When migration leaves a clean trace: Decoupling migration from coalescence in the Structured Serial Coalescent

Hao Shen[1] and John Novembre[1,2]

[1]Department of Human Genetics, University of Chicago
[2]Department of Ecology and Evolution, University of Chicago

Abstract

With the rapid accumulation of population genomic data across space and time, there is an urgent need for demographic inference methods that incorporate explicit time-series modeling, achieve fine spatial resolution, and ensure clear identifiability between migration and coalescence rates. To address this need, we investigate pairwise genealogical processes under the structured serial coalescent, deriving evolution equations for pairwise branch length distributions and related statistics. By classifying the resulting relationships according to their parameter dependencies and computational complexity, we identify a class that is not only computationally tractable but also determined solely by migration rates. Building on this theoretical basis, we propose an inference framework for fine-resolution, time-varying migration rates inference and demonstrate its feasibility through simulation. We further outline how this framework can be extended to the joint estimation of migration and coalescence rates.

Introduction

Since Kingman introduced the coalescent (Kingman, 1982a,b), it has been widely used in population genetic theory and inferences. However, the standard coalescent theory typically oversimplifies reality by assuming contemporaneous sampling and panmixia, whereas empirical samples can come from different time periods and geographically structured populations with variable lineage migration and coalescence rates. These limitations have motivated key extensions: the serial coalescent for temporal stratification (Rodrigo and Felsenstein, 1999) and the structured coalescent for population subdivision (Notohara, 1990; Takahata, 1991). For models combining both temporal and spatial structure, terminology varies in the literature. Some authors retain "structured coalescent" (Müller et al., 2017) while others use "structured serial coalescent" (Ewing and Rodrigo, 2007). For clarity and distinction, we adopt the latter term throughout this paper.

These extensions of the standard coalescent have enabled powerful demographic inference tools, particularly for estimating effective population sizes (inverse of the instantaneous local coalescence rates) and migration rates. Among these tools, phylogeographic methods have the longest history and remain the most widely used. For the serial coalescent, phylogeographic approaches were developed to estimate effective population sizes from temporally sampled data (Drummond et al., 2002, 2005). For the structured coalescent and structured serial coalescent, phylogeographic methods (phylogeographic methods in a spatial context) have been proposed to jointly infer migration rates and effective population sizes (Beerli and Felsenstein, 2001; Ewing et al., 2004; Volz, 2012; Vaughan et al., 2014; De Maio et al., 2015; Müller et al., 2017, 2018).

However, because these methods rely on information of the full genealogy, their computational scalability with respect to the number of lineages and demes remains severely limited—even with modern algorithmic advances. Consequently, demographic inference for systems with hundreds or thousands of demes remains impractical under these frameworks.

One effective way to solve the scalability problem is to use pairwise coalescence information instead of the whole genealogy. In the context of the structured coalescent, this means developing methods based on pairwise coalescence times. The distribution of pairwise coalescence times depends on both the migration rates and coalescence rates, while also being directly connected to empirical data through pairwise summary statistics. This dual relationship makes pairwise coalescence times a powerful tool for demographic inference.

For example, EEMS (Petkova et al., 2016), FEEMS (Marcus et al., 2021), and FRAME (Shen and Novembre, 2025) all exploit the connection between expected pairwise coalescence times and the sample covariance matrix (McVean, 2009). Meanwhile, MAPS (Al-Asadi et al., 2019) uses the fact that the length distribution of long pairwise shared coalescent segments (LPSC segments, also know as the identity-by-descent or IBD tracts) is determined by the distribution of pairwise coalescence times (Palamara et al., 2012; Carmi et al., 2013). All of these methods are scalable to hundreds of demes, achieving resolutions that are impossible with traditional phylogeographic methods. While earlier methods like EEMS, FEEMS, and MAPS relied heavily on the assumption of symmetric migration, recent computational advances in FRAME have largely solved this problem. These improvements now enable fine-resolution inference even for cases with highly asymmetric gene flow.

Despite these methodological advances, several critical limitations persist. First, research across many domains requires explicit time-series modeling to accommodate serial sampling and time-varying migration rates. In conservation genetics, samples collected at multiple time points can reveal how gene flow patterns change through time. In epidemiology, serially sampled viral genomes can be used to reconstruct transmission dynamics. In human evolutionary biology, ancient DNA provides genetic snapshots from different epochs, enabling reconstruction of migration histories within each period. FRAME, however, assumes migration–drift equilibrium under the structured coalescent and therefore cannot handle serial sampling or time-varying parameters. While it is possible to analyze data by dividing it into temporal slices and fitting each slice with an equilibrium model (as done in MAPS and FRAME), it remains suboptimal compared to direct modeling under the structured serial coalescent framework. Second, these methods can have difficulty in distinguish the effects of migration rates and coalescence rates, leading to identifiability issues. (MAPS is an exception, but its effectiveness in asymmetric migration scenarios remains uncertain.)

These limitations motivate the need for a new demographic inference method that can: (1) preserve the scalability advantages of pairwise coalescence-based approaches, (2) directly incorporate serial sampling and time-varying parameters under the structured serial coalescent framework, and (3) effectively decouple the inference of migration rates from that of coalescence rates. While such a method might appear overly ambitious, in this paper we demonstrate that it is indeed theoretically achievable.

Unlike the structured coalescent, where the equation for expected pairwise coalescence times in the structured coalescent was established (Strobeck, 1987) even before the theory's formal development, the analogous theory for pairwise branch lengths in the structured serial coalescent - defined as the summed branch lengths from two sampled lineages to their most recent common ancestor (the serial sampling counterpart to pairwise coalescence times) - remains largely undeveloped. To bridge this theoretical gap and facilitate new inference methods, we develop the foundational theory for pairwise branch length dynamics under the structured serial coalescent framework.

Specifically, we derive evolution equations governing the probability density functions (PDFs) of pairwise branch lengths. By solving these equations, we uncover fundamental relationships among the PDFs and systematically classify them into distinct categories based on their parameter dependencies and computational complexity. This classification framework extends naturally to other quantities derived from pairwise branch lengths, as we demonstrate through parallel analysis of mean pairwise branch length dynamics and length distribution dynamics of LPSC segments.

The general applicability of this classification enables a unified evaluation of each relationship class's potential for demographic inference. Crucially, we identify the a promising class that exhibit precisely the properties needed to realize our proposed ideal inference method: exclusive dependence on lineage migration rates with no influence from coalescence parameters, and computational tractability scaling as $O(d^3)$ where d represents the number of demes. These properties directly address the fundamental challenges identified earlier - they resolve the migration-coalescence identifiability problem while maintaining the scalability required for fine-resolution analysis.

