has k lineages at time t, it must have accumulated at least $v(k)t$ towards the total tree length. It can be shown that the solution to equation (3.5) exhibits this property. Moreover,

$$\mathbb{P}\{A(t) = k, L(t) \leq x\} = \mathbb{P}\{A(t) = k\}, \qquad \text{if } x \geq n \cdot t,$$

since the process can have accumulated at most nt. Thus, we only have to use equation (3.5) to compute the solution

$$F_k(t, x) = \mathbb{P}\{A(t) = k, L(t) \leq x\}$$

when $v(k)t \leq x < n \cdot t$. Note that $v(1) = 0$. Moreover, for $k = n$, the region $v(k)t \leq x < n \cdot t$ is empty, and thus $F_n(t, x)$ has a discontinuity along the line $n \cdot t$. See Figure 1 for a visualization of the different regions for different values of k. To devise an accurate and efficient numerical scheme for computing the time-dependent CDFs in the interior region, we use the method of characteristics to solve the respective PDE

$$\partial_t F_k(t, x) + v(k)\partial_x F_k(t, x) = F_k(t, x)Q_{k,k}(t) + F_{k+1}(t, x)Q_{k+1,k}(t). \tag{3.8}$$

Since for $k = n$, the interior region is empty, we consider $k \neq n$ and introduce the family of characteristics

$$\tau \to \left(t_0 + \tau, x_0 + v(k)\tau\right)^\top \quad \text{with} \quad t_0 = \frac{x_0}{n}$$

11

Taking the derivative of $F_k(t,x)$ along such a characteristic yields

$$
\begin{aligned}
\frac{d}{d\tau} &F_k\Big(t_0 + \tau, x_0 + v(k)\tau\Big) \\
&= \left.\left(\frac{d}{d\tau}\Big[t_0 + \tau\Big] \cdot \partial_t F_k(t,x) + \frac{d}{d\tau}\Big[x_0 + v(k)\tau\Big] \cdot \partial_x F_k(t,x)\right)\right|_{(t,x)=(\frac{x_0}{n}+\tau, x_0 + v(k)\tau)} \\
&= \left.\Big(\partial_t F_k(t,x) + v(k)\partial_x F_k(t,x)\Big)\right|_{(t,x)=(\frac{x_0}{n}+\tau, x_0 + v(k)\tau)} \\
&= F_k\Big(t_0 + \tau, x_0 + v(k)\tau\Big)Q_{k,k}(t_0 + \tau) + F_{k+1}\Big(t_0 + \tau, x_0 + v(k)\tau\Big)Q_{k+1,k}(t_0 + \tau).
\end{aligned}
\tag{3.9}
$$

Here we used the chain rule and the fact that the third line is equal to the left-hand side of equation (3.8). Formally, the derivations (3.9) do not hold for all τ. It can be shown, however, that the equality holds for almost all τ; we omit the technical details here for readability. Thus, for given x_0, as a function of τ, the function $\tau \to F_k(t_0 + \tau, x_0 + v(k)\tau)$ solves the equation

$$
\frac{d}{d\tau} F_k\Big(t_0 + \tau, x_0 + v(k)\tau\Big) = -q_k^{(1)}(\tau)F_k\Big(t_0 + \tau, x_0 + v(k)\tau\Big) + g_k^{(1)}(\tau),
$$

with

$$
q_k^{(1)}(\tau) := -Q_{k,k}(t_0 + \tau) = \frac{k(k-1)}{2}\lambda(t_0 + \tau)
$$

and

$$
\begin{aligned}
g_k^{(1)}(\tau) &:= F_{k+1}\Big(t_0 + \tau, x_0 + v(k)\tau\Big)Q_{k+1,k}(t_0 + \tau) \\
&= F_{k+1}\Big(t_0 + \tau, x_0 + v(k)\tau\Big)\frac{(k+1)k}{2}\lambda(t_0 + \tau),
\end{aligned}
$$

Since this is a non-homogeneous linear first-order ODE, the solution can be readily obtained as

$$
F_k\Big(t_0 + \tau, x_0 + v(k)\tau\Big) = e^{-H_k^{(1)}(\tau)}\left(\int_0^\tau g_k^{(1)}(\alpha)e^{H_k^{(1)}(\alpha)}d\alpha + F_k\big(t_0, x_0\big)\right),
\tag{3.10}
$$

with

$$
H_k^{(1)}(\tau) := \int_0^\tau q_k^{(1)}(\alpha)d\alpha = \frac{k(k-1)}{2}\big(\Lambda(u) - \Lambda(t_0)\big).
\tag{3.11}
$$

The initial conditions for $\tau = 0$ are given by the boundary values of the associated PDE as

$$
F_k(t_0, x_0) = \mathbb{P}\big\{A(t_0) = k\big\}.
$$

Now, to obtain the value of the function $F_k(t,x)$, for given t and x, one just needs to identify the right characteristic and the parameters x_0 and τ such that $(t_0 + \tau, x_0 + v(k)\tau)^\top = (t,x)^\top$. Since the characteristics are parallel, it can be uniquely identified. Using these values of x_0 and τ in the solution (3.10) yields $F_k(t,x)$. However, we will not pursue this strategy to compute the requisite values of $F_k(t,x)$. Instead, we present a numerical upstream scheme in Appendix B.1 that can be used to compute $F_k(t,x)$ efficiently on a suitable grid to ultimately obtain values for the CDF $\mathbb{P}\{\mathcal{L} \leq x\}$.

3.2   Joint Distribution of the Total Tree Length

In this section we present a method to compute the joint CDF of the total tree length

$$
\mathbb{P}\{\mathcal{L}^a \leq x, \mathcal{L}^b \leq y\}
$$

12

at two loci a and b separated by a given recombination distance ρ. Again, we approach this problem by first computing the time-dependent joint CDF

$$F_s(t, x, y) = \mathbb{P}\{\bar{A}^\rho(t) = s, L^a(t) \leq x, L^b(t) \leq y\}.$$

We will follow closely along the lines of the method presented in Section 3.1, where we replace the ancestral process A by the ancestral process with limited recombination $\bar{A}^\rho$, and compute the integrals (2.13) and (2.14), to ultimately compute the joint CDF.

The analog to Lemma 3.1 is as follows.

Lemma 3.4. *With definition* (3.4), *the relation*

$$\begin{aligned}
\mathbb{P}\{\mathcal{L}^a \leq x, \mathcal{L}^b \leq y\} &= \mathbb{P}\{\bar{A}^\rho(\bar{t}) \in \Delta, L^a(\bar{t}) \leq x, L^b(\bar{t}) \leq y\} \\
&= F_{(1,0,0,0)}(\bar{t}, x, y) + F_{(1,0,0,1)}(\bar{t}, x, y)
\end{aligned}$$

holds for $x, y \in \mathbb{R}_+$, $\bar{t} \geq \max\{x, y\}/2$, *and* $\Delta = \{(1,0,0,0), (1,0,0,1)\}$, *the absorbing states of* $\bar{A}^\rho$.

Proof. The proof is similar to the proof of lemma 3.1. With

$$\bar{A}^\rho(t) = \big(\bar{K}_{ab}(t), \bar{K}_a(t), \bar{K}_b(t), \bar{R}(t)\big),$$

note that

$$2T_{\text{MRCA}}^a \leq \int_0^{T_{\text{MRCA}}^a} \mathbb{1}_{\{\bar{K}_{ab}(s) + \bar{K}_a(s) > 1\}} \big(\bar{K}_{ab}(s) + \bar{K}_a(s)\big) ds = \mathcal{L}^a,$$

and similarly $2T_{\text{MRCA}}^b \leq \mathcal{L}^b$. Thus, on the event $\{\mathcal{L}^a \leq x, \mathcal{L}^b \leq y\}$, the relations $T_{\text{MRCA}}^a \leq \max\{x, y\}/2 \leq \bar{t}$ and $T_{\text{MRCA}}^b \leq \bar{t}$ hold. This implies $\bar{K}_{ab}(\bar{t}) + \bar{K}_a(\bar{t}) = 1$ and $\bar{K}_{ab}(\bar{t}) + \bar{K}_a(\bar{t}) = 1$, which in turn implies $\bar{A}^\rho(\bar{t}) \in \Delta = \{(1,0,0,0), (1,0,0,1)\}$, since these two states are the only admissible states that can satisfy these conditions. Incidentally, these are also the absorbing states of $\bar{A}^\rho$. Thus,

$$\{\mathcal{L}^a \leq x, \mathcal{L}^b \leq y\} = \{\bar{A}^\rho(\bar{t}) \in \Delta, \mathcal{L}^a \leq x, \mathcal{L}^b \leq y\}$$

holds. Furthermore, on the event $\{\bar{A}^\rho(\bar{t}) \in \Delta\}$, $T_{\text{MRCA}}^a \leq \bar{t}$ and $T_{\text{MRCA}}^b \leq \bar{t}$ hold, which imply $L^a(\bar{t}) = \mathcal{L}^a$ and $L^b(\bar{t}) = \mathcal{L}^b$. This in turn implies

$$\{\bar{A}^\rho(\bar{t}) \in \Delta, \mathcal{L}^a \leq x, \mathcal{L}^b \leq y\} = \{\bar{A}^\rho(\bar{t}) \in \Delta, L^a(\bar{t}) \leq x, L^b(\bar{t}) \leq y\}.$$

Finally, note that

$$\{\bar{A}^\rho(\bar{t}) = (1,0,0,1)\} \cap \{\bar{A}^\rho(\bar{t}) = (1,0,0,0)\} = \emptyset,$$

which proves the statement of the lemma. $\qquad\square$

Again, Lemma 3.4 shows that the joint CDF of $\mathcal{L}^a$ and $\mathcal{L}^b$ can be computed from the time-dependent CDFs $F_{(1,0,0,0)}(t, x, y)$, and $F_{(1,0,0,1)}(t, x, y)$. In order to derive a system of PDEs like (3.5) for the time-dependent joint CDF of the tree length at two loci, we again apply Proposition A.14, for dimension $d = 2$. To this end, define the functions

$$v^a(k_{ab}, k_a, k_b, r) := \mathbb{1}_{\{k_{ab} + k_a > 1\}}(k_{ab} + k_a)$$

and

$$v^b(k_{ab}, k_a, k_b, r) := \mathbb{1}_{\{k_{ab} + k_b > 1\}}(k_{ab} + k_b)$$

that yield for $(k_{ab}, k_a, k_b, r) \in \bar{\mathcal{S}}^\rho$ the number of lineages ancestral to locus a and b, respectively, and define

$$V^a := \texttt{diag}\Big(\big(v^a(s)\big)_{s \in \bar{\mathcal{S}}^\rho}\Big) \qquad \text{and} \qquad V^b := \texttt{diag}\Big(\big(v^b(s)\big)_{s \in \bar{\mathcal{S}}^\rho}\Big).$$

We then have the following corollary.

Corollary 3.5. *The time-dependent joint CDF of the tree lengths*

$$\mathbf{F}(t, x, y) = \big(\mathbf{F}_s(t, x, y)\big)_{s \in \bar{\mathcal{S}}^\rho}$$

can be obtained for all points in $U = \Big\{(t, x, y) : 0 < x < nt, 0 < y < nt, \, t > 0\Big\}$ *as the strong solution of*

$$\partial_t \mathbf{F}(t, x, y) + \partial_x \mathbf{F}(t, x, y) V^a + \partial_y \mathbf{F}(t, x, y) V^b = \mathbf{F}(t, x, y) \bar{Q}(t), \tag{3.12}$$

with boundary conditions

$$\mathbf{F}(t, x, y) = \begin{cases} \Big(\mathbb{P}\{\bar{A}^\rho(t) = s, L^b(t) \le y\}\Big)_{s \in \bar{\mathcal{S}}^\rho} \cdot \mathbb{1}_{\{v^a(s) \ne n\}}, & \text{if } x = nt, \\[1.5ex] \Big(\mathbb{P}\{\bar{A}^\rho(t) = s, L^a(t) \le x\}\Big)_{s \in \bar{\mathcal{S}}^\rho} \cdot \mathbb{1}_{\{v^b(s) \ne n\}}, & \text{if } y = nt, \\[1.5ex] 0, & \text{if } x = 0 \text{ or } y = 0, \end{cases}$$

for $(x, y, t) \in \partial U$ *and* $\bar{Q}(t)$ *as defined in* (2.11).

Proof. Define the function $\mathbf{v} : \bar{\mathcal{S}}^\rho \to \mathbb{R}^2$ as

$$\mathbf{v}(k_{ab}, k_a, k_b, r) := \big(\mathbb{1}_{\{k_{ab} + k_a > 1\}}(k_{ab} + k_a), \mathbb{1}_{\{k_{ab} + k_b > 1\}}(k_{ab} + k_b)\big).$$

This function and the generator $\bar{Q}(t)$ satisfy the requirements of Proposition A.14, and thus, the statement of the corollary follows from Proposition A.14 and Remark A.16. □

Remark 3.6. *Note that due to symmetry of* $\bar{A}^\rho$,

$$\mathbb{P}\{\bar{A}^\rho(t) = s, L^a(t) \le x\} = \mathbb{P}\{\bar{A}^\rho(t) = s, L^b(t) \le x\}$$

holds.