This theoretical breakthrough establishes a foundation for two inference approaches. First, for studies focusing specifically on gene flow dynamics, these relationships enable highly scalable migration rate estimation that is decoupled from coalescence parameters. Second, joint estimation of migration and coalescence rates may also benefit from these relationships. A two-step approach—first estimating migration rates using these classes, then inferring coalescence rates conditioned on the migration parameters—could improve both identifiability and scalability, though the latter gain may be less pronounced than in migration-only inference. We outline inference frameworks for both approaches through a proof-of-concept example and demonstrate the feasibility of the first approach through preliminary simulation studies.

The model

The structured serial coalescent process models the combined effects of backward migration between demes and coalescence events occurring when lineages reside in the same deme. We represent backward migration as a continuous-time jump process on a weighted directed graph with nodes $\{1, 2, \ldots, d\}$, each corresponding to a deme. The weight on the edge from node i to node j is denoted $m_{ij}(t)$, representing the backward migration rate from i to j at time t. Biologically, $m_{ij}(t)$ is the rate at which deme i receives ancestry from deme j. The transition rate matrix of this jump process is $Q(t)$, and the associated Laplacian is defined as $L(t) = -Q(t)$. The migration rates are encoded in the weighted adjacency matrix $M(t) = Q(t) - \operatorname{diag} Q(t)$. The coalescence rate in deme i is denoted by $\gamma_i(t)$, and all coalescence rates are collected into the vector $\gamma(t) = (\gamma_i(t))$.

In our model, we study the pairwise genealogical process using the pairwise branch length, defined as the sum of the branch lengths of the two lineages up to their MRCA. This measure generalizes pairwise coalescence time by explicitly accounting for temporal offsets between samples and can be viewed as the distance of the two lineages in terms of branch length. For an illustration of how pairwise branch length is computed, see Figure 1.

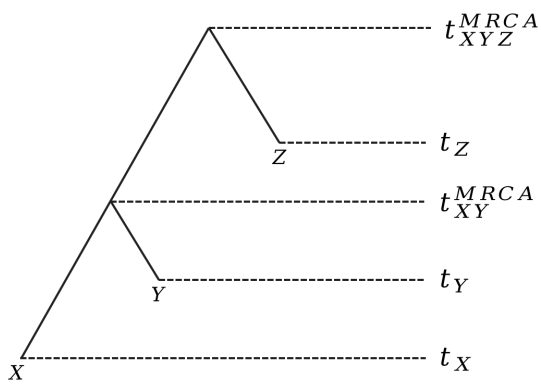

Fig. 1. Pairwise branch length. A toy example showing how pairwise branch lengths are computed. Assume we have three samples X, Y, and Z sampled at times t_X, t_Y, and t_Z, respectively. The pairwise branch length between X and Y is $(t_{XY}^{MRCA} - t_X) + (t_{XY}^{MRCA} - t_Y)$. Similarly, the pairwise branch length between X and Z is $(t_{XYZ}^{MRCA} - t_X) + (t_{XYZ}^{MRCA} - t_Z)$, and between Y and Z is $(t_{XYZ}^{MRCA} - t_Y) + (t_{XYZ}^{MRCA} - t_Z)$.

Now let $B_{ij}^{x,y}$ denote the random variable representing the pairwise branch length between two lineages, one sampled from deme i at time x and the other from deme j at time y, where $x \leq y$ and time is measured backward from the present ($t = 0$). Let $F_{ij}^{x,y}(b) = F_{ij}^{x,y,b} = P(B_{ij}^{x,y} \leq b)$ be the corresponding cumulative distribution function ; $f_{ij}^{x,y}(b) = f_{ij}^{x,y,b} = \frac{\partial F_{ij}^{x,y,b}}{\partial b}$ be its probability density function; $f^{x,y}(b) = f^{x,y,b} = (f_{ij}^{x,y,b})$ be the matrix of the probability density functions (some times we will neglect b and just write $f^{x,y}$ to refer to the function). All the above functions have support in $[y - x, +\infty)$ and will be 0 for $b < y - x$.

In the next three sections, we will explore the relationships among the probability density functions as well as two related statistics (expected pairwise branch length and survival function of LPSC segment length).

Relationships among probability density functions of pairwise branch length

We consider a toy example with three epochs: $[t_0, t_1), [t_1, t_2)$, and $[t_2, \infty)$, which we label as epoch 0, epoch 1, and epoch 2, respectively. The migration and coalescence rates are assumed to be piecewise constant in the first two epoches, that is, $L(x) = L^0, \gamma(x) = \gamma^0$ for $x \in [t_0, t_1)$ and $L(x) = L^1, \gamma(x) = \gamma^1$ for $x \in [t_1, t_2)$ (The migration and coalescence rates in the last epoch are not specified, as they are not used in this section). We now investigate relationships among the following time-ordered pairs of pairwise branch length pdf matrices: $(f^{t_0,t_1}, f^{t_1,t_1})$;$(f^{t_0,t_2}, f^{t_1,t_2})$; $(f^{t_1,t_2}, f^{t_2,t_2})$; $(f^{t_0,t_0}, f^{t_1,t_1})$; $(f^{t_1,t_1}, f^{t_2,t_2})$; $(f^{t_0,t_0}, f^{t_0,t_1})$; $(f^{t_1,t_1}, f^{t_1,t_2})$;$(f^{t_0,t_1}, f^{t_0,t_2})$. By abuse of notation, we will also use these pairs to denote the corresponding relationships. These 8 relationships correspond to the 8 edges in Fig. 2.

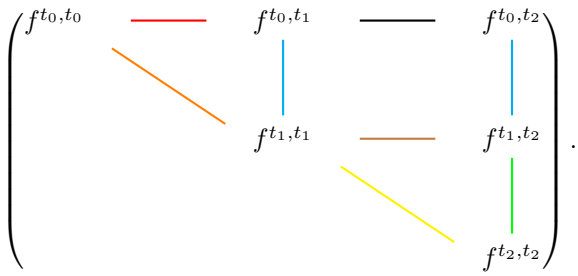

Fig. 2. Graphical representation of the relationships among pairwise branch length probability density functions

Graphically, these 8 edges in Fig. 2 fall naturally into three categories: the vertical edges, the diagonal edges and the horizontal edges. In what follows, we show that this graphical classification corresponds to a classification of the relationships themselves in terms of representational and computational complexity.