The process $\big(A(t), L^a(t), L^b(t)\big)_{t \in \mathbb{R}_+}$ is a piecewise-deterministic Markov process as well (see Remark A.17), where $\bar{Q}(t)$ captures the stochastic dynamics, and ∂_x and ∂_y the deterministic dynamics. The numerical scheme to compute the time-dependent joint CDF is again an upstream scheme based on the method of characteristics and follows essentially along the lines of the scheme presented for the marginal case. The relation $\prec$ defined in (2.12) implies a partial ordering on the state space $\bar{\mathcal{S}}^\rho$, and the matrix $\bar{Q}(t)$ is triangular with respect to this ordering. Thus, again, the values of F_s only depends on $F_{s'}$ with $s \prec s'$, and they can be computed for each s separately. For given $s \in \bar{\mathcal{S}}^\rho$,

$$F_s(t, x, y) = \mathbb{P}\{\bar{A}^\rho(t) = s, L^a(t) \le x, L^b(t) \le y\}$$

$$= \begin{cases} 0, & \text{if } x < v^a(s) \cdot t \text{ or } y < v^b(s) \cdot t, \\ \text{solution to (3.12)}, & \text{if } v^a(s) \cdot t \le x < n \cdot t \text{ and } v^b(s) \cdot t \le y < n \cdot t, \\ \mathbb{P}\{\bar{A}^\rho(t) = s, L^a(t) \le x\}, & \text{if } v^a(s) \cdot t \le x < n \cdot t \text{ and } n \cdot t \le y, \\ \mathbb{P}\{\bar{A}^\rho(t) = s, L^b(t) \le y\}, & \text{if } n \cdot t \le x \text{ and } v^b(s) \cdot t \le y < n \cdot t, \\ \mathbb{P}\{\bar{A}^\rho(t) = s\}, & \text{if } n \cdot t \le x \text{ and } n \cdot t \le y \end{cases} \tag{3.13}$$

holds. Figure 2 shows the different regions of $F_s(t, x, y)$ for a fixed t. Moreover, for each $s \in \bar{\mathcal{S}}^\rho$, the PDE that has to be satisfied in the region $v^a(s) \cdot t \le x < n \cdot t$ and $v^b(s) \cdot t \le y < n \cdot t$ can be re-written as

$$\partial_t F_s(t, x, y) + \big(v^a(s), v^b(s)\big) \nabla F_s(t, x, y) = F_s(t, x, y) \bar{Q}_{s,s}(t) + \sum_{s \prec s'} F_{s'}(t, x, y) \bar{Q}_{s',s}(t), \tag{3.14}$$
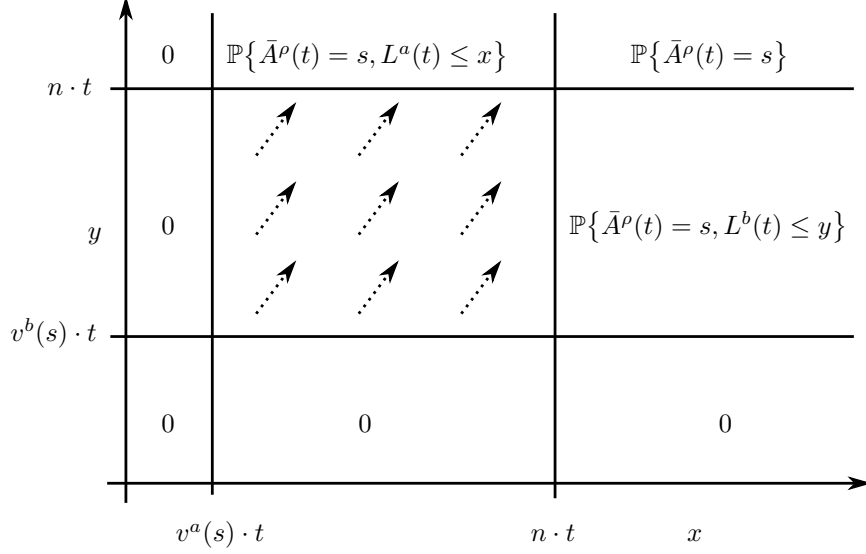
14

Figure 2: The different regions and (projected) characteristics of $F_s(t, x, y)$ (defined in equation (3.4)) for an intermediate state $s \in \bar{\mathcal{S}}^\rho$ at a given time t. The characteristics also extend in the t-direction at unit speed. Note that for the states s with $v^a(s) = n$ or $v^b(s) = n$ the interior region is empty.

where $\nabla f = (\partial_x f, \partial_y f)^\top$. Again, taking the derivative of $F_s(t, x, y)$ along the characteristics

$$\tau \to \left(t_0 + \tau, \mathbf{x}_0 + \tau \mathbf{v}(s)\right)^\top,$$

with $t_0 := \frac{1}{n} \max\{x_0, y_0\}$, $\mathbf{x}_0 := (x_0, y_0)$, and $\mathbf{v}(s) := \left(v^a(s), v^b(s)\right)$, yields the right-hand side of equation (3.14). Thus, $F_s(\cdot, \cdot, \cdot)$ satisfies the ODE

$$\frac{d}{d\tau} F_s\left(t_0 + \tau, \mathbf{x}_0 + \tau \mathbf{v}(s)\right) = -q_s^{(2)}(\tau) F_s\left(t_0 + \tau, \mathbf{x}_0 + \tau \mathbf{v}(s)\right) + g_s^{(2)}(\tau),$$

with

$$q_s^{(2)}(\tau) = -\bar{Q}_{s,s}(t_0 + \tau)$$

and

$$g_s^{(2)}(\tau) = \sum_{s \prec s'} F_{s'}\left(t_0 + \tau, \mathbf{x}_0 + \tau \mathbf{v}(s)\right) \bar{Q}_{s',s}(t_0 + \tau).$$

The characteristics for $F_s(t, x, y)$ are depicted in Figure 2. Like in the marginal case, this is a non-homogeneous linear first-order ODE and can be readily solved. The solution involves integrating $q_s^{(2)}(\tau)$, which leads to

$$F_s\left(t_0 + \tau, x_0 + \mathbf{v}(s)\tau\right) = e^{-H_k^{(2)}(\tau)}\left(\int_0^\tau g_s^{(2)}(\alpha) e^{H_k^{(2)}(\alpha)} d\alpha + F_s\left(t_0, x_0\right)\right), \tag{3.15}$$

with

$$H_s^{(2)}(\tau) = \int_0^\tau q_s^{(2)}(\alpha) d\alpha = -\bar{Q}_{s,s}^\rho (u - t_0) - \bar{Q}_{s,s}^c\left(\Lambda(u) - \Lambda(t_0)\right). \tag{3.16}$$

We provide the details of our numerical upstream scheme to efficiently and accurately compute solutions to equation (3.15) in Appendix B.2.
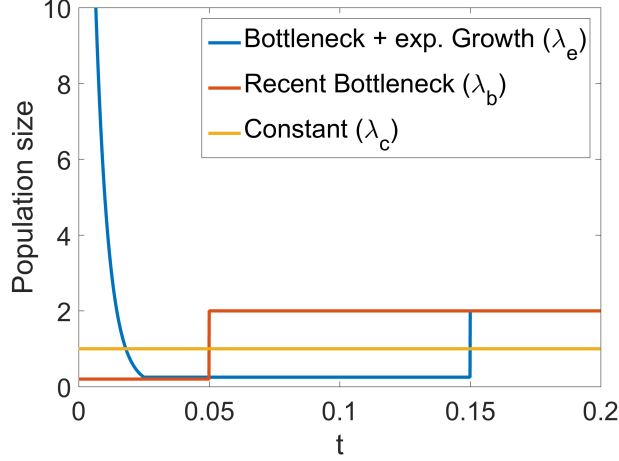
15

Figure 3: The three population size histories we will mainly consider in this paper: A constant population size (λ_c), an ancient bottleneck followed by exponential growth (λ_e), and a recent bottleneck (λ_b).

4 Empirical evaluation

In this section, we demonstrate that the numerical algorithms presented in Section B.1 and B.2 can be used to accurately and efficiently compute the time-dependent marginal CDF (3.3) and joint CDF (3.4), as well as the regular marginal CDF (3.1) and joint CDF (3.2), for different population size histories and different recombination rates. Furthermore, we show how our method can be used to study properties of the marginal and joint distributions, and compute their moments. We implemented the numerical algorithms in MATLAB, and the code is available upon request.

For ease of exposition, we use a sample size of $n = 10$ in the remainder of this paper, unless mentioned otherwise. We mainly focus on three population size histories, depicted in Figure 3. The first is a history of constant size 1, and we refer to the corresponding rate function as λ_c. Secondly, we consider a history with an ancient bottleneck, followed by exponential growth up to the present. Specifically, for $t > 0.15$, the relative population size is set to 2, and for $0.025 < t < 0.15$, it is set to 0.25. Then, the population grows exponentially from size 0.25 at $t = 0.025$ up to $t = 0$ (the present), at an exponential rate of g. We refer to this population size history by λ_e, and if not mentioned otherwise, the growth rate is set to $g = 200$. This size history is a rough sketch of the human population size history, with an out-of-Africa bottleneck, followed by recent exponential growth at a rate of 1% per generation. In addition, we consider a pure bottleneck, where the relative ancestral size is 2 until time $t = 0.05$, and N_B from $t = 0.05$ until the present. We refer to this size history by λ_b, and if not otherwise mentioned, we set $N_\text{B} = 0.2$.

4.1 Accuracy

In this section we demonstrate that the numerical algorithms presented in this paper can be used to compute the requisite CDFs accurately. Naturally, the accuracy will depend on the exact choice of the grid for the numerical algorithm. We will present results for a particular grid here, and discuss the issues for choosing an adequate grid in Section 5. We set $n = 5$ and compute the time-dependent marginal CDF

$$\mathbb{P}\big\{A(t) = k, L(t) \leq x\big\}$$

for $k = 5$, 3, and the absorbing state 1, and show the respective surfaces as functions of t and x in Figure 4. Here we used the population size history with exponential growth λ_e. These surfaces exhibit the properties sketched in Figure 1, and the different regions can be observed. Below the line $x = nt$, the functions are
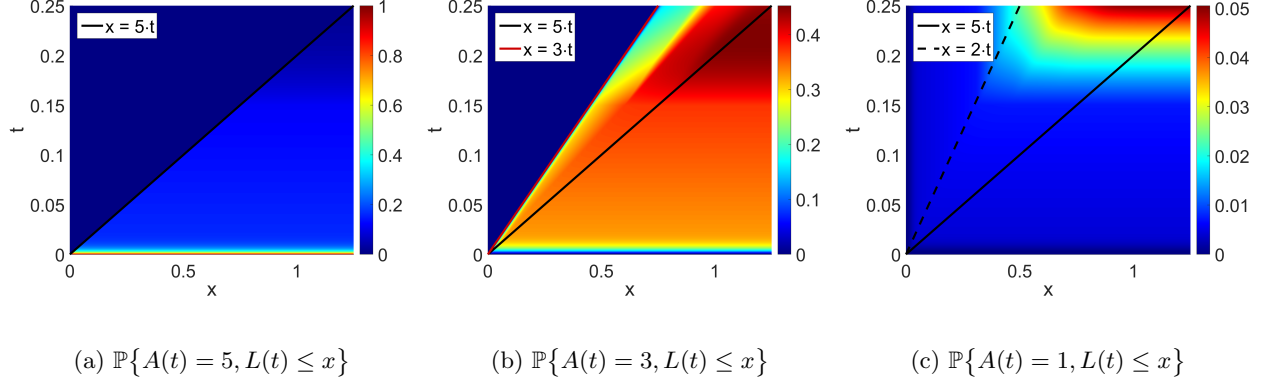
16

(a) $\mathbb{P}\{A(t) = 5, L(t) \leq x\}$ (b) $\mathbb{P}\{A(t) = 3, L(t) \leq x\}$ (c) $\mathbb{P}\{A(t) = 1, L(t) \leq x\}$

Figure 4: Heatmaps of $\mathbb{P}\{A(t) = k, L(t) \leq x\}$ (defined in equation (3.3)) as a function of t and x, for different k, computed using our numerical algorithm.

x	y	p	$\hat{p}$ ($N = 256{,}000$)	$\hat{p}$ ($N = 16{,}384{,}000$)
1.5	3.0	0.075326	0.074914 ($\pm$ 0.002)	0.075030 ($\pm$ 0.0002)
3.0	6.0	0.213703	0.213324 ($\pm$ 0.002)	0.213565 ($\pm$ 0.0002)
6.0	6.0	0.522821	0.521578 ($\pm$ 0.002)	0.522707 ($\pm$ 0.0003)
12.0	18.0	0.873357	0.872840 ($\pm$ 0.002)	0.873319 ($\pm$ 0.0002)
30.0	30.0	0.998499	0.998504 ($\pm$ 0.0002)	0.998516 ($\pm$ 0.00002)

Table 1: The CDF $\mathbb{P}\{\mathcal{L}^a \leq x, \mathcal{L}^b \leq y\}$ (defined in equation (3.2)) for different values of x and y under λ_1, with $n = 10$ and $\rho = 0.001$. p is computed using the numeric algorithm, and $\hat{p}$ is estimated from simulations for different N. The confidence bounds are indicated in parentheses.

independent of x. Furthermore, the functions are zero above the line $t = kt$, except for $k = 1$, where the function is independent of t above the line $x = 2t$.

As shown in Section 3, the marginal CDF of the total tree length

$$\mathbb{P}\{\mathcal{L} \leq x\}$$

and the joint CDF

$$\mathbb{P}\{\mathcal{L}^a \leq x, \mathcal{L}^b \leq y\}$$

can be computed from the respective time-dependent CDFs. To demonstrate the accuracy of our numerical algorithm, we compared the numerical values from the algorithm to simulations under the respective ancestral processes A and $\bar{A}^\rho$. To this end, we simulated a certain number N of trajectories from these processes, and estimated the respective probabilities. Figure 5 shows the marginal CDFs for $n = 10$ under exponential growth (λ_e) and the bottleneck scenario (λ_b). The simulations can also be used to bound the difference $d(P_{\text{pde}}, P_{\text{T}})$ between the values computed using the numerical scheme P_{pde} and the true value P_{T}. These bounds are indicated in Figure 5 for different values of N and decrease as N gets larger, as expected. For the joint CDF, we present numerical values for different x and y, and compare them to the respective estimates from the simulations, including confidence bounds for these estimates. We set $n = 10$, and used $\rho = 0.001$. The values for the model with exponential growth (λ_e) are shown in Table 1, and for the bottleneck scenario (λ_b) in Table 2. The values computed using the numeric algorithm always fall into the confidence bounds, demonstrating that our algorithm computes the respective values accurately. In these tables, it becomes particularly apparent that in order to guarantee a high accuracy using simulations, a very large number of trajectories needs to be simulated, which is time-consuming. Our numerical scheme yields a high accuracy, and does not suffer from these issues.

17
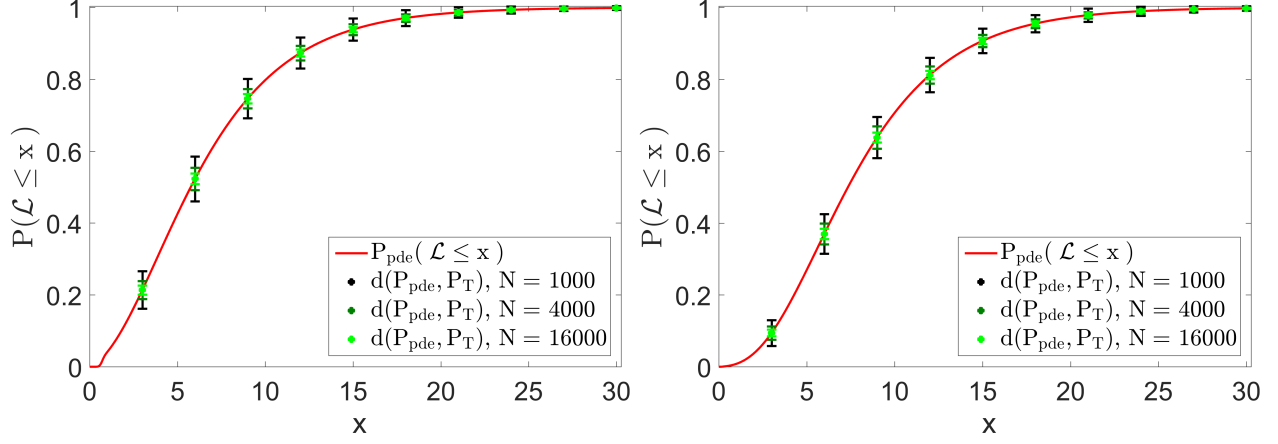
(a) $\mathbb{P}\{\mathcal{L} \leq x\}$ under exponential population growth (λ_e).

(b) $\mathbb{P}\{\mathcal{L} \leq x\}$ in the bottleneck scenario (λ_b).

Figure 5: The CDF $\mathbb{P}\{\mathcal{L} \leq x\}$ (defined in equation (3.1)) as a function of x is depicted by the red line. Additionally, the green bars indicate the bound on the distance between the numerical value P_{pde} and the true value P_{T} for different N, thus the true value is guaranteed to fall within these bounds.

x	y	p	$\hat{p}$ ($N = 256,000$)	$\hat{p}$ ($N = 16,384,000$)
1.5	3.0	0.019794	0.019238 ($\pm$ 0.0006)	0.019579 ($\pm$ 0.00007)
3.0	6.0	0.094393	0.094414 ($\pm$ 0.002)	0.094172 ($\pm$ 0.0002)
6.0	6.0	0.369581	0.369059 ($\pm$ 0.002)	0.369544 ($\pm$ 0.0003)
12.0	18.0	0.812236	0.812328 ($\pm$ 0.002)	0.812109 ($\pm$ 0.0002)
30.0	30.0	0.997696	0.997922 ($\pm$ 0.0002)	0.997721 ($\pm$ 0.00003)

Table 2: The CDF $\mathbb{P}\{\mathcal{L}^a \leq x, \mathcal{L}^b \leq y\}$ (defined in equation (3.2)) for different values of x and y under λ_2, with $n = 10$ and $\rho = 0.001$. p is computed using the numeric algorithm, and $\hat{p}$ is estimated from simulations for different N. The confidence bounds are indicated in parentheses.

(a) The expected value of $\mathcal{L}$.
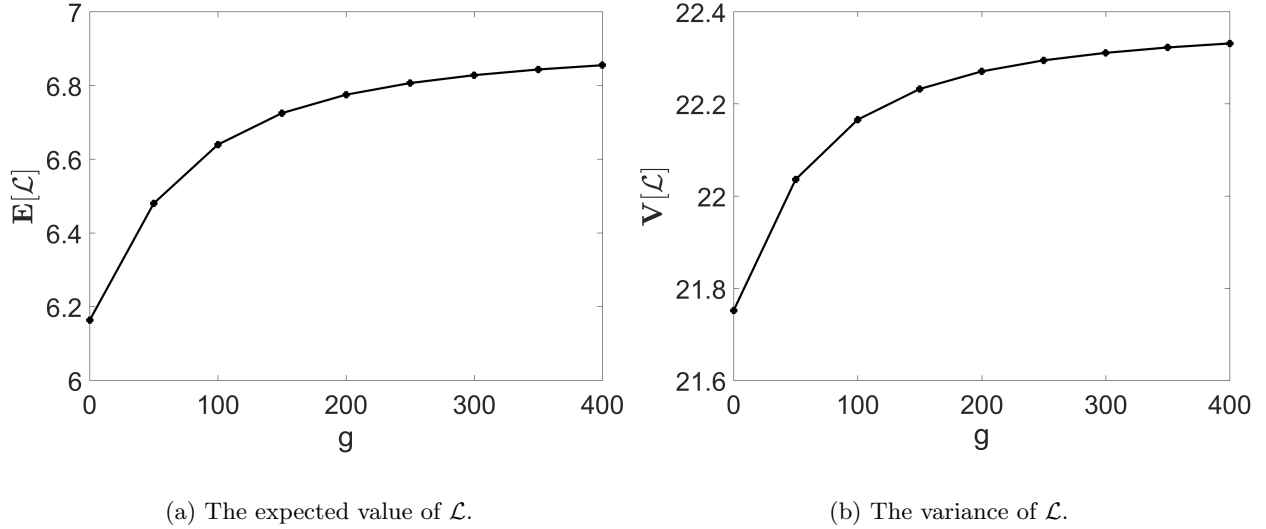
(b) The variance of $\mathcal{L}$.

Figure 6: Approximations to the expected value and the variance of the total tree length $\mathcal{L}$ (defined in equation (2.7)) computed using our numerical procedure, under the model for exponential growth (λ_e), with different values for the growth-rate g.

4.2 Properties of the Distributions

The results provided in the previous section show that our numerical algorithm can be used to accurately and efficiently compute the marginal and joint CDF of the total tree length in populations with variable size. We will now demonstrate the utility of our numerical method to study properties of the respective distributions.

The numerical values of the marginal CDF $\mathbb{P}\{\mathcal{L} \leq x\}$ can be readily applied to compute approximations of the expected value and the variance of the total tree length $\mathcal{L}$. Figure 6 shows the different values of the expectation and the variance under exponential growth (λ_e), with varying growth-rates g. Recall that a rate of $g = 0$ corresponds to no growth. Figure 7 shows the expected value and the variance under the bottleneck model (λ_b) for different values of the bottleneck size N_B. In both scenarios, the expected value and the variance are smallest in the models with the smallest contemporary population size, corresponding to the largest recent coalescent rate. They increase as g, respectively N_B, increases, but level off, indicating that increasing the population size has diminishing effects for large values. The absolute value of the expectation is higher in the bottleneck scenario, because, independent of the growth parameter, there is a substantial bottleneck in the growth-scenario.

Figure 8 shows the joint CDF $\mathbb{P}\{\mathcal{L}^a \leq x, \mathcal{L}^b \leq y\}$ as a function of x and y for different population size scenarios and different recombination rates ρ, computed on a suitable grid using our numerical algorithm. Naturally, the CDF converges towards 1 as x and y increase, and due to the symmetry of the ancestral process $\bar{A}^\rho$ the CDF is symmetric when interchanging x and y. Furthermore, note that the isolines in the plots for $\rho = 0.0001$ show pronounced right angles along the line $x = y$, because for small ρ the trees at the two loci are highly correlated. As the recombination rate increases, the two tree lengths become increasingly uncorrelated, and these angles soften. In all plots, the isoline for 0.2 is around $x = y = 5$, for the case λ_e even lower. Thus, under λ_e, there is an elevated probability for very short trees, likely due to the strong bottleneck, which favors short trees. Under the constant population size model λ_c, the CDF increases rapidly as x and y increase, whereas the function is less steep for λ_e and λ_b. This behavior seems to be dominated by the ancient population sizes.

Finally, we employ our numerical values of the joint CDF to compute approximations to the correlation
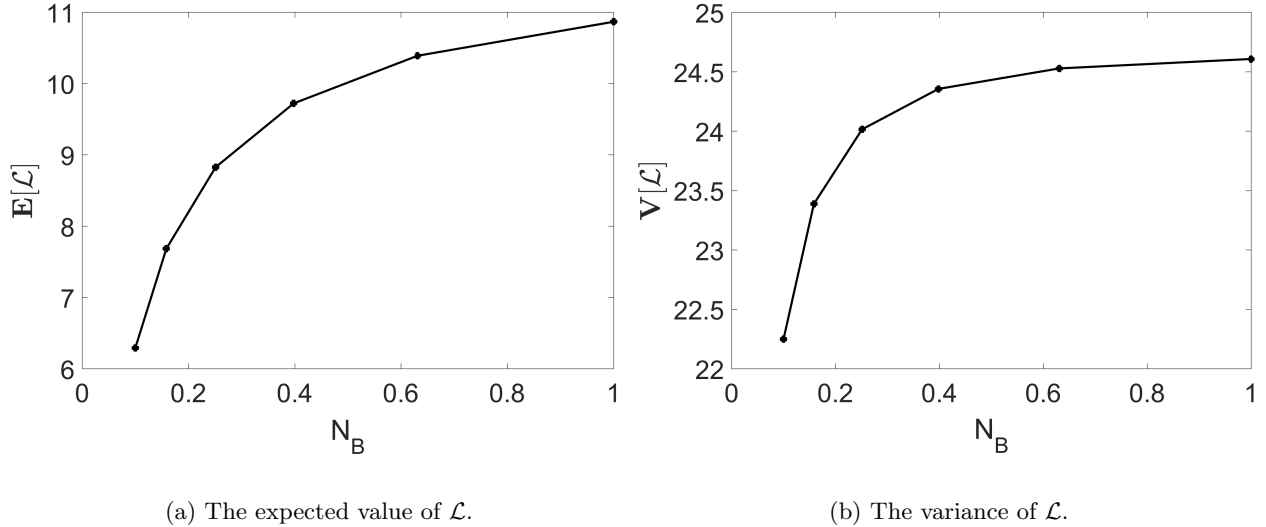
(a) The expected value of $\mathcal{L}$.

(b) The variance of $\mathcal{L}$.

Figure 7: Approximations to the expected value and the variance of the total tree length $\mathcal{L}$ (defined in equation (2.7)), under the bottleneck model (λ_b), with different values for the bottleneck size N_B.

coefficient between the tree lengths

$$\operatorname{corr}(\mathcal{L}^a, \mathcal{L}^b) := \frac{\operatorname{cov}(\mathcal{L}^a, \mathcal{L}^b)}{\sqrt{\mathbb{V}\mathcal{L}^a}\sqrt{\mathbb{V}\mathcal{L}^b}},$$

where $\operatorname{cov}(\cdot, \cdot)$ denotes the covariance. Figure 9 shows this correlation coefficient under the population size history λ_e for different values of ρ, and sample sizes $n = 5$ and $n = 20$, respectively. Recall that our numerical procedure was derived using the approximate ancestral process $\bar{A}^\rho$ for computational efficiency, where we limited the number of recombination events to 1. To compare the correlation under the process $\bar{A}^\rho$ with the correlation under the regular ancestral process with recombination A^ρ, we estimated the latter from repeated simulations using the widely applied coalescent-simulation tool `ms` (Hudson, 2002), which is based on the regular coalescent with recombination (using $N = 10^7$ repetitions). Naturally, the correlation is close to 1 for small recombination rates, and it decreases with increasing recombination rate. The values are basically indistinguishable until they start separating around $\rho = 0.05$. This is to be expected, since the approximation we introduced limits the number of recombination events to 1, and thus, as the recombination rate increases, the approximation error also increases.

To further investigate how restrictive the assumption of at most one recombination event is, we also used the simulated trajectories to estimate the probability that two or more recombination events occur under the regular ancestral process A^ρ. The results are shown in Figure 10. These probabilities increase with increasing ρ and n. However, they remain small for $\rho \leq 0.05$, which is in good agreement with the observation that the correlation is well approximated for ρ up to 0.05. In conclusion, Figure 9 and Figure 10 show that the approximate process can be used without loss of accuracy for a large range of recombination rates relevant for human genetics, where recombination rates between neighboring sites are on the order of 10^{-3}.

5 Discussion

In this paper, we presented a novel computational framework to compute the marginal and joint CDF of the total tree length in populations with variable size. To our knowledge, these distributions have not been addressed in the literature before, especially in populations of variable size. We introduced a system of linear hyperbolic PDEs and showed that the requisite CDFs can be obtained from the solution of this
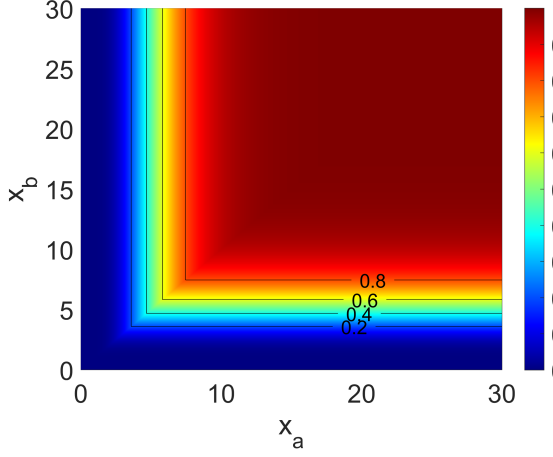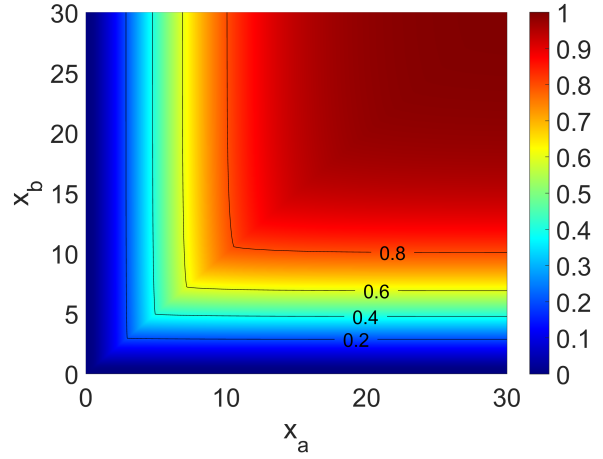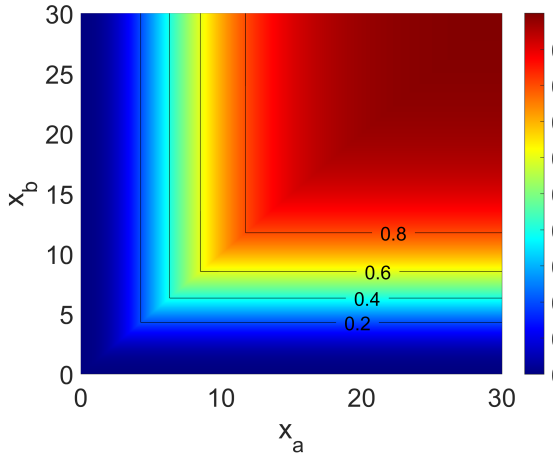
(a) Constant population size (λ_c), $\rho = 0.0001$.
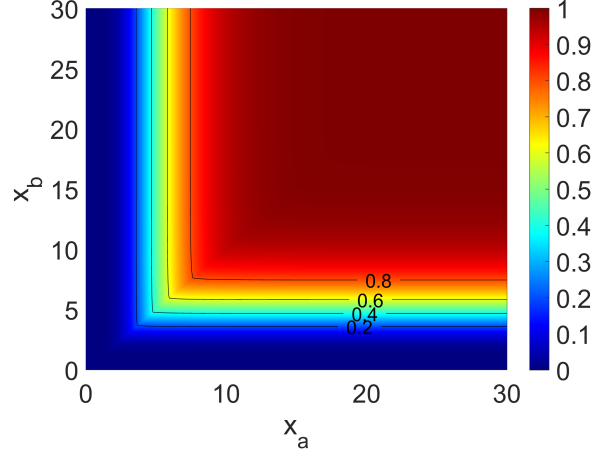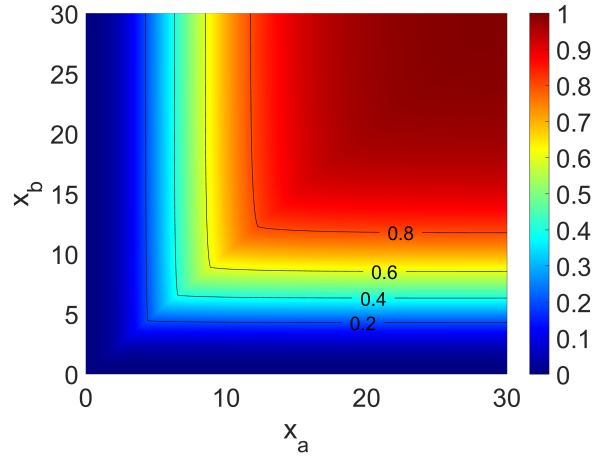
(b) Constant population size (λ_c), $\rho = 0.1$.
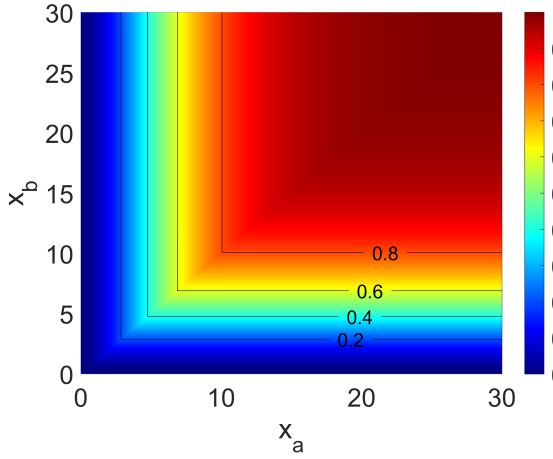
(c) Exponential growth (λ_e), $\rho = 0.0001$.