We first examine the relationship $(f^{t_0,t_1}, f^{t_1,t_1})$, which is represented by a vertical edge. To this end, we study the evolution of $f^{x,t_1,b}$ for $x \in [t_0, t_1)$. As we move from x to $x + \Delta t$, coalescence does not occur, and only migration of the first lineage is relevant. This yields the following equation

$$f_{ij}^{x,t_1,b} = \sum_{k \neq i} m_{ik}(x)\Delta t f_{kj}^{x+\Delta t,t_1,b-\Delta t} + \left(1 - \sum_{k \neq i} m_{ik}(x)\Delta t\right) f_{ij}^{x+\Delta t,t_1,b-\Delta t}. \tag{1}$$

Expanding $f_{\alpha j}^{x+\Delta t,t_1,b-\Delta t}$ as $f_{\alpha j}^{x,t_1,b} + \frac{\partial f_{\alpha j}^{x,t_1,b}}{\partial x}\Delta t - \frac{\partial f_{\alpha j}^{x,t_1,b}}{\partial b}\Delta t$ and ignoring higher-order terms, we cancel Δt on both sides and obtain

$$\frac{\partial f_{ij}^{x,t_1,b}}{\partial x} - \frac{\partial f_{ij}^{x,t_1,b}}{\partial b} = \left(\sum_{k \neq i} m_{ik}(x)\right) f_{ij}^{x,t_1,b} - \sum_{k \neq i} m_{ik}(x) f_{kj}^{x,t_1,b}. \tag{2}$$

In matrix form, this partial differential equation becomes

$$\frac{\partial f^{x,t_1,b}}{\partial x} - \frac{\partial f^{x,t_1,b}}{\partial b} = L(x) f^{x,t_1,b} = L^0 f^{x,t_1,b}. \tag{3}$$

with the boundary condition of known $f^{t_1,t_1,b}$. This PDE has the explicit solution

$$f^{x,t_1,b} = e^{-L^0(t_1-x)} f^{t_1,t_1,b-(t_1-x)}. \tag{4}$$

Taking the limit as $x \to t_0$, we obtain the relationship $(f^{t_0,t_1}, f^{t_1,t_1})$. Following similar logic, the relationships $(f^{t_0,t_2}, f^{t_1,t_2})$ and $(f^{t_1,t_2}, f^{t_2,t_2})$, which are likewise represented by vertical edges, can also be computed. These three relationships can be jointly expressed as

$$f^{t_0,t_1,b} = e^{-L^0(t_1-t_0)} f^{t_1,t_1,b-(t_1-t_0)}, \tag{5a}$$

$$f^{t_0,t_2,b} = e^{-L^0(t_1-t_0)} f^{t_1,t_2,b-(t_1-t_0)}, \tag{5b}$$

$$f^{t_1,t_2,b} = e^{-L^1(t_2-t_1)} f^{t_2,t_2,\,b-(t_2-t_1)}. \tag{5c}$$

These relationships are particularly valuable for two reasons. First, they depend only on backward migration rates—$(f^{t_0,t_1}, f^{t_1,t_1})$ and $(f^{t_0,t_2}, f^{t_1,t_2})$ depend on L^0 (we represent the corresponding edges in blue in Fig. 2), while $(f^{t_1,t_2}, f^{t_2,t_2})$ depends on L^1. This makes them a promising tool for disentangling migration dynamics from coalescence effects (or effective population sizes) in complex migration–drift scenarios. Second, the computational complexity of these relationships is at most $O(d^3)$, and can often be reduced further by exploiting the (potentially) sparse structure of the Laplacian matrices. This scalability makes it possible to perform gene flow or migration inference at an unprecedented fine resolution, involving hundreds or even thousands of demes. As we shall see, the class of relationships represented by the vertical edges is the only one that possesses these favorable properties, and should therefore be fully exploited to facilitate inference.

Now we turn to the relationships $(f^{t_0,t_0}, f^{t_1,t_1})$, which corresponds to a diagonal edge in Figure.2. Let $f^{x,b} = f^{x,x,b}$ denote the density matrix when the two lineages are sampled at the same time. For $x \in [t_0, t_1)$, the associated partial differential equation is

$$\frac{\partial f^{x,b}}{\partial x} - 2\frac{\partial f^{x,b}}{\partial b} = \mathrm{diag}\{\gamma^0\}\,\mathrm{diag}\{f^{x,b}\} + L^0 f^{x,b} + f^{x,b}(L^0)^T, \tag{6}$$

with boundary condition of known $f^{t_1,b}$ and $f^{x,0} = \mathrm{diag}\{\gamma^0\}$ for $x \in [t_0, t_1)$. This equation is derived in a manner similar to equation (3), with the key difference being that both lineages can migrate independently, and coalescence may occur when they are in the same deme. The PDE remains linear once we vectorize both sides.

Let E_i denote the matrix with a 1 in the (i,i)-th entry and zeros elsewhere, and let $\epsilon_i = \mathrm{vec}(E_i)$ be its vectorization. Define $S^0 = L^0 \otimes I + I \otimes L^0 + \sum_{i=1}^{d} \gamma_i^0 \epsilon_i \epsilon_i^T$, where $\otimes$ denotes the Kronecker product, define and $G^0 = \sum_{i=1}^{d} \gamma_i^0 \epsilon_i$. Then the solution to equation (6) is given by

$$\mathrm{vec}(f^{x,b}) = \begin{cases} e^{-\frac{1}{2}S^0 b} G^0 & b \in [0, 2(t_1 - x)) \\ e^{-S^0(t_1 - x)}\,\mathrm{vec}(f^{t_1, b-2(t_1-x)}) & b \in [2(t_1 - x), +\infty) \end{cases} \tag{7}$$

Taking the limit as $x \to t_0$, we obtain the relationship $(f^{t_0,t_0}, f^{t_1,t_1})$. A similar analysis applies to the relationship $(f^{t_1,t_1}, f^{t_2,t_2})$, which is represented by another diagonal edge. Together, the relationships represented by the diagonal edges can be written jointly as

$$\mathrm{vec}(f^{t_0,b}) = \begin{cases} e^{-\frac{1}{2}S^0 b} G^0 & b \in [0, 2(t_1 - t_0)) \\ e^{-S^0(t_1 - t_0)}\,\mathrm{vec}(f^{t_1, b-2(t_1-t_0)}) & b \in [2(t_1 - t_0), +\infty) \end{cases} \tag{8a}$$

$$\mathrm{vec}(f^{t_1,b}) = \begin{cases} e^{-\frac{1}{2}S^1 b} G^1 & b \in [0, 2(t_2 - t_1)) \\ e^{-S^1(t_2 - t_1)}\,\mathrm{vec}(f^{t_2, b-2(t_2-t_1)}) & b \in [2(t_2 - t_1) + \infty) \end{cases} \tag{8b}$$

where $S^1 = L^1 \otimes I + I \otimes L^1 + \sum_{i=1}^{d} \gamma_i^1 \epsilon_i \epsilon_i^T, G^1 = \sum_{i=1}^{d} \gamma_i^1 \epsilon_i$.