(d) Exponential growth (λ_e), $\rho = 0.1$.

(e) Bottleneck model (λ_b), $\rho = 0.0001$.

(f) Bottleneck model (λ_b), $\rho = 0.1$.

Figure 8: The joint CDF $\mathbb{P}\{\mathcal{L}^a \leq x, \mathcal{L}^b \leq y\}$ (defined in equation (3.2)) for the three population models (rows), with different recombination rates ρ (columns). Again, we use $n = 10$.
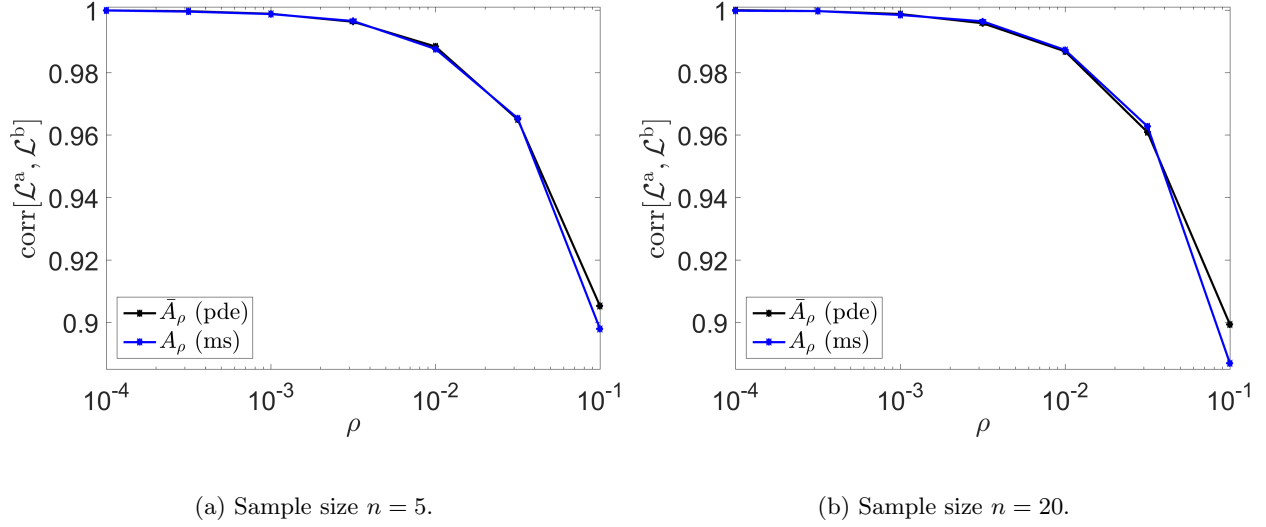
21

(a) Sample size $n = 5$.

(b) Sample size $n = 20$.

Figure 9: Correlation between $\mathcal{L}^a$ and $\mathcal{L}^b$ (defined in equations (2.13) and (2.14)) under the exponential growth model (λ_e) for different sample sizes n and different recombination rates ρ. The black lines show the values computed using our method under $\bar{A}^\rho$, and the blue lines show values estimated from coalescent simulations under A^ρ using the popular tool `ms` (using $N = 10^7$ repetitions).
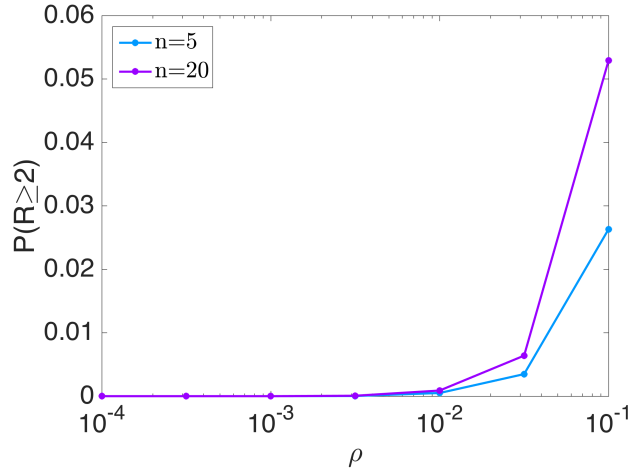

Figure 10: Probability of two or more recombination events R in the regular ancestral process A^ρ (Definition 2.4), under exponential growth (λ_e), for different different sample sizes n and different recombination rates ρ. These values were estimated using the coalescent simulation tool `ms` (using $N = 10^7$ repetitions).

system. We introduced a numerical algorithm to compute the solution of this system based on the method of characteristics and demonstrated its accuracy in a wide range of biologically relevant scenarios.

The numerical algorithm that we introduced is an upstream-method that computes the requisite solutions step-wise on a grid. We presented the algorithm for a regular, equidistantly spaced grid. We used the trapezoidal rule for the integration steps in the method, and also used linear interpolation to interpolate values that do not fall onto the specified grid. We used these basic approaches for ease of exposition. Using higher order interpolation and integration schemes, combined with adaptive grids that have more points in regions where the coalescent-rate function is large will most certainly increase the accuracy. However, such higher order schemes come with additional computational cost. This opens numerous avenues for future research to optimize the balance between accuracy and efficiency that is required in the respective applications.

Moreover, for reasons of computational efficiency, we introduced the first-order approximation $\bar{A}^\rho$ to the regular ancestral process with recombination A^ρ, and computed the joint CDF under this approximate process. We demonstrated that this approximation is accurate for a large range of relevant recombination rates. It is straightforward to use higher order approximations, including more recombination events, to gain additional accuracy, but computing the joint CDF under the regular ancestral process is desirable. Proposition A.14 guarantees that we can use our numerical procedure to compute the requisite CDF under $\bar{A}^\rho$, but it is conceivable that it can be extended to more general processes like A^ρ in future work.

Another research direction is to use our novel framework to study higher order correlations between trees at multiple loci. On the one hand, this could again be correlations between the total tree lengths, but the distribution of other summary statistics of the genealogical trees could be included, for example, the length of the external branches or the length of all branches subtending k leaves. Statistics that have been successfully used in the literature, like the coalescence time between two lineages (Li and Durbin, 2011; Terhorst et al., 2017) or the time of the first coalescent event (Schiffels and Durbin, 2014) could be used as well. Our framework is flexible enough to compute the distribution of multiple path integrals along the trajectories of a given Markov chain. Thus, to implement these additions, one needs to define and implement an appropriate ancestral process and compute suitable integrals along the trajectories.

In this paper, we studied the ancestral process in a single panmictic population. However, in recent years, researchers have gathered an increasing amount of genomic datasets that contain individuals from multiple sub-populations, and studied historical events like migration or population subdivision using these datasets. In light of these studies, it is important to augment our framework to compute joint CDFs of the total tree length in structured populations with complex migration histories. Again, this can be done by introducing suitable ancestral processes and suitable integrals along their trajectories.

Acknowledgements

Appendix

A Path-integrals of Markov chains

Since the marginal and joint distributions of the tree length can be obtained by integrating a certain function of the ancestral processes, we now consider distributions of path integrals for Markov chains. We will introduce these distributions assuming Lipschitz continuity in Section A.1, and then show in Section A.2 that these assumptions can be relaxed if the state space is monotone.

A.1 Path-integrals under regularity assumptions

Let the Markov chain X be defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We will be using the following assumptions throughout the section.

(A1) $\{X(t, \omega), t \in \mathbb{R}_+, \omega \in \Omega\}$ is a regular jump Markov process with values in a finite state space $\mathcal{S}_n$, for convenience labeled $\mathcal{S}_n = \{1, 2, \ldots, n\}$, satisfying

$$\mathbb{P}\{X(t+h) = j | X(t) = i\} = q_{ij}(t)h + o(h), \quad i, j \in \mathcal{S}_n \quad \text{as} \quad h \to 0^+.$$

We assume that the trajectories of $t \to X(\cdot, \omega)$ are right-continuous.

(A2) The infinitesimal generator $Q(t) = \{q_{ij}(t)\}_{i,j=1,\ldots,n}$ is conservative, that is

$$q_i(t) := -q_{ii}(t) = \sum_{i \neq j} q_{ij}(t),$$

and satisfies $Q \in C(\mathbb{R}_+; M^{n \times n}) \bigcap L^\infty(\mathbb{R}_+; M^{n \times n})$. In addition, for each $i \in \mathcal{S}_n$ either $q_i(t) = 0$ for all $t \geq 0$ or $q_i(t) > 0$ for all $t \geq 0$. In the latter case, we require $\int_0^\infty q_i(s)\, ds = \infty$.

Definition A.1. *Let $X(t)$ satisfy (A1)-(A2). Given a function*

$$\mathbf{v} = (v_1, v_2, \ldots, v_d) : \mathcal{S}_n \to \mathbb{R}^d$$

we define a vector-valued path-integral over the interval $[0, t]$ by

$$\mathbf{L}^{\mathbf{v}}(t, \omega) := \int_0^t \mathbf{v}\big(X(s, \omega)\big)\, ds \, \in \, \mathbb{R}^d, \quad t \in \mathbb{R}_+.$$

Definition A.2. *Let $X(t)$ satisfy (A1)-(A2) and $\mathbf{v} : \mathcal{S}_n \to \mathbb{R}^d$ be some real-valued function defined on the state space. We define a distribution vector-function associated with $\big(X(t), \mathbf{L}^{\mathbf{v}}(t)\big)$ by*

$$\mathbf{F}^{\mathbf{v}} = \big(F_1^{\mathbf{v}}, \ldots, F_n^{\mathbf{v}}\big) : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}_+^n \tag{A.1}$$

with

$$F_k^{\mathbf{v}}(t, \mathbf{x}) := \mathbb{P}\Big\{X(t) = k, \mathbf{L}^{\mathbf{v}}(t) \leq \mathbf{x}\Big\},$$

$k \in \{1, \ldots, n\}$, $\mathbf{x} \in \mathbb{R}^d$, and the comparison is understood componentwise.

Definition A.3. *Let $\mathbf{v} : \mathcal{S}_n \to \mathbb{R}^d$ be some real-valued function defined on the state space. Define*

$$\mathbf{m}^{\mathbf{v}} := \big(\min_{j \in \mathcal{S}_n} v_1(j), \ldots, \min_{j \in \mathcal{S}_n} v_d(j)\big)$$

and

$$\mathbf{M}^{\mathbf{v}} := \big(\max_{j \in \mathcal{S}_n} v_1(j), \ldots, \max_{j \in \mathcal{S}_n} v_d(j)\big).$$

as the componentwise minima and maxima.

Remark A.4. Since $X(t)$ is a regular jump process, it is separable. Thus, for each $t_0 \in \mathbb{R}_+$ the random variable $\mathbf{L}^{\mathbf{v}}(t_0, \cdot)$ is well-defined and $\mathcal{F}$-measurable, and for each $\omega_0 \in \Omega$ the map $t \to \mathbf{L}^{\mathbf{v}}(t, \omega_0)$ is Lipschitz continuous. This in turn implies that the process $\mathbf{L}^{\mathbf{v}}(t)$ is measurable and separable (see Chapter 12 of Koralov and Sinay (2007) and Chapter 2 of Doob (1953)).

Proposition A.5 (Locally Lipschitz). *Let (A1)-(A2) hold. Let $\mathbf{v} : \mathcal{S}_n \to \mathbb{R}^d$ and $\mathbf{F}^{\mathbf{v}}$ be defined by (A.1). Suppose that $\mathbf{F}^{\mathbf{v}}$ is Lipschitz continuous in an open set $\mathcal{U} \subset \mathbb{R}^+ \times \mathbb{R}^d$. Then*

$$\partial_t \mathbf{F}^{\mathbf{v}}(t, \mathbf{x}) + \sum_{j=1}^{d} \partial_{x_j} \mathbf{F}^{\mathbf{v}}(t, \mathbf{x}) V_j = \mathbf{F}^{\mathbf{v}}(t, \mathbf{x}) Q(t) \quad a.e. \ in \ \ \mathcal{U}, \tag{A.2}$$

where $V_j = \mathbf{diag}(v_j(1), \ldots, v_j(n))$.

To prove this statement, we first need to introduce the following lemmas.