Before turning to the relationships represented by the horizontal edges, we would like to note two points. First, when the two lineages evolve together in $f^{x,x,b}$, the model simplifies to a structured coalescent model. The probability density function of the pairwise coalescence times can be derived using the fact that the pairwise branch lengths are twice the pairwise coalescence times in the structured coalescent. Second, as t_2 goes to infinity in equation (8b, we obtain the stationary distribution of pairwise branch length under a structured coalescent model with migration rates coded by L^1 and coalescence rates coded by γ^1. The stationary distribution of pairwise branch length distribution satisfies

$$\mathrm{vec}(f^{1*,b}) = e^{-\frac{1}{2}S^1 b} G^1, \tag{9}$$

where $f^{1*,b} = \lim_{t_2 \to +\infty} f^{t_1,b}$.

Because S^0 and S^1 depend on both migration and coalescence rates, the relationships represented by the diagonal edges are generally more complex and computationally intensive than those represented by the vertical edges. However, as we shall see next, they remain relatively straightforward to derive and interpret compared with the relationships represented by the horizontal edges, and thus serve as useful candidates for inferring coalescence rates or effective population sizes.

For the relationships represented by the horizontal edges, namely $(f^{t_0,t_0}, f^{t_0,t_1})$, $(f^{t_1,t_1}, f^{t_1,t_2})$, and $(f^{t_0,t_1}, f^{t_0,t_2})$, it is not easy to directly formulate and solve the corresponding PDEs. Since we have already established the relationships corresponding to the vertical and diagonal edges, we derive those for the

horizontal edges indirectly (for details of the derivation, see Supplementary Information A), which yields

$$\mathrm{vec}(f^{t_0,t_0,b}) = \begin{cases} e^{-\frac{1}{2}S^0 b} G^0 & b \in [0, 2(t_1 - t_0)) \\ e^{-S^0(t_1-t_0)} \mathrm{vec}(e^{L^0(t_1-t_0)} f^{t_0,t_1,b-(t_1-t_0)}) & b \in [2(t_1 - t_0), +\infty) \end{cases} \tag{10a}$$

$$\mathrm{vec}(f^{t_1,t_1,b}) = \begin{cases} e^{-\frac{1}{2}S^1 b} G^1 & b \in [0, 2(t_2 - t_1)) \\ e^{-S^1(t_2-t_1)} \mathrm{vec}(e^{L^1(t_2-t_1)} f^{t_1,t_2,b-(t_2-t_1)}) & b \in [2(t_2 - t_1), +\infty) \end{cases} \tag{10b}$$

$$\begin{aligned} &\mathrm{vec}(e^{L^0(t_1-t_0)} f^{t_0,t_1,b}) \\ &= \begin{cases} e^{-\frac{1}{2}S^1[b-(t_1-t_0)]} G^1 & b \in [t_1 - t_0, 2t_2 - t_1 - t_0) \\ e^{-S^1(t_2-t_1)} \mathrm{vec}(e^{L^1(t_2-t_1)} e^{L^0(t_1-t_0)} f^{t_0,t_2,b-(t_2-t_1)}) & b \in [2t_2 - t_1 - t_0, +\infty) \end{cases} \end{aligned} \tag{10c}$$

These equations are highly complex both in terms of representation and computation. The difficulty arises from the fact that when we trace lineage 2 in $f^{x,y,b}$ along y, the process is not purely migrational: we must also account for the location of lineage 1 at the same time point. If lineage 1 happens to be in the same deme as lineage 2, there is a positive probability that they coalesce. Moreover, for lineage 1 to evolve from time x to time y, it must pass through all changes in the migration rates during the interval $[x, y]$. This is why the relationship $(f^{t_0,t_1}, f^{t_0,t_2})$, represented by the black horizontal edge in Fig. 2, is particularly intricate, as it depends simultaneously on L^0, L^1, and γ^1.

Since all relationships represented by the horizontal edges can be effectively derived from those represented by the vertical and diagonal edges, and are typically much more complicated than the other two categories, it is important to avoid using them for inference. In a demographic inference framework based on pairwise branch lengths, one should rely only on relationships represented by vertical edges when estimating backward migration rates, and use both vertical and diagonal edges for the joint inference of migration rates and coalescence rates.

Relationships among expected pairwise branch lengths

The relationships among the probability density functions (pdfs) of pairwise branch lengths naturally extend to summary statistics derived from them. For example, consider the expected pairwise branch length $\overline{B^{x,y}}$. In the structured coalescent framework, the analogous quantity - expected pairwise coalescence times - are directly related to sample covariance structure (McVean, 2009), a connection exploited by methods including EEMS (Petkova et al., 2016), FEEMS (Marcus et al., 2021), and FRAME (Shen and Novembre, 2025). As we demonstrate in Supplementary Information D, such connection can extend to serially sampled data with slight modifications. Furthermore, advances in tree sequence reconstruction now allow for the direct estimation of $\overline{B^{x,y}}$ from inferred branch lengths, though this requires careful consideration of the demographic assumptions inherent in the reconstruction process. Here, we establish the relationships among expected pairwise branch lengths in the structured serial coalescent.

Similar to last section, we study a three-epoch toy model. The 8 relationships we would like to study are $(\overline{B^{t_0,t_1}}, \overline{B^{t_1,t_1}})$; $(\overline{B^{t_0,t_2}}, \overline{B^{t_1,t_2}})$; $(\overline{B^{t_1,t_2}}, \overline{B^{t_2,t_2}})$; $(\overline{B^{t_0,t_0}}, \overline{B^{t_1,t_1}})$; $(\overline{B^{t_1,t_1}}, \overline{B^{t_2,t_2}})$; $(\overline{B^{t_0,t_0}}, \overline{B^{t_0,t_1}})$; $(\overline{B^{t_1,t_1}}, \overline{B^{t_1,t_2}})$; $(\overline{B^{t_0,t_1}}, \overline{B^{t_0,t_2}})$. These relationships corresponds to edges in Fig. 3.