Lemma A.6. *Let $X(t)$ satisfy (A1)-(A2), $\mathbf{v} : \mathcal{S}_n \to \mathbb{R}^d$. Then, for any $i, k \in \mathcal{S}_n$, any $\varepsilon > 0$, and each $\mathbf{x} \in \mathbb{R}^d$, we have*

$$\mathbb{P}\{X(t) = i, \ X(t + \varepsilon) = k\} F_i^{\mathbf{v}}(t, \mathbf{x} - \varepsilon \mathbf{M}^{\mathbf{v}})$$

$$\leq \mathbb{P}\{X(t) = i, X(t + \varepsilon) = k, \mathbf{L}^{\mathbf{v}}(t + \varepsilon) \leq \mathbf{x}\} \mathbb{P}\{X(t) = i\}$$

$$\leq \mathbb{P}\{X(t) = i, X(t + \varepsilon) = k\} F_i^{\mathbf{v}}(t, \mathbf{x} - \varepsilon \mathbf{m}^{\mathbf{v}}),$$

and the comparison is understood componentwise.

Proof. We suppress the superscript $\mathbf{v}$ in our calculations. Take any $\varepsilon > 0$. Observe that

$$\mathbf{L}(t) + \varepsilon \mathbf{m} \leq \mathbf{L}(t + \varepsilon) = \mathbf{L}(t) + \int_t^{t+\varepsilon} \mathbf{v}(X(s)) \, ds \leq \mathbf{L}(t) + \varepsilon \mathbf{M}$$

and therefore

$$\mathbb{P}\{X(t) = i, \ X(t + \varepsilon) = k, \mathbf{L}(t) \leq \mathbf{x} - \varepsilon \mathbf{M}\}$$

$$\leq \mathbb{P}\{X(t) = i, \ X(t + \varepsilon) = k, \mathbf{L}(t + \varepsilon) \leq \mathbf{x}\}$$

$$\leq \mathbb{P}\{X(t) = i, \ X(t + \varepsilon) = k, \mathbf{L}(t) \leq \mathbf{x} - \varepsilon \mathbf{m}\}.$$

Let $\mathbf{y} \in \mathbb{R}^d$. Suppose that $0 < F_i(t, \mathbf{y})$. Observe that for any $t_0 > 0$ the path-integral $\mathbf{L}(t_0)$ is fully determined by $(X(t), t \in [0, t_0))$. This fact (and the separability of the process) enables us to use the Markov property and we obtain

$$\mathbb{P}\{X(t) = i, \ X(t + \varepsilon) = k, \mathbf{L}(t) \leq \mathbf{y}\}$$

$$= \mathbb{P}\{X(t + \varepsilon) = k | X(t) = i, \mathbf{L}(t) \leq \mathbf{y}\} \mathbb{P}\{X(t) = i, \mathbf{L}(t) \leq \mathbf{y}\}$$

$$= \mathbb{P}\{X(t + \varepsilon) = k | X(t) = i\} F_i(t, \mathbf{y}).$$

This together with the previous inequality finishes the proof. $\qquad\square$

Lemma A.7. *Let $X(t)$ satisfy (A1)-(A2) and $\mathbf{v} : \mathcal{S}_n \to \mathbb{R}^d$. For every $(t, \mathbf{x}) \in \mathbb{R}_+ \times \mathbb{R}^d$ and each $k \in \mathcal{S}_n$, we have*

$$\lim_{\varepsilon \to 0^+} \frac{1}{\varepsilon} \left| \mathbb{P}\{X(t) = k, X(t + \varepsilon) = k, \mathbf{L}^{\mathbf{v}}(t + \varepsilon) \leq \mathbf{x} + \varepsilon \mathbf{v}(k)\} - (1 - q_k(t)\varepsilon) F_k(t, \mathbf{x}) \right| = 0.$$

Proof. We suppress the superscript $\mathbf{v}$ in what follows. Due to (A1) the process $X(t)$ is separable and therefore (see (Karlin, 1981, p. 146))

$$\mathbb{P}\{X(s) = k, s \in [t, t + \varepsilon]\} = \exp\left\{ - \int_t^{t+\varepsilon} q_i(s) \, ds \right\} \mathbb{P}\{X(t) = k\}. \tag{A.3}$$

Suppose now that $F_k(t, \mathbf{x}) > 0$. Then, using the Markov property, we obtain

$$\mathbb{P}\{X(s) = k, s \in [t, t + \varepsilon], \mathbf{L}(t + \varepsilon) \leq \mathbf{x} + \varepsilon \mathbf{v}(k)\}$$

$$= \mathbb{P}\{X(s) = k, s \in [t, t + \varepsilon] \, | X(t) = k, \mathbf{L}(t) \leq \mathbf{x}\} \mathbb{P}\{X(t) = k, \mathbf{L}(t) \leq \mathbf{x}\}$$

$$= \mathbb{P}\{X(s) = k, s \in [t, t + \varepsilon] | X(t) = k\} F_k(t, \mathbf{x}) = \exp\left(- \int_t^{t+\varepsilon} q_k(s) \, ds \right) F_k(t, \mathbf{x}).$$

If $F_k(t, \mathbf{x}) = 0$, then the first and the last term in the above identity are zero. Thus, employing (A.3) and recalling that Q is continuous, we conclude

$$\frac{1}{\varepsilon}\left|\mathbb{P}\Big\{X(t) = k, X(t + \varepsilon) = k, \mathbf{L}(t + \varepsilon) \leq \mathbf{x} + \varepsilon\,\mathbf{v}(k)\Big\} - \exp\Big(-\int_t^{t+\varepsilon} q_k(s)\,ds\Big)F_k(t, \mathbf{x})\right|$$

$$= \frac{1}{\varepsilon}\left(\mathbb{P}\big\{X(t) = k, X(t + \varepsilon) = k, \mathbf{L}(t + \varepsilon) \leq \mathbf{x} + \varepsilon\,\mathbf{v}(k)\big\} - \mathbb{P}\big\{X(s) = k\,, s \in [t, t + \varepsilon]\,, \mathbf{L}(t + \varepsilon) \leq \mathbf{x} + \varepsilon\,\mathbf{v}(k)\big\}\right)$$

$$\leq \frac{1}{\varepsilon}\left(\mathbb{P}\Big(X(t) = k, X(t + \varepsilon) = k\Big) - \mathbb{P}\Big(X(s) = k\,, s \in [t, t + \varepsilon]\Big)\right)$$

$$\leq \frac{1}{\varepsilon}\left(\mathbb{P}\big\{X(t) = k, X(t + \varepsilon) = k\big\} - \mathbb{P}\big\{X(t) = k\big\}\right)$$

$$- \frac{1}{\varepsilon}\left[\exp\Big(-\int_t^{t+\varepsilon} q_i(s)\,ds\Big) - 1\right]\mathbb{P}\big\{X(t) = k\big\} \to 0 \quad \text{as} \quad \varepsilon \to 0^+.$$

Since $(1 - q_k(t)\varepsilon)F_k(t, \mathbf{x}) = \exp\Big(-\int_t^{t+\varepsilon} q_k(s)\,ds\Big)F_k(t, \mathbf{x}) + o(\varepsilon^2)$, the statement of the lemma follows. $\qquad\square$

We can now turn back to the proof of Proposition A.5

Proof of Proposition A.5. Let us suppress the superscript $\mathbf{v}$ in our calculations. Let $\widetilde{\mathcal{U}}$ denote the set of all points in $\mathcal{U}$ at which $\mathbf{F}$ is differentiable. Since $\mathbf{F}$ is Lipschitz continuous in $\mathcal{U}$, Rademacher's Theorem (Federer, 1969) implies that $\mathbf{F}$ is Lebesgue almost surely differentiable in $\mathcal{U}$ and therefore $\mathcal{U} \backslash \widetilde{\mathcal{U}}$ is of Lebesgue measure zero. Take any $k \in \mathcal{S}_n$. Fix any $(t, \mathbf{x}) \in \widetilde{\mathcal{U}}$. For any $\varepsilon > 0$ we have

$$F_k\big(t + \varepsilon, \mathbf{x} + \varepsilon\,\mathbf{v}(k)\big) - F_k(t, \mathbf{x})$$
$$= \Big(\sum_{i=1}^n \mathbb{P}\{X(t) = i\,, X(t + \varepsilon) = k\,, \mathbf{L}(t + \varepsilon) \leq \mathbf{x} + \varepsilon\,\mathbf{v}(k)\}\Big) - F_k(t, \mathbf{x}). \tag{A.4}$$

Consider first the terms with $i \neq k$. By Lemma A.6, we have

$$\frac{1}{\varepsilon}\mathbb{P}\big\{X(t + \varepsilon) = k, X(t) = i\big\}F_i\big(t + \varepsilon, \mathbf{x} + \varepsilon\,\mathbf{v}(k) - \varepsilon\,\mathbf{M}\big)$$

$$\leq \frac{1}{\varepsilon}\mathbb{P}\big\{X(t) = k, X(t + \varepsilon) = i\,, \mathbf{L}(t + \varepsilon) \leq \mathbf{x} + \varepsilon\,\mathbf{v}(k)\big\}\mathbb{P}\big\{X(t) = i\big\}$$

$$\leq \frac{1}{\varepsilon}\mathbb{P}\big\{X(t + \varepsilon) = k, X(t) = i\big\}F_i\big(t + \varepsilon, \mathbf{x} + \varepsilon\,\mathbf{v}(k) - \varepsilon\,\mathbf{m}\big)$$

where $\mathbf{m}$ and $\mathbf{M}$ are as in Definition A.3. Since Q is continuous we must have

$$\lim_{\varepsilon \to 0^+} \frac{1}{\varepsilon}\mathbb{P}\big\{X(t + \varepsilon) = i, X(t) = k\big\} = q_{ki}(t)\mathbb{P}\big\{X(t) = k\big\}$$

and hence, employing the continuity of $\mathbf{F}$, we conclude

$$\lim_{\varepsilon \to 0^+} \frac{1}{\varepsilon}\mathbb{P}\Big\{X(t) = i, X(t + \varepsilon) = k,\ \mathbf{L}(t + \varepsilon) \leq \mathbf{x} + \varepsilon\,\mathbf{v}(k)\Big\} = q_{ki}(t)F_i(t, \mathbf{x})\,. \tag{A.5}$$

Next consider the case $i = k$. By Lemma A.7 we have

$$\lim_{\varepsilon \to 0^+} \frac{1}{\varepsilon}\Big(\mathbb{P}\big\{X(t) = k, X(t + \varepsilon) = k,\ \mathbf{L}(t + \varepsilon) \leq \mathbf{x} + \varepsilon\,\mathbf{v}(k)\big\} - F_k(t, \mathbf{x})\Big) = -q_k(t)F_k(t, \mathbf{x})\,. \tag{A.6}$$

Combining (A.4) with (A.5) and (A.6) we obtain

$$\lim_{\varepsilon \to 0^+} \frac{1}{\varepsilon}\Big(F_k\big(t + \varepsilon, \mathbf{x} + \varepsilon\,\mathbf{v}(k)\big) - F_k(t, \mathbf{x})\Big)$$
$$= -q_k(t)F_k(t, \mathbf{x}) + \sum_{i \neq k} q_{ik}(t)F_i(t, \mathbf{x}) = \big(\mathbf{F}(t, \mathbf{x})Q(t)\big)_k\,. \tag{A.7}$$

Since $\mathbf{F}$ is differentiable at $(t, \mathbf{x}) \in \widetilde{\mathcal{U}}$ and the map $\varepsilon \to (t + \varepsilon, \mathbf{x} + \varepsilon\, \mathbf{v}(k))$ is differentiable with the image contained in $\mathcal{U}$ for sufficiently small ε, the chain rule is applicable (see Rudin (1976)[Theorem 9.15]) and we conclude

$$\lim_{\varepsilon \to 0^+} \frac{1}{\varepsilon} \Big(F_k\big(t + \varepsilon, \mathbf{x} + \varepsilon\, \mathbf{v}(k)\big) - F_k(t, \mathbf{x}) \Big) = \frac{d}{d\varepsilon} F_k\big(t + \varepsilon, \mathbf{x} + \varepsilon\, \mathbf{v}(k)\big)\Big|_{\varepsilon=0}$$

$$= \partial_t F_k(t, \mathbf{x}) + \sum_{j=1}^{d} v_j(k)\, \partial_{x_j} F_k(t, \mathbf{x}).$$

Since both $k \in \mathcal{S}_n$ and $(t, \mathbf{x}) \in \widetilde{\mathcal{U}}$ were arbitrary, (A.7) implies (A.2). $\qquad\square$

We next show that $\mathbf{F^v}$ as $t \to 0^+$ has certain continuity properties.

Proposition A.8. *Let (A1)-(A2) hold. Let $\mathbf{v} : \mathcal{S}_n \to \mathbb{R}^d$ and $\mathbf{F^v}$ as defined by (A.1). Then*

$$\lim_{t \to 0^+} F_k^{\mathbf{v}}(t, \mathbf{x}) = \mathbb{1}_{\mathbb{R}^d_+}(\mathbf{x})\, \mathbb{P}\{X(0) = k\} = F_k^{\mathbf{v}}(0, \mathbf{x}) \quad for \quad \mathbf{x} \notin \partial \mathbb{R}^d_+.$$

Proof. First, we note that $\mathbf{L}(0) = 0$ for all $\omega \in \Omega$ and hence $F_k^{\mathbf{v}}(0, \mathbf{x}) = \mathbb{1}_{\mathbb{R}^d_+}(\mathbf{x}) \mathbb{P}\{X(0) = k\}$.
  Observe that

$$\mathbf{m^v} t \le \mathbf{L}(t) = \int_0^t \mathbf{v}(X(s))\, ds \le \mathbf{M^v} t,$$

where $\mathbf{m^v}$ and $\mathbf{M^v}$ as in Definition A.3. Fix $\delta > 0$. Then for all $0 < t < \delta(1 + \max(||m^v||_\infty, ||M^v||_\infty))^{-1}$, and every $\mathbf{x} \in \mathbb{R}^d$ such that $||\mathbf{x} - \mathbf{y}|| > \delta$ for all $y \in \partial \mathbb{R}^d_+$, we have

$$F_k(t, \mathbf{x}) = \mathbb{1}_{\mathbb{R}^d_+}(\mathbf{x}) \mathbb{P}\{X(t) = k, \mathbf{L}(t) \le \mathbf{x}\}$$

$$= \mathbb{1}_{\mathbb{R}^d_+}(\mathbf{x}) \mathbb{P}\{X(t) = k\} \to \mathbb{1}_{\mathbb{R}^d_+}(\mathbf{x}) P\{X(0) = k\} \quad \text{as} \quad t \to 0^+.$$

$\qquad\square$

Remark A.9. *From Proposition A.8 it follows that the 'initial values' of $\mathbf{F^v}$ are discontinuous. Since the system (A.2) a is linear hyperbolic system, discontinuities present at time $t = 0$ will travel in space as time t increases and therefore $\mathbf{F^v}$ is not C^1 or even continuous. Nevertheless, one can show that (A.2) holds in a weaker sense. To do that one needs to employ the notion of weak solutions, and we will pursue this avenue in an upcoming paper. However, for certain type of state space functions $\mathbf{v}$, relevant to our application, one can show additional regularity properties of $\mathbf{F^v}$. We provide more details in Appendix A.2.*

A.2   Path-integrals for monotone state space functions

Hyperbolic systems of partial differential equations admit in general solutions that are not classical even if the initial (or boundary data) is smooth. Typically there are two distinct classes of solutions: strong solutions, which are Lipschitz continuous (see Dafermos (2010)), and weak solutions, which allow for discontinuities. Here, we will be using the first type of solutions.