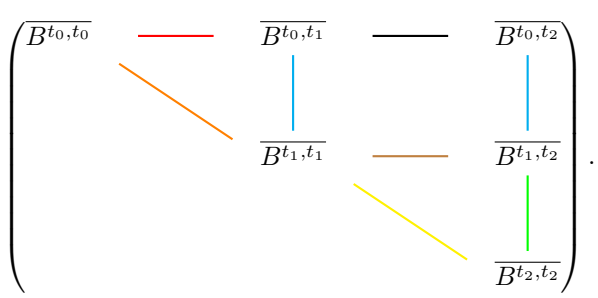

Fig. 3. Graphical representation of the relationships among expected pairwise branch lengths

We now study $(\overline{B^{t_0,t_1}}, \overline{B^{t_1,t_1}})$. When $x \in [t_0, t_1)$, this expectation satisfies the partial differential equation

$$\frac{\partial \overline{B^{x,t_1}}}{\partial x} = L^0 \overline{B^{x,t_1}} - 1_{d\times d}, \tag{11}$$

with boundary condition of known $\overline{B^{t_1,t_1}}$. Using the identities $L1_{d\times d} = 0$ and $e^{L^0\alpha}1_{d\times d} = 1_{d\times d}$ for any α, the solution is given by

$$\overline{B^{x,t_1}} = e^{-L^0(t_1-x)}\overline{B^{t_1,t_1}} + (t_1 - x)1_{d\times d}. \tag{12}$$

Taking the limit as $x \to t_0$, we get the relationship $(\overline{B^{t_0,t_1}}, \overline{B^{t_1,t_1}})$. Following the same logic, the relationships represented by other vertical edges can also be computed. These relationships can be jointly written as

$$\overline{B^{t_0,t_1}} = e^{-L^0(t_1-t_0)}\overline{B^{t_1,t_1}} + (t_1 - t_0)1_{d\times d}, \tag{13a}$$

$$\overline{B^{t_0,t_2}} = e^{-L^0(t_1-t_0)}\overline{B^{t_1,t_2}} + (t_1 - t_0)1_{d\times d}, \tag{13b}$$

$$\overline{B^{t_1,t_2}} = e^{-L^1(t_2-t_1)}\overline{B^{t_2,t_2}} + (t_2 - t_1)1_{d\times d}. \tag{13c}$$

The relationships corresponding to the diagonal edges can also be derived directly. Let $\overline{B^x} = \overline{B^{x,x}} = \int_0^{+\infty} b f^{x,b}\, db$ denote the matrix of expected pairwise branch lengths when both lineages are sampled at time x. For $x \in [t_0, t_1)$, this satisfies the ordinary differential equation

$$\frac{d\overline{B^x}}{dx} = \mathrm{diag}\{\gamma\}\,\mathrm{diag}\{\overline{B^x}\} + L^0\overline{B^x} + \overline{B^x}(L^0)^T - 2_{d\times d}. \tag{14}$$

Vectorizing and solving this equation gives

$$\mathrm{vec}(\overline{B^x} - \overline{B^{0*}}) = e^{-S^0(t_1-x)}\,\mathrm{vec}(\overline{B^{t_1}} - \overline{B^{0*}}), \tag{15}$$

where $\overline{B^{0*}}$ is the equilibrium expected pairwise branch length under the parameter set (L^0, γ^0) and satisfies the equation

$$\mathrm{diag}\{\gamma^0\}\,\mathrm{diag}\{\overline{B^{0*}}\} + L^0\overline{B^{0*}} + \overline{B^{0*}}(L^0)^T = 2_{d\times d}. \tag{16}$$

Now letting $x \to t_0$ and applying the same logic to the epoch $[t_1, t_2)$, we obtain the joint equation for the relationships represented by the diagonal edges

$$\mathrm{vec}(\overline{B^{t_0}} - \overline{B^{0*}}) = e^{-S^0(t_1-t_0)}\,\mathrm{vec}(\overline{B^{t_1}} - \overline{B^{0*}}), \tag{17a}$$

$$\mathrm{vec}(\overline{B^{t_1}} - \overline{B^{1*}}) = e^{-S^1(t_2-t_1)}\,\mathrm{vec}(\overline{B^{t_2}} - \overline{B^{1*}}), \tag{17b}$$

where $\overline{B^{1*}}$ satisfies

$$\mathrm{diag}\{\gamma^1\}\,\mathrm{diag}\{\overline{B^{1*}}\} + L^1\overline{B^{1*}} + \overline{B^{1*}}(L^1)^T = 2_{d\times d}. \tag{18}$$

The relationships corresponding to the horizontal edges can also be derived indirectly (for a detailed derivation, see Supplementary Information B), and can be written as

$$\mathrm{vec}(\overline{B^{t_0,t_0}} - \overline{B^{0*}}) = e^{-S^0(t_1-t_0)}\,\mathrm{vec}(e^{L^0(t_1-t_0)}\overline{B^{t_0,t_1}} - (t_1 - t_0)1_{d\times d} - \overline{B^{0*}}), \tag{19a}$$

$$\mathrm{vec}(\overline{B^{t_1,t_1}} - \overline{B^{1*}}) = e^{-S^1(t_2-t_1)}\,\mathrm{vec}(e^{L^1(t_2-t_1)}\overline{B^{t_0,t_1}} - (t_2 - t_1)1_{d\times d} - \overline{B^{1*}}), \tag{19b}$$

$$\begin{aligned}&\mathrm{vec}(e^{L^0(t_1-t_0)}\overline{B^{t_0,t_1}} - (t_1 - t_0)1_{d\times d} - \overline{B^{1*}})\\ &= e^{-S^1(t_2-t_1)}\,\mathrm{vec}(e^{L^1(t_2-t_1)}e^{L^0(t_1-t_0)}\overline{B^{t_0,t_2}} - (t_2 - t_0)1_{d\times d} - \overline{B^{1*}}).\end{aligned} \tag{19c}$$

As expected, the relationships corresponding to the vertical edges depend only on migration rates and are relatively easy to compute, whereas those corresponding to the horizontal edges are the most complex in terms of both representation and computation.

Relationships among survival functions of LPSC segment lengths

In classical coalescent as well structured coalescent, it is well known that the length distribution of long pairwise shared coalescence segments (LPSC segments, or identity-by-descent tracts) is determined by the distribution of pairwise coalescence time (Palamara et al., 2012; Carmi et al., 2013) and can thus be used for demographic inference. This is the idea underlying MAPS (Al-Asadi et al., 2019), which infers the migration surfaces and coalescence rates based on the LPSC length distribution. In structured serial coalescent, following the logic of the previous two sections, we establish the relationships among the survival functions of the LPSC segment lengths.

Let $\rho^{x,y,\mu}$ be the matrix of survival functions such that it's $(i,j)_{\text{th}}$ element is the probability that a LPSC segment between two randomly sampled lineages from deme i at time x and deme j at time y has length larger than μ. The probability of a random LPSC segment has length larger than μ given the pairwise branch length b and recombination rate r is simply e^{-rbu}. And we obtain

$$\rho^{x,y,\mu} = \int_0^{+\infty} e^{-rb\mu} f^{x,y,b} db. \tag{20}$$

Here again we employ the three epoch toy model, and we would like to investigate the 8 relationships $(\rho^{t_0,t_1,\mu}, \rho^{t_1,t_1,\mu})$; $(\rho^{t_0,t_2,\mu}, \rho^{t_1,t_2,\mu})$; $(\rho^{t_1,t_2,\mu}, \rho^{t_2,t_2,\mu})$; $(\rho^{t_0,t_0,\mu}, \rho^{t_1,t_1,\mu})$; $(\rho^{t_1,t_1,\mu}, \rho^{t_2,t_2,\mu})$; $(\rho^{t_0,t_0,\mu}, \rho^{t_0,t_1,\mu})$; $(\rho^{t_1,t_1,\mu}, \rho^{t_1,t_2,\mu})$; $(\rho^{t_0,t_1,\mu}, \rho^{t_0,t_2,\mu})$. Again, these 8 relationships correspond to 8 edges in Fig. 4.

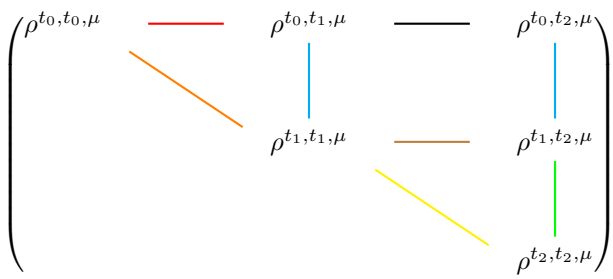

Fig. 4. Graphical representation of the relationships among survival functions of LPSC segment lengths

The easiest way to study the relationships represented by the vertical edges here to directly write out the integration out and use the relationships among the probability density functions, instead of solving the PDEs. Writing $L^0 + r\mu I, L^1 + r\mu I$ as L^0_μ, L^1_μ and using equation (5a), (5b) and (5c) yields

$$\rho^{t_0,t_1,\mu} = e^{-L^0_\mu (t_1-t_0)} \rho^{t_1,t_1,\mu}, \tag{21a}$$

$$\rho^{t_0,t_2,\mu} = e^{-L^0_\mu (t_1-t_0)} \rho^{t_1,t_2,\mu}, \tag{21b}$$

$$\rho^{t_1,t_2,\mu} = e^{-L^1_\mu (t_2-t_1)} \rho^{t_2,t_2,\mu}. \tag{21c}$$

For the relationships represented by the diagonal edges, we can also derive analogous expressions. Applying equation (8a), (8b), with $\rho^{x,x,\mu} = \rho^{x,\mu}$, $S^0 + 2r\mu I = S^0_\mu$, $S^1 + 2r\mu I = S^1_\mu$, we obtain

$$\text{vec}(\rho^{t_0,\mu}) = 2[I - e^{-S^0_\mu (t_1-t_0)}](S^0_\mu)^{-1} G^0 + e^{-S^0_\mu (t_1-t_0)} \text{vec}(\rho^{t_1,\mu}), \tag{22a}$$

$$\text{vec}(\rho^{t_1,\mu}) = 2[I - e^{-S^1_\mu (t_2-t_1)}](S^1_\mu)^{-1} G^1 + e^{-S^1_\mu (t_2-t_1)} \text{vec}(\rho^{t_2,\mu}). \tag{22b}$$

Finally the relationships corresponding to the horizontal edges can also be derived indirectly, which are given by

$$\text{vec}(\rho^{t_0,t_0,\mu}) = 2[I - e^{-S^0_\mu (t_1-t_0)}](S^0_\mu)^{-1} G^0 + e^{-S^0_\mu (t_1-t_0)} \text{vec}(e^{L^0_\mu (t_1-t_0)} \rho^{t_0,t_1,\mu}), \tag{23a}$$

$$\text{vec}(\rho^{t_1,t_1,\mu}) = 2[I - e^{-S^1_\mu (t_2-t_1)}](S^1_\mu)^{-1} G^1 + e^{-S^1_\mu (t_2-t_1)} \text{vec}(e^{L^1_\mu (t_2-t_1)} \rho^{t_1,t_2,\mu}), \tag{23b}$$

$$\begin{aligned} &e^{L^0_\mu (t_1-t_0)} \text{vec}(\rho^{t_0,t_1,\mu}) \\ &= 2[I - e^{-S^1_\mu (t_2-t_1)}](S^1_\mu)^{-1} G^1 + e^{-S^1_\mu (t_2-t_1)} \text{vec}(e^{L^1_\mu (t_2-t_1)} e^{L^0_\mu (t_1-t_0)} \rho^{t_0,t_2,\mu}). \end{aligned} \tag{23c}$$

A step-by-step derivation of the equations presented in this section can be found in the Supplementary Information C. Once again, we observe that the relationships corresponding to the vertical edges are the most well-behaved, while those corresponding to the horizontal edges are the most complex in terms of both representation and computation.

Inference framework

Here we study how the above relationships can be used to establish inference frameworks. Similar to the previous sections, we set up a proof-of-concept example by assuming that samples are available at three time points, t_0, t_1, and t_2, and that the goal is to infer the 'effective' demographic parameters in $[t_0, t_1)$ and $[t_1, t_2)$ using the relationships among the expected pairwise coalescence times, i.e., the $\overline{B^{t_i,t_j}}$ matrices.

Although the $\overline{B^{t_i,t_j}}$ matrices are not directly observable, they are naturally connected to sample variance structures and reconstructed tree sequences, as discussed in previous sections. The protocols below do not detail the establishment of these connections but instead present a higher-level inference framework. For convenience, we refer to sample variance structures and reconstructed tree sequences collectively as 'corresponding data', and on this basis propose two demographic inference approaches: one specialized for inferring migration rates, and another designed for the joint estimation of migration and coalescence rates.

Pure migration rates inference

For pure migration rates inference in the proof-of concept example, we follow a simple three-step protocol:

1. Estimate the $\overline{B^{t_2,t_2}}$ matrix from the corresponding data. This estimation can be straightforward if genetic samples are available from every deme. However, for cases involving vacant demes (those without samples), imputation becomes necessary. One strategy is to use spatial imputation, which leverages geographical proximity to infer values for the vacant demes based on the directly estimated submatrix of $\overline{B^{t_2,t_2}}$ that represents the demes with samples. Alternatively, a model-based imputation approach can be employed. This method operates under the assumption that the stochastic process governing lineages before time t_2 was driven by a set of equilibrium migration rates (L^∞) and coalescence rates (γ^∞). In other words, it assumes $\overline{B^{t_2,t_2}} = \overline{B^\infty}$, where the equilibrium matrix is defined by the equation:

$$\operatorname{diag}\gamma^\infty \operatorname{diag}\overline{B^\infty} + L^\infty\overline{B^\infty} + \overline{B^\infty}(L^\infty)^T = 2_{d\times d}. \tag{24}$$

Under this assumption, established equilibrium-based methods, such as FRAME (Shen and Novembre, 2025), can be used to first infer the parameters L^∞ and γ^∞, which in turn provides an estimate for the complete $\overline{B^{t_2,t_2}}$ matrix.

2. Estimate $\overline{B^{t_1,t_2}}$ and L^1 based on corresponding data and estimated $\overline{B^{t_2,t_2}}$ using equation (13c). From equation (13c), we know $\overline{B^{t_1,t_2}}$ is a function of $\overline{B^{t_2,t_2}}$ and L^1, since in the first step we already have the inferred $\overline{B^{t_2,t_2}}$, we can write $\overline{B^{t_1,t_2}}$ solely as a function of L^1. Consequently, a connection between L^1 and the corresponding data is mediated through $\overline{B^{t_1,t_2}}$. This established relationship allows for the estimation of L^1. Notably, if vacant demes prevent the direct estimation of $\overline{B^{t_1,t_2}}$ from corresponding data, a model based imputation step can be performed simultaneously with the estimation of L^1.

3. Estimate $\overline{B^{t_0,t_2}}$, and L^0 based on corresponding data and estimated $\overline{B^{t_1,t_2}}$ using equation (5b). Here the logic is similar to that in step 2. One thing to be mentioned here is that equation (5a) can also be used to help estimate L^0 if $\overline{B^{t_1,t_2}}$ can be properly estimated from corresponding data. In practice people can either use a more convenient one or use both of them.

Joint inference of migration rates and coalescence rates

While the joint inference of migration and coalescence rates involves additional steps, the procedure for estimating migration rates remains consistent with the gene-flow-only protocol. Therefore, for these established steps, we will simply repeat the instructions without further explanation. New procedures specific to joint inference will be explained in detail. The protocol in the proof-of-concept example is as follows:

1. Estimate the $\overline{B^{t_2,t_2}}$ matrix from the corresponding data.

2. Estimate $\overline{B^{t_1,t_2}}$ and L^1 based on corresponding data and estimated $\overline{B^{t_2,t_2}}$ using equation (13c).

3. Estimate $\overline{B^{t_1,t_1}}$ and γ^1 based on corresponding data, estimated $\overline{B^{t_2,t_2}}$ and estimated L^1 using equation (8b). The idea here again is to use $\overline{B^{t_1,t_1}}$ as a mediater to connect γ^1 to corresponding data. An important thing to mention is that we directly use the estimate of L^1 from last step instead of estimating it jointly with L^1.

4. Estimate $\overline{B^{t_0,t_1}}$, $\overline{B^{t_0,t_2}}$, L^0 based on corresponding data, estimated $\overline{B^{t_1,t_1}}$ and estimated $\overline{B^{t_1,t_2}}$ using equation (5a) and (5b). The underlying logic is analogous to step 3 of the gene-flow inference protocol. The key distinction is that $\overline{B^{t_1,t_1}}$ is now already estimated from the previous step, which simplifies the direct application of equation (5a). Again, in practice people can either use a more convenient equation or use both of them.

5. Estimate $\overline{B^{t_0,t_0}}$ and γ^0 based corresponding data and estimated $\overline{B^{t_1,t_1}}$ using equation (8a). This step follows the same logic as step 3.

While the proof-of-concept example demonstrates the core framework, it can be generalized to cases with more than two time slices and other data types, such as sample covariance structures or LPSC segment data.

Simulation

Here, we test the feasibility of our first inference approach—inferring pure migration rates—through simulation studies. Similar to the previous analysis, we consider a model with three epochs, $[t_0, t_1)$, $[t_1, t_2)$, and $[t_2, t_3)$. In each epoch, we impose a different migration topology and assign random effective population sizes (i.e., coalescence rates). For each combination of topologies, we sample 10 haplotype lineages per deme per epoch and use msprime (Baumdicker et al., 2022) to simulate $10,000$ tree sequences. Parameters are then inferred from the pairwise branch length data of these simulated datasets as follows:

1. Estimate the $\overline{B^{t_2,t_2}}$ matrix from the simulated dataset, which we denote by $\widehat{B^{t_2,t_2}}$. For simplicity, in this step we directly use the sample mean of pairwise branch length, i.e. $\widehat{B^{t_2,t_2}_{\text{sample}}}$, since we have samples from all the demes.

2. Estimate $\overline{B^{t_1,t_2}}$ and L^1 based on corresponding data and estimated $\overline{B^{t_2,t_2}}$ using equation (13c). From last step, we already have $\widehat{B^{t_2,t_2}}$, and we first obtain a preliminary estimate of $\overline{B^{t_1,t_2}}$ by taking the sample mean of the pairwise branch lengths, denoted $\widehat{B^{t_1,t_2}_{\text{sample}}}$. We then infer L^1 by solving an optimization problem with objective $\| \varepsilon^1_r \|^2 + \lambda\Psi$, where ε^1_r is the relative error matrix of equation (13c) computed from $\widehat{B^{t_2,t_2}}$ and $\widehat{B^{t_1,t_2}_{\text{sample}}}$, Ψ is a smoothness penalty on edge weights, and λ is a hyperparameter selected by cross-validation (see Supplementary Information E for details).

Once we obtain $\widehat{L^1}$, we refine our estimate of $\overline{B^{t_1,t_2}}$ using

$$\widehat{B^{t_1,t_2}} = e^{\widehat{L^1}(t_2-t_1)}\big[\widehat{B^{t_2,t_2}} - (t_2 - t_1)1_{d\times d}\big]. \tag{25}$$

3. Estimate $\overline{B^{t_0,t_2}}$, and L^0 based on corresponding data and estimated $\overline{B^{t_1,t_2}}$ using equation (5b). The procedure follows the same logic as in last step: we first obtain a direct sample-mean estimate of $\overline{B^{t_0,t_2}}$, and then infer L^0 by solving an analogous optimization problem. Finally, we refine $\overline{B^{t_0,t_2}}$ using the resulting $\widehat{L^0}$.

The results of the simulation study are shown in Fig. 5. Figures 5(a–c) present the ground-truth topologies for each epoch. In Fig. 5a, the sequence of topologies is as follows: during epoch $[t_0, t_1]$, we have a topology of large-scale directionally migrating lineages, where lineages from the left boundary migrate backward to the right boundary; during epoch $[t_1, t_2]$, the topology is large-scale spatially converging lineages, where lineages converge backward to the center; and during epoch $[t_2, +\infty)$, the topology consists of a mixture of

small-scale patterns. We denote these by topology 1, topology 2, and topology 3, respectively. Thus, Fig. 5a corresponds to the topology sequence $1 \to 2 \to 3$. Figures 5b and 5c show two other settings obtained by rotating this sequence ($2 \to 3 \to 1$ for Fig. 5b and $3 \to 1 \to 2$ for Fig. 5c). The detailed setting and visualization of simulation parameters are provided in Supplementary Information F and Supplementary Fig. 1–3.