Definition A.10. *Let $\mathcal{U} \subset \mathbb{R}_+ \times \mathbb{R}^d$ be open and let $A_1, \ldots, A_d, B \in L^\infty(\mathcal{U}; \mathbb{R}^{n \times n})$. We say that $\mathbf{u}(t, \mathbf{x}) : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}^n$ is a strong solution of*

$$\partial_t \mathbf{u}(t, \mathbf{x}) + \sum_{j=1}^{d} \partial_{x_j} \big\{ \mathbf{u}(t, \mathbf{x}) \big\} A_j(t, \mathbf{x}) = \mathbf{u}(t, \mathbf{x}) B(t, \mathbf{x}) \quad in \quad \mathcal{U} \subset \mathbb{R}_+ \times \mathbb{R}^d, \qquad (A.8)$$

if $\mathbf{u}$ is Lipschitz continuous in $\mathcal{U}$, and the equation (A.8) holds for Lebesgue almost all points $(t, \mathbf{x})$ in $\mathcal{U}$.

Remark A.11. *By the Rademacher's Theorem (Federer, 1969) a function $\mathbf{u}(t, \mathbf{x})$ that is Lipschitz continuous in an open domain $\mathcal{U}$ is Lebesgue almost sure differentiable in $\mathcal{U}$. In fact, its pointwise partial derivatives, which exist almost everywhere, coincide with its corresponding weak partial derivatives (see Evans (2010)).*

Regularity of solutions to hyperbolic problems depends on both the initial (or boundary) data and the domain itself. For linear hyperbolic problems as long as the initial data is smooth and the domain has a smooth boundary one may expect a solution to be (locally) smooth. Typically one studies solutions to hyperbolic problems on the domain $\mathcal{U} = \mathbb{R}_+ \times \mathbb{R}^d$ with initial data $\mathbf{u}_0(\mathbf{x})$ at $t = 0$ (Cauchy problem). The initial data for the vector of probabilities $\mathbf{F}^{\mathbf{v}}$ (studied in $\mathcal{U}$) are unfortunately discontinuous (which is shown below). To avoid unnecessary difficulties, in Proposition A.14 we split the space-time domain into two regions $\mathcal{U}_I$ and $\mathcal{U}_E$. In $\mathcal{U}_E$ the values of $\mathbf{F}^{\mathbf{v}}$ admit a simpler form while in $\mathcal{U}_I$ the vector $\mathbf{F}^{\mathbf{v}}$ is obtained via solving a linear hyperbolic system with smooth initial data. We note that components of $\mathbf{F}^{\mathbf{v}}$ are in general merely Lipschitz continuous in $\mathcal{U}_I$. This is not surprising for two reasons. Firstly, the domain is singular because it has a 'corner' and the discontinuities of the derivatives of $\mathbf{F}^{\mathbf{v}}$ originating at points $\partial \mathbb{R}_+^d$ travel along the corresponding characteristics. Secondly, the vector $\mathbf{F}^{\mathbf{v}}$ solves the same system of equations in the domain $\mathcal{U}$ with discontinuous initial data and hence it is in general not smooth.

Definition A.12. *Let (A1)-(A2) hold. Let $\mathbf{v} : \mathcal{S}_n \to \mathbb{R}^d$. We say that $\mathbf{v}$ is monotone along the process $X(t)$ if the map $t \to \mathbf{v}(X(t))$ is either non-increasing $\mathbb{P}$-almost surely or non-decreasing $\mathbb{P}$-almost surely.*

Definition A.13. *Define the following regions in $\mathbb{R}_+ \times \mathbb{R}^d$:*

$$\mathcal{U}_I := \left\{ (t, \mathbf{x}) : t > 0, \, \mathbf{x} < \mathbf{M}^{\mathbf{v}} t \right\}$$

and

$$\mathcal{U}_E := \left(\mathcal{U}_I \cup \partial \mathcal{U}_I \right)^c,$$

where the comparison is understood componentwise.

Proposition A.14. *Let $X(t)$ satisfy (A1)-(A2), $X(0) = n$, and $Q(t)$ be lower triangular. Let $\mathbf{v} : \mathcal{S}_n \to \mathbb{R}^d$ be monotone along $X(t)$. Suppose that $t \to \mathbf{v}(X(t))$ is non-increasing on Ω, and $\mathbf{m}^{\mathbf{v}} < \mathbf{M}^{\mathbf{v}}$. For $\mathbf{x} \in \mathbb{R}^d$, define $J(t, \mathbf{x}) = \left\{ j : x_j < M_j^{\mathbf{v}} t \right\}$. Then $\mathbf{F}^{\mathbf{v}}$ defined by (A.1) has the following properties:*

(i) *For each $i \in \mathcal{S}_n$, with $\mathbf{v}(i) < \mathbf{M}^{\mathbf{v}}$, $F_i^{\mathbf{v}}(t, \mathbf{x})$ is Lipschitz continuous on $\mathbb{R}_+ \times \mathbb{R}^d$.*

(ii) *For each $i \in \mathcal{S}_n$, with $\mathbf{v}(i) \not< \mathbf{M}^{\mathbf{v}}$, $F_i^{\mathbf{v}}(t, \mathbf{x}) = \mathbb{1}_{\overline{\mathcal{U}_E}}(\mathbf{x}) \, \mathbb{P}\left\{ X(t) = i, L_j^{\mathbf{v}}(t) \leq x_j, j \in J(t, \mathbf{x}) \right\}$.*

(iii) $\mathbf{F}^{\mathbf{v}}$ *is a strong solution of*

$$\partial_t \mathbf{F}^{\mathbf{v}}(t, \mathbf{x}) + \sum_{j=1}^d \partial_{x_j} \mathbf{F}^{\mathbf{v}}(t, \mathbf{x}) V_j = \mathbf{F}^{\mathbf{v}}(t, \mathbf{x}) Q(t), \tag{A.9}$$

where $V_j = \mathbf{diag}(v_j(1), \ldots, v_j(n))$, in the open region $\mathcal{U}_I$. Furthermore, let $(t, \mathbf{x}) \in \partial \mathcal{U}_I$. Then

$$\lim_{(\bar{t}, \bar{\mathbf{x}}) \in \mathcal{U}_I \to (t, \mathbf{x})} F_i^{\mathbf{v}}(\bar{t}, \bar{\mathbf{x}}) = \begin{cases} \mathbb{P}\left\{ X(t) = i, L_j^{\mathbf{v}}(t) \leq x_j, j \in J(t, \mathbf{x}) \right\}, & \text{if } \mathbf{v}(i) < \mathbf{M}^{\mathbf{v}}, \\ 0, & \text{otherwise}, \end{cases} \tag{A.10}$$

and

$$F_i^{\mathbf{v}}(\bar{t}, \bar{\mathbf{x}}) = \mathbb{P}\left\{ X(\bar{t}) = i, L_j^{\mathbf{v}}(\bar{t}) \leq \bar{x}_j, j \in J(t, \bar{\mathbf{x}}) \right\} \quad \text{for all } (\bar{t}, \bar{\mathbf{x}}) \in \mathcal{U}_E. \tag{A.11}$$

Remark A.15. *Computing the solution in Proposition A.14 for a given number d of path-integrals requires computing solutions for $\bar{d}$ integrals with $\bar{d} < d$ on the boundary. These can be obtained by straightforwardly applying the Proposition in lower dimensions. Note that for $d = 1$, the values on the boundary can be directly obtained from the distribution of $X(t)$.*

Remark A.16. *Note that for each $i \in \mathcal{S}_n$ we have*

$$F_i^{\mathbf{v}}(t, \mathbf{x}) = 0,$$

for $\mathbf{x} \leq \mathbf{v}(i)t$.

Remark A.17. *The process $\big(X(t), \mathbf{L}^{\mathbf{v}}(t)\big)_{t \in \mathbb{R}_+}$ is a time-inhomogeneous piecewise-deterministic strong Markov process (Davis, 1993, Chapter 2), and Proposition A.14 essentially shows that the generator is given by*

$$\mathcal{G}_t \mathbf{H}(\mathbf{x}) = -\sum_{j=1}^{d} \partial_{x_j} \mathbf{H}(\mathbf{x}) V_j + \mathbf{H}(\mathbf{x}) Q(t),$$

for suitably defined functions $\mathbf{H}(\mathbf{x})$. The stochastic transitions of $X(t)$ are described by $Q(t)$ and the deterministic evolution of $\mathbf{L}^{\mathbf{v}}(t)$ in each dimension is governed by the terms $V_j \, \partial_{x_j}$. However, in addition, Proposition A.14 establishes regularity of $\mathbf{F}^{\mathbf{v}}(t, \mathbf{x})$, which is important for numerical computations.

Remark A.18. *Then ancestral process with limited recombination satisfies assumptions (A1)-(A2), and thus, we focus on this case here. It is conceivable that these assumptions could be relaxed and Proposition A.14 could be extended to more general Markov chains $X(t)$ with a (countably) infinite state space, and more general dynamics, for example, a non-triangular rate matrix $Q(t)$, or $\int_0^\infty q_i(s) \, ds < \infty$. However, the approach presented here in the proof of Proposition A.14 to show the necessary regularity of $\mathbf{F}^{\mathbf{v}}(t, \mathbf{x})$ uses the fact that $X(t)$ has absorbing states, and reaches them in finite time, after a finite number of jumps. For a more general version, this strategy would need to be adapted, or a different strategy used.*

Proof of Proposition A.14. Let Δ denote the set of absorbing states of the process $X(t)$. Since Q is lower triangular, $1 \in \Delta$ and thus Δ is not empty.

Take any $i \in \mathcal{S}_n$ with $\mathbf{v}(i) \not< \mathbf{M}^{\mathbf{v}}$. Since $\mathbf{v}$ is monotone along the process we conclude that

$$L_j(t, \omega) = \int_0^t v_j\big(X(s, \omega)\big) \, ds = M_j^{\mathbf{v}} t \quad \text{for all} \quad j \notin J\big(1, \mathbf{v}(i)\big), \, \omega \in \{\tilde{\omega} : X(t, \tilde{\omega}) = i\}$$

and this yields (ii).

Recall next that for a time-inhomogeneous Markov processes $X(t)$ (under the assumptions (A1)-(A2)) the jumping times $T_1, T_2, T_3, \ldots$ of $X(t)$ satisfy $\mathbb{P}\{T_1 > \alpha\} = \exp\big(-\int_0^\alpha q_1(s) \, ds\big)$ and for $k \geq 2$

$$\mathbb{P}\Big\{T_k > t + \alpha \,\Big|\, T_{k-1} = t, X(T_{k-1}) = i\Big\} = \exp\Big(-\int_t^{t+\alpha} q_i(s) \, ds\Big). \tag{A.12}$$

Take any $i \in \mathcal{S}_n$ with $\mathbf{v}(i) < \mathbf{M}^{\mathbf{v}}$, in which case $i < n$. Since $Q(t)$ is lower triangular, each trajectory of the process has at most $n - 1$ jumps before it enters into the absorbing set Δ. Thus we obtain

$$\Big\{\omega : X(\cdot, \omega) \text{ enters the state } i\Big\} = \bigcup_{k=1}^{n-1} \Omega_k^{(i)}, \quad \Omega_k^{(i)} = \Big\{\omega : X(\cdot, \omega) \text{ enters the state } i \text{ on the } k\text{-th jump}\Big\}.$$

We next denote $T_0 = 0$, $s_0 = n$, $s_i^{(k)} = (s_1, s_2, \ldots, s_{k-1}, s_k = i) \in (\mathcal{S}_n)^k$, with $k \geq 1$, and

$$A\big(s_i^{(k)}\big) = \Big\{\omega : X(T_1) = s_1, \ldots, X(T_{k-1}) = s_{k-1}, X(T_k) = s_k = i\Big\} \subset \Omega_k^{(i)}.$$

First, suppose that $i \notin \Delta$. For $(t, \mathbf{x}) \in \mathbb{R}_+ \times \mathbb{R}^d$, using the above partitioning, we write

$$\begin{aligned}
F_i^{\mathbf{v}}(t, \mathbf{x}) &= \mathbb{P}\{X(t) = i, \mathbf{L}(t) \leq \mathbf{x}\} \\
&= \sum_{k=1}^{n-1} \sum_{s_i^{(k)} \in \mathcal{S}^k} \mathbb{P}\Big\{A\big(s_i^{(k)}\big), T_k < t < T_{k+1}, \mathbf{L}(t) \leq \mathbf{x}\Big\} \\
&= \sum_{k=1}^{n-1} \sum_{s_i^{(k)} \in \mathcal{S}^k} \mathbb{P}\Big\{A\big(s_i^{(k)}\big), T_k < t < T_{k+1}, \sum_{j=1}^{k} T_j\big(\mathbf{v}(s_{j-1}) - \mathbf{v}(s_j)\big) \leq \mathbf{x} - \mathbf{v}(i)t\Big\}.
\end{aligned} \tag{A.13}$$

We now show that $F_i^{\mathbf{v}}$ is Lipschitz continuous. To this end, consider the function

$$G\big(t, \mathbf{x}; s_i^{(k)}\big) = \mathbb{P}\Big\{ A\big(s_i^{(k)}\big), T_k < t < T_{k+1}, \sum_{j=1}^{k} T_j\big(\mathbf{v}(s_{j-1}) - \mathbf{v}(s_j)\big) \le \mathbf{x} \Big\}.$$

Observe that $G\big(t, \mathbf{x}; s_i^{(k)}\big)$ is well-defined for $(t, \mathbf{x}) \in \mathbb{R}^{1+d}$. Moreover, since $i \notin \Delta$, the assumption (A2) implies that the process after entering the state i leaves this state in finite time $\mathbb{P}$-almost surely. Thus $\Omega_k^{(i)} \subset \{T_k < \infty\} \subset \{T_{k+1} < \infty\}$ and therefore

$$G\big(t, \mathbf{x}; s_i^{(k)}\big) = \mathbb{P}\Big\{ A\big(s^{(k)}\big), T_k < t, \sum_{j=1}^{k} T_j\big(\mathbf{v}(s_{j-1}) - \mathbf{v}(s_j)\big) \le \mathbf{x} \Big\}$$

$$- \mathbb{P}\Big\{ A\big(s^{(k)}\big), T_{k+1} < t, \sum_{j=1}^{k} T_j\big(\mathbf{v}(s_{j-1}) - \mathbf{v}(s_j)\big) \le \mathbf{x} \Big\}$$

$$=: G_1\big(t, \mathbf{x}; s_i^{(k)}\big) - G_2\big(t, \mathbf{x}; s_i^{(k)}\big).$$

Now, using (A.12) and induction, one can show that for each $r \in \mathcal{S}_n$ and $k \ge 1$

$$\mathbb{P}\Big\{ A\big(s_r^{(k)}\big), T_{k+1} \le z \Big\} = \int_{-\infty}^{z} f_{k+1}\big(\alpha; s_r^{(k)}\big)\, d\alpha \tag{A.14}$$

where $f_{k+1}\big(\cdot; s_r^{(k)}\big)$ is a globally bounded function. Thus, we conclude that the map

$$z \to \mathbb{P}\Big\{ A\big(s_r^{(k)}\big), T_{k+1} \le z \Big\}$$

is globally Lipschitz for each $k \ge 1$.

Since $\mathbf{v}$ is non-increasing along the process, for each $s_i^{(k)}$ we have

$$\mathbf{v}(n) = \mathbf{M}^{\mathbf{v}} \ge \mathbf{v}(s_1) \cdots \ge \mathbf{v}(s_k) = \mathbf{v}(i).$$

By assumption $\mathbf{v}(i) < \mathbf{M}^{\mathbf{v}}$ and hence for each $l \in \{1, \dots, d\}$ there exists $k_l \in \{1, \dots, k\}$ such that $v_l(s_{k_l - 1}) - v_l(s_{k_l}) > 0$, which guarantees that not all terms in the nonnegative sum $\sum_{j=1}^{k} T_j\big(v_l(s_{j-1}) - v_l(s_j)\big)$ vanish. Then, in view of the fact that the event $A\big(s_i^{(k)}\big)$ does not depend on the $(t, \mathbf{x})$-variable, we can use (A.14) and induction to conclude that

$$\mathbb{P}\Big\{ A\big(s_i^{(k)}\big), \sum_{j=1}^{k} T_j\big(v_l(s_{j-1}) - v_l(s_j)\big) \le x_l \Big\} = \int_{-\infty}^{x_l} \tilde{f}_{kl}(s)\, ds, \quad k \ge 1, l \in \{1, \dots, d\}$$

for some globally bounded function $\tilde{f}_{kl}$. It can be shown that this implies that

$$\mathbf{x} \to \mathbb{P}\Big\{ A\big(s_i^{(k)}\big), \sum_{j=1}^{k} T_j\big(\mathbf{v}(s_{j-1}) - \mathbf{v}(s_j)\big) \le \mathbf{x} \Big\} \tag{A.15}$$

globally Lipschitz. Combining (A.14) with (A.15) and using the definition of the Lipschitz continuity we conclude that $G_1\big(t, \mathbf{x}; s_i^{(k)}\big)$ and $G_2\big(t, \mathbf{x}; s_i^{(k)}\big)$ are globally Lipschitz and hence $G\big(t, \mathbf{x}; s_i^{(k)}\big)$ is as well.

Furthermore, any Lipschitz continuous function composed with a linear map is also Lipschitz continuous. Thus $\bar{G}\big(t, \mathbf{x}; s_i^{(k)}\big) := G\big(B(t, \mathbf{x}); s_i^{(k)}\big)$, where $B(t, \mathbf{x}) = \big(t, \mathbf{x} - \mathbf{v}(i)t\big)$, is globally Lipschitz. In (A.13) each of the terms in the sum is one of the functions $\bar{G}\big(x, t; s_i^{(k)}\big)$. Hence F_i which is restricted to $(t, \mathbf{x}) \in [0, \infty) \times \mathbb{R}^d$ is globally Lipschitz on this domain.

Lastly, if $i \in \Delta$, observe that

$$\Big\{ T_k < \infty, X(T_k) = i \Big\} \subset \Big\{ T_{k+1} = \infty \Big\}$$

30

and therefore

$$F_i^{\mathbf{v}}(t, \mathbf{x}) = \mathbb{P}\{X(t) = i, \mathbf{L}(t) \leq \mathbf{x}\}$$

$$= \sum_{k=1}^{n-1} \sum_{s_i^{(k)} \in \mathcal{S}^k} \mathbb{P}\Big\{A\big(s_i^{(k)}\big), T_k < t, \sum_{j=1}^{k} T_j\big(\mathbf{v}(s_{j-1}) - \mathbf{v}(s_j)\big) \leq \mathbf{x} - \mathbf{v}(i)t\Big\}.$$

Using an analogous approach (to the one in the case $i \notin \Delta$) one can show that each term in the above expression is globally Lipschitz continuous. This yields (i).

From (i) and (ii) it follows that $\mathbf{F}^{\mathbf{v}}$ is Lipschitz continuous in the open region $\mathcal{U}_I$. Then, by Proposition A.8 we conclude that $\mathbf{F}^{\mathbf{v}}$ is a strong solution of (A.9) in $\mathcal{U}_I$. The boundary conditions (A.10) and equation (A.11) follow directly from the definition of $\mathbf{F}^{\mathbf{v}}$. This proves (iii).

$\square$

B   Numerical Schemes

B.1   Upstream Numerical Scheme for Single-Locus Case

Here we present a numerical algorithm for computing solutions to the system (3.5). The numerical scheme is an upstream scheme based on the method of characteristics. In particular, the numerical scheme we develop makes use of the integral representation formulas (3.10), (3.11).

To define a grid in the (t, x)-space suitable for computation, choose $x_{\max}$, the maximum value that the CDF $\mathbb{P}\{\mathcal{L} \leq x_{\max}\}$ should be computed for. Due to Lemma 3.1, the relation $\mathbb{P}\{\mathcal{L} \leq x\} = F_1(t_{\max}, x)$ holds for all $x \leq x_{\max}$, with $t_{\max} := \frac{x_{\max}}{2}$. Thus $t_{\max}$ is set as the maximal gridpoint for t. In addition to the maximum gridpoints, choose small step sizes Δt and Δx. The number of gridpoints in the t dimension is then given by $M := \lceil \frac{t_{\max}}{\Delta t} \rceil + 1$, and the set of gridpoints is given as

$$T := \big\{0, \Delta t, 2\Delta t, \dots, (M-1)\Delta t, \min(M\Delta t, t_{\max})\big\}. \tag{B.1}$$

For each point T_i, define a grid in the x-dimension as

$$X_i := \big\{0, \Delta x, \dots, \min(U\Delta x, nT_i)\big\} \cup \big\{2\Delta t + \bar{X}_{i-1}, 3\Delta t + \bar{X}_{i-1}, \dots, n\Delta t + \bar{X}_{i-1}\big\} \cup \big\{2T_i, 3T_i, \dots, nT_i\big\}, \tag{B.2}$$

with $U = \lceil \frac{nT_i}{\Delta x} \rceil$ and $\bar{X}_{i-1} := \max(X_{i-1})$. Furthermore, set $U_i := |X_i|$. The same grid will be used for all $k \in \{1, \dots, n\}$. The points $k\Delta t + \bar{X}_{i-1}$ and kT_i are added for numerical stability reasons, to improve the accuracy of the interpolation we will perform in subsequent steps.

Now fix $i \in \{0, \dots, M\}$ and $k \in \{1, \dots, n\}$, and assume that $F_\ell(T_{i-1}, X_{i-1,j})$ has been computed for all $\ell \in \{1, \dots, n\}$ and $X_{i-1,j} \in X_{i-1}$. Furthermore, assume that $F_\ell(T_i, X_{i,j})$ has been computed for all $\ell \in \{k+1, \dots, n\}$ and $X_{i,j} \in X_i$. Under these assumptions, $F_k(T_i, X_{i,j})$ can be computed for all $X_{i,j} \in X_i$ as follows. If $X_{i,j} < v(k)T_i$, then

$$F_k(T_i, X_{i,j}) = 0. \tag{B.3}$$

If $X_{i,j} = nT_i$, the maximal value of X_i, then

$$F_k(T_i, X_{i,j}) = \mathbb{P}\Big\{A\big(T_i\big) = k\Big\}. \tag{B.4}$$

The values on the right-hand side can be pre-computed for all k and $T_i \in T$ by solving the ODE (2.3) numerically. In the general case, note that the characteristic of F_k that goes through the point $(T_i, X_{i,j})^\top$ and the boundary $x = nt$ intersect at the point $(T_x, nT_x)^\top$, with $T_x := \frac{X_{i,j} - v(k)T_i}{n - v(k)}$. Thus, define

$$(X_{i,j}^\downarrow, T_{i,j}^\downarrow)^\top := \begin{cases} (T_{i-1}, X_{i,j} - v(k)\Delta t)^\top, & \text{if } T_x < T_{i-1}, \\ (T_x, nT_x)^\top, & \text{otherwise}, \end{cases}$$
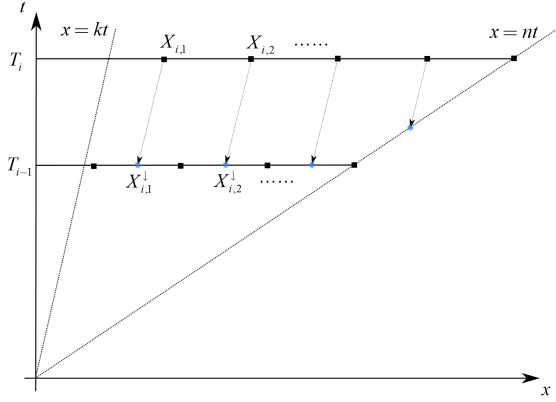
(a) Projection the grid points backwards along the characteristics from time layer T_i to T_{i-1}.
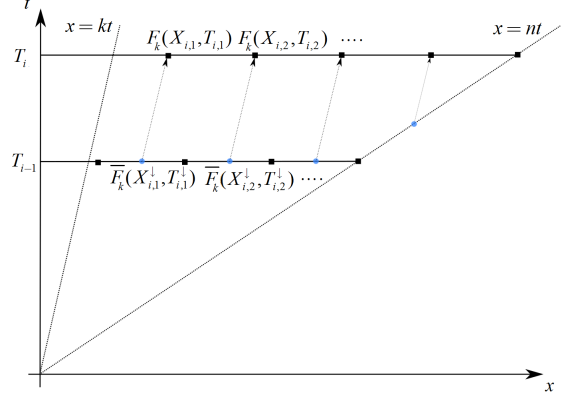
(b) Propagating the interpolated values of the function F_k via numerical integration.

Figure 11: The back-tracing and propagation step of the upstream numerical scheme to compute F_k at all points of the grid.

the projection of $(T_i, X_{i,j})^\top$ back along the corresponding characteristic to the previous time-slice T_{i-1}, or onto the boundary $x = nt$, whichever has the larger t-component. This backward projection step is illustrated in Figure 11(a). Then, according to equation (3.10)

$$F_k(T_i, X_{i,j}) = e^{-(H_k^{(1)}(T_i) - H_k^{(1)}(T_{i,j}^\downarrow))} \left(\int_{T_{i,j}^\downarrow}^{T_i} g_k^{(1)}(\alpha) e^{(H_k^{(1)}(\alpha) - H_k^{(1)}(T_{i,j}^\downarrow))} d\alpha + F_k(X_{i,j}^\downarrow, T_{i,j}^\downarrow) \right) \tag{B.5}$$

holds.

The right-hand side of the equation (B.5) can now be computed using two approximations. Note that the point $(T_{i,j}^\downarrow, X_{i,j}^\downarrow)^\top$ is in general not on the grid X_i, and thus $F_k(T_{i,j}^\downarrow, X_{i,j}^\downarrow)$ has not been pre-computed. If $T_x < T_{i-1}$, the point is equal to $(T_{i-1}, X_{i,j} - v(k)\Delta t)^\top$. In that case, identify the two grid points in X_{i-1} that are closest to $X_{i,j} - v(k)\Delta t$ to the right and ~~to the~~ left. Then let $\bar{F}_k(T_{i,j}^\downarrow, X_{i,j}^\downarrow)$ be the linear interpolation between the values of F_k at those gridpoints. If $T_x \geq T_{i-1}$, then the point is given by $(T_{i,j}^\downarrow, X_{i,j}^\downarrow)^\top = (T_x, nT_x)^\top$ and is located on the boundary. Thus

$$F_k(X_{i,j}^\downarrow, T_{i,j}^\downarrow) = \mathbb{P}\Big\{ A(T_x) = k \Big\},$$

which is also not pre-computed. However, $\mathbb{P}\Big\{ A(T_{i-1}) = k \Big\}$ and $\mathbb{P}\Big\{ A(T_i) = k \Big\}$ have been pre-computed, and thus set $\bar{F}_k(T_{i,j}^\downarrow, X_{i,j}^\downarrow)$ as the linear interpolation between these two values.

The second approximation is to compute the integral on the right-hand side of equation (B.5) using the trapezoidal rule. Thus, the values of F_k on the grid can be computed using

$$F_k(T_i, X_{i,j}) = \frac{\Delta t}{2} \left(g_k^{(1)}(T_i) + e^{-(H_k^{(1)}(T_i) - H_k^{(1)}(T_{i,j}^\downarrow))} g_k^{(1)}(T_{i,j}^\downarrow) \right) + e^{-(H_k^{(1)}(T_i) - H_k^{(1)}(T_{i,j}^\downarrow))} \bar{F}_k(T_{i,j}^\downarrow, X_{i,j}^\downarrow)$$
$$+ o(\Delta t^3) + o(\Delta x^2) \tag{B.6}$$

The terms $g_k(\cdot)$ depend on values of F_ℓ with $k < \ell \leq n$ that might not have been pre-computed on the grid either. However, the same interpolation schemes as for $\bar{F}_k$ can be applied. At this stage it is important though to strictly set F_ℓ to 0 if it should be 0 according to equation (3.7). Lastly, the values for $H_k^{(1)}(\cdot)$ can either be obtained using analytic formulas in equation (3.11) for certain classes of coalescent-speed functions

we will consider (e.g. piece-wise constant), or by computing the requisite integrals using the trapezoidal rule, which can be done incrementally. The integration step of out numerical scheme is illustrated in Figure 11(b).