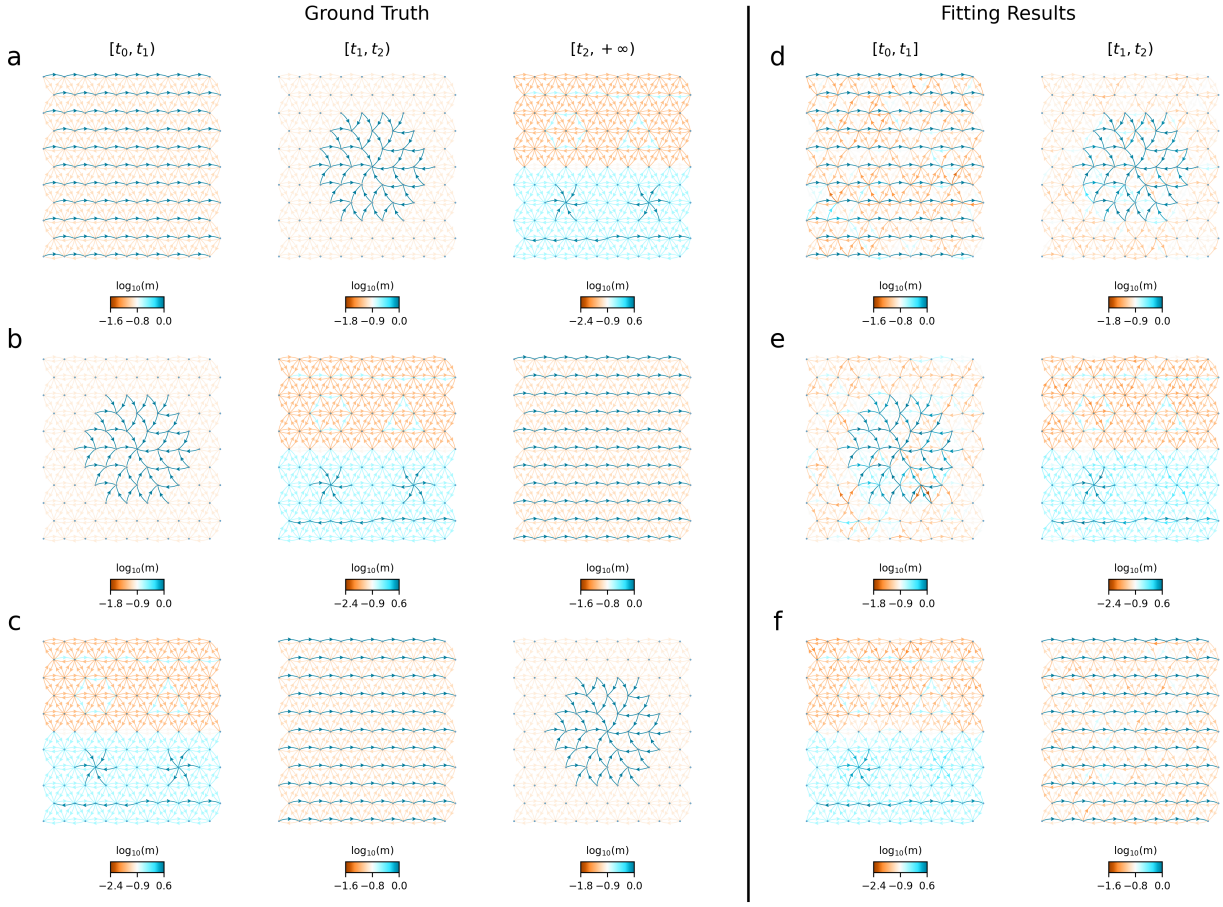

Fig. 5. Simulation results. The first row (**a**–**c**) presents the ground truth topology sequences across three epochs. We define three distinct migration topologies: topology 1 is large-scale directionally migrating lineages, topology 2 is large-scale spatially converging lineages, and topology 3 is a mixutre of small scale patterns. Panel (a) shows the sequence $1 \to 2 \to 3$, panel (b) the sequence $2 \to 3 \to 1$, and panel (c) the sequence $3 \to 1 \to 2$. The second row (**d**–**f**) shows the corresponding inferred migration patterns for each of the three settings.

Figures 5(d–f) show the inferred migration patterns. In all three settings, our method largely recovers the underlying structures, including many of the small-scale patterns. The only exception is the spatially diverging lineages in topology 3, which are less clearly recovered. This might be due to fact that the center of the spatially diverging lineages is visited much less than the center of the spatially converging lineages.

In previous studies such as EEMS (Petkova et al., 2016), FEEMS (Marcus et al., 2021), and FRAME (Shen and Novembre, 2025), migration rates could only be inferred in a relative sense. Here since we directly use the data of pairwise branch lengths, we can estimate the absolute migration rates. We provide a more detailed discussion of this issue in the discussion section.

Discussion

This study establishes an analytical foundation for pairwise genealogical processes under the structured serial coalescent. By deriving and solving evolution equations for pairwise branch length distributions and their

expectations, we characterize fundamental relationships governing the interplay of migration and coalescence across time intervals. Our systematic classification identifies distinct relationship categories— represented graphically as edge classes in Figs. 2-4 —and evaluates their inference utility through parametric dependencies and computational complexity. This analysis reveals a pivotal insight: relationships represented by vertical edges depend exclusively on migration rates, which is also validated in our analysis of relationships among length distributions of LPSC segments. This decoupling of spatial dynamics from coalescence effects provide a powerful tool to resolve the identifiability problem due to the interaction between migration and coalescence and enables fine-resolution inference of migration rates dynamics.

Leveraging this decoupling, we propose inference frameworks that fully exploits the migration-exclusive nature of relationships represented by vertical edges. Our approach employs a forward in time sequential inference strategy: beginning with inferring/imputing the focal quantity related to pairwise branch lengths and parameters in the oldest time point, we sequentially reconstruct migration rates (and coalescence rates) using relationships represented by vertical (and diagonal) edges. This inference framework— illustrated through a proof-of-concept example based on expected pairwise branch lengths —can possibly overcome the identifiability challenges that have historically hindered demographic inferences based on coalescent models. The architecture naturally accommodates time-stratified sampling and can be adapted to alternative summary statistics.

Despite these advances, several challenges remain for future work. First, the optimization and inference procedure we employ is straightforward and simple. While this may suffice for the proof-of-concept example, a more thorough investigation of the objective, penalty, and cross-validation procedures will be necessary to develop formal inference methods based on different statistics derived from pairwise branch lengths. What's more, for expected pairwise branch length based inference, estimation accuracy depends critically on reliable estimates of expected pairwise branch lengths