These equations lead naturally to a dynamic programming algorithm to compute F_k on the specified grid. To this end, iterate through the values $T_i \in T$ in increasing order. For each T_i, iterate through $k \in \{1, 2, \ldots, n\}$ in decreasing order, starting with $k = n$. Then, for each fixed T_i and k, $F_k(T_i, X_{i,j})$ can be computed for every $X_{i,j} \in X_i$ using equations (B.3), (B.4), and (B.6). The order of iteration guarantees that all necessary quantities have been pre-computed. This dynamic program can be employed to compute F_k on the specified grid for all k. Due to Lemma 3.1, the relation

$$\mathbb{P}\{\mathcal{L} \leq X_{M,j}\} = F_1(T_M, X_{M,j})$$

holds, which yields the values of the CDF $\mathbb{P}\{\mathcal{L} \leq x\}$ on the specified grid X_M.

B.2 Upstream Numerical Scheme for Two-Locus Case

In the two-locus case, we can compute F_s efficiently on a chosen grid similar to the marginal case. To this end, we again choose $x_{\max} = y_{\max}$, set $t_{\max} := \frac{1}{n} x_{\max}$, and choose step sizes Δt and $\Delta x = \Delta y$. Then, define the grid T as in definition (B.1), $M = |T|$ and for each T_i, define X_i as in definition (B.2). Furthermore, set $Y_i := X_i$ and $U_i := |Y_i|$. Thus, we use the regular grid $X_i \times Y_i$ in the (x, y)-space.

Now fix T_i and $s \in \bar{\mathcal{S}}^\rho$, and assume that $F_{s'}(T_{i-1}, X_{i-1,j}, Y_{i-1,\ell})$ has been computed for all $s' \in \bar{\mathcal{S}}^\rho$, $X_{i-1,j} \in X_{i-1}$, and $Y_{i-1,\ell} \in Y_{i-1}$. Furthermore, assume that $F_{s'}(T_i, X_{i,j}, Y_{i,\ell})$ has been computed for all s' with $s \prec s'$, $X_{i,j} \in X_i$, and $Y_{i,\ell} \in Y_i$. To compute $F_s(T_i, X_{i,j}, Y_{i,\ell})$, first check using equation (3.13) whether the requisite point lies on the boundary, or in the zero region. The values on the boundary according to equation (3.13) are computed as time-dependent CDFs of marginal integrals along trajectories of the process $\bar{A}^\rho$, and thus they can be computed using exactly the same procedure as detailed in Section B.1, replacing A by $\bar{A}^\rho$. In the interior region, applying the trapezoidal rule to the solution of the first-order ODE, for all $X_{i,j} \in X_i$, and $Y_{i,\ell} \in Y_i$ the value of $F_s(T_i, X_{i,j}, Y_{i,\ell})$ can be computed using

$$
\begin{aligned}
&F_s(T_i, X_{i,j}, Y_{i,\ell}) \\
&= \frac{\Delta t}{2} \left(g_s^{(2)}(T_i) + e^{-(H_s^{(2)}(T_i) - H_s^{(2)}(T_{i,j,\ell}^\downarrow))} g_s^{(2)}(T_{i,j,\ell}^\downarrow) \right) + e^{-(H_s^{(2)}(T_i) - H_s^{(2)}(T_{i,j,\ell}^\downarrow))} \bar{F}_s(T_{i,j,\ell}^\downarrow, X_{i,j}^\downarrow, Y_{i,\ell}^\downarrow) \\
&\quad + o(\Delta t^3) + o(\Delta x^2) + o(\Delta y^2).
\end{aligned}
$$

Here

$$
(T_{i,j,\ell}^\downarrow, X_{i,j}^\downarrow, Y_{i,\ell}^\downarrow)^\top := \begin{cases}
(T_{i-1}, X_{i,j} - v^a(s)\Delta t, Y_{i,\ell} - v^b(s)\Delta t)^\top, & \text{if } \max(T_x, T_y) < T_{i-1}, \\
(T_x, nT_x, Y_{i,\ell} - v^b(s) \cdot (T_i - T_x))^\top, & \text{if } \max(T_{i-1}, T_y) \leq T_x, \\
(T_y, X_{i,j} - v^a(s) \cdot (T_i - T_y), nT_y)^\top, & \text{if } \max(T_x, T_{i-1}) \leq T_y,
\end{cases}
$$

with

$$T_x := \frac{X_{i,j} - v^a(s)T_i}{n - v^a(s)}$$

being the t-coordinate of the point of intersection between the characteristic through the point $(T_i, X_{i,j}, Y_{i,\ell})^\top$ and the boundary $x = nt$, and

$$T_y := \frac{Y_{i,\ell} - v^b(s)T_i}{n - v^a(s)}$$

likewise for the boundary $y = nt$.

The points $(T_{i,j,\ell}^\downarrow, X_{i,j}^\downarrow, Y_{i,\ell}^\downarrow)^\top$ will in general not be on the grid of pre-computed values, and thus the approximation $\bar{F}_s(T_{i,j,\ell}^\downarrow, X_{i,j}^\downarrow, Y_{i,\ell}^\downarrow)$ has to be used. In the case $\max(T_x, T_y) < T_{i-1}$, this value can be obtained by identifying the four points in $X_{i-1} \times Y_{i-1}$ surrounding $(X_{i,j}^\downarrow, Y_{i,\ell}^\downarrow)$, and interpolating the respective values

of $F_s(T_{i-1}, \cdot, \cdot)$ linearly. In the case $\max(T_{i-1}, T_y) \leq T_x$, the point $(T_{i,j,\ell}^{\downarrow}, X_{i,j}^{\downarrow}, Y_{i,\ell}^{\downarrow})^{\top}$ is on the boundary $x = nt$, and

$$F_s(T_{i,j,\ell}^{\downarrow}, X_{i,j}^{\downarrow}, Y_{i,\ell}^{\downarrow}) = \mathbb{P}\{\bar{A}^{\rho}(T_x) = s, L^b(T_x) \leq Y_{i,\ell} - v^b(s) \cdot (T_i - T_x)\}$$

holds. The value of the time-dependent CDF on the right-hand can be obtained as the linear interpolation between the values $\mathbb{P}\{\bar{A}^{\rho}(T_{i-1}) = s, L^b(T_{i-1}) \leq Y_{i,\ell} - v^b(s)\Delta t\}$ and $\mathbb{P}\{\bar{A}^{\rho}(T_i) = s, L^b(T_i) \leq Y_{i,\ell}\}$, which we pre-compute (or approximations thereof) using the numerical scheme for the marginal case (see Appendix B.1) on the boundary. By symmetry, the case $\max(T_x, T_{i-1}) \leq T_y$ can be handled in the same way. Computing $g_s^{(2)}(\cdot)$ will require some $F_{s'}$ with $s \prec s'$, which can be obtained by similar interpolation procedures, or setting it to zero in the appropriate regions. The values of $H_s^{(2)}(\cdot)$ can be computed according to equation (3.16) analytically or numerically, as before.

Again, we can implement these formulas in an efficient dynamic programming algorithm to compute the values of $F_s(t, x, y)$ on the specified grid for all $s \in \bar{\mathcal{S}}^{\rho}$, and thus compute

$$\mathbb{P}\{\mathcal{L}^a \leq X_{M,j}, \mathcal{L}^b \leq Y_{M,\ell}\} = F_{(1,0,0,0)}(t_{\max}, X_{M,j}, Y_{M,\ell}) + F_{(1,0,0,1)}(t_{\max}, X_{M,j}, Y_{M,\ell}),$$

the joint CDF of the total tree length at two linked loci evaluated on the specified grid.

References

Bhaskar, A., Wang, Y. R., and Song, Y. S. (2015). Efficient inference of population size histories and locus-specific mutation rates from large-sample genomic variation data. *Genome Res.*, **25**(2), 268–279.

Dafermos, C. M. (2010). *Hyperbolic conservation laws in continuum physics.* Springer, New York.

Davis, M. H. A. (1993). *Markov Models & Optimization (Chapman & Hall/CRC Monographs on Statistics & Applied Probability).* Chapman and Hall/CRC.

Doob, J. (1953). *Stochastic processes.* Wiley, New York.

Dutheil, J. Y., Ganapathy, G., Hobolth, A., Mailund, T., Uyenoyama, M. K., and Schierup, M. H. (2009). Ancestral population genomics: the coalescent hidden markov model approach. *Genetics*, **183**(1), 259–274.

Eriksson, A., Mahjani, B., and Mehlig, B. (2009). Sequential markov coalescent algorithms for population models with demographic structure. *Theor. Popul. Biol.*, **76**(2), 84–91.

Evans, L. (2010). *Partial Differential Equations.* Graduate studies in mathematics. American Mathematical Society.

Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V. C., and Foll, M. (2013). Robust demographic inference from genomic and snp data. *PLoS Genet.*, **9**(10), e1003905.

Federer, H. (1969). *Geometric measure theory.* Grundlehren der mathematischen Wissenschaften. Springer.

Ferretti, L., Disanto, F., and Wiehe, T. (2013). The effect of single recombination events on coalescent tree height and shape. *PLoS ONE*, **8**(4), 1–15.

Griffiths, R. C. (2003). The frequency spectrum of a mutation, and its age, in a general diffusion model. *Theor. Popul. Biol.*, **64**(2), 241–251.

Griffiths, R. C. and Marjoram, P. (1996). Ancestral inference from samples of dna sequences with recombination. *J. Comp. Biol.*, **3**(4), 479–502.

Griffiths, R. C. and Tavaré, S. (1994). Simulating probability distributions in the coalescent. *Theor. Popul. Biol.*, **46**(2), 131 – 159.

Griffiths, R. C. and Tavaré, S. (1998). The age of a mutation in a general coalescent tree. *Comm. Statist. Stochastic Models*, **14**(1-2), 273–295.

Griffiths, R. C. and Marjoram, P. (1997). An ancestral recombination graph. In *Progress in Population Genetics and Human Evolution*, Donnelly, P. and Tavaré, S., editors, volume 87, pages 257–270. Springer, Berlin.

Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H., and Bustamante, C. D. (2009). Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet.*, **5**, e1000695.

Hobolth, A. and Jensen, J. L. (2014). Markovian approximation to the finite loci coalescent with recombination along multiple sequences. *Theor. Popul. Biol.*, **98**, 48–58.

Hudson, R. R. (1990). Gene genealogies and the coalescent process. *J. Evol. Biol.*, **7**, 1–44.

Hudson, R. R. (2002). Generating samples under a wright–fisher neutral model of genetic variation. *Bioinformatics*, **18**(2), 337–338.

Kamm, J. A., Terhorst, J., and Song, Y. S. (2017). Efficient computation of the joint sample frequency spectra for multiple populations. *J. Comp. Graph. Stat.*, **26**(1), 182–194.

Karlin, S. (1981). *A second course in stochastic processes*. Academic Press, New York.

Keinan, A. and Clark, A. G. (2012). Recent explosive human population growth has resulted in an excess of rare genetic variants. *Science*, **336**(6082), 740–743.

Kingman, J. F. (1982). The coalescent. *J. Evol. Biol.*, **13**(3), 235–248.

Koralov, L. and Sinay, Y. (2007). *Theory of probability and random processes*. Springer, Berlin.

Li, H. and Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. *Nature*, **475**, 493–496.

Marjoram, P. and Wall, J. D. (2006). Fast "coalescent" simulation. *BMC Genet.*, **7**, 16.

McVean, G. A. and Cardin, N. J. (2005). Approximating the coalescent with recombination. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, **360**, 1387–1393.

Paul, J. S., Steinrücken, M., and Song, Y. S. (2011). An accurate sequentially markov conditional sampling distribution for the coalescent with recombination. *Genetics*, **187**(4), 1115–1128.

Pfaffelhuber, P., Wakolbinger, A., and Weisshaupt, H. (2011). The tree length of an evolving coalescent. *Probab. Theory Related Fields*, **151**(3), 529–557.

Polanski, A., Bobrowski, A., and Kimmel, M. (2003). A note on distributions of times to coalescence, under time-dependent population size. *Theor. Popul. Biol.*, **63**(1), 33–40.

Rasmussen, M. D., Hubisz, M. J., Gronau, I., and Siepel, A. (2014). Genome-wide inference of ancestral recombination graphs. *PLoS Genet.*, **10**(5), 1–27.

Renardy, M. and Rogers, R. C. (2004). *An Introduction to Partial Differential Equations*. Springer, 2nd edition.

Rudin, W. (1976). *Principles of Mathematical Analysis*. International series in pure and applied mathematics. McGraw-Hill.

Schiffels, S. and Durbin, R. (2014). Inferring human population size and separation history from multiple genome sequences. *Nat. Genet.*, **46**(8), 919–25.

Sheehan, S., Harris, K., and Song, Y. S. (2013). Estimating variable effective population sizes from multiple genomes: A sequentially markov conditional sampling distribution approach. *Genetics*, **194**(3), 647–662.

Simonsen, K. L. and Churchill, G. A. (1997). A markov chain model of coalescence with recombination. *Theor. Popul. Biol.*, **52**(1), 43–59.

Steinrücken, M., Kamm, J. A., and Song, Y. S. (2015). Inference of complex population histories using whole-genome sequences from multiple populations. *bioRxiv*. Preprint at: http://dx.doi.org/10.1101/026591.

Stroock, D. W. (2008). *An Introduction to Markov Processes (Graduate Texts in Mathematics)*. Springer.

Tavaré, S. and Zeitouni, O. (2004). *Lectures on Probability Theory and Statistics: Ecole d'Eté de Probabilités de Saint-Flour XXXI - 2001 (Lecture Notes in Mathematics)*. Springer.

Terhorst, J., Kamm, J. A., and Song, Y. S. (2017). Robust and scalable inference of population history from hundreds of unphased whole genomes. *Nat. Genet.*, **49**(2), 303–309.

Wakeley, J. (2008). *Coalescent Theory: An Introduction*. W. H. Freeman.

Wiuf, C. and Hein, J. (1999). Recombination as a point process along sequences. *Theor. Popul. Biol.*, **55**, 248–259.

Živković, D. and Wiehe, T. (2008). Second-order moments of segregating sites under variable population size. *Genetics*, **180**(1), 341–357.

Živković, D., Steinrücken, M., Song, Y. S., and Stephan, W. (2015). Transition densities and sample frequency spectra of diffusion processes with selection and variable population size. *Genetics*, **200**(2), 601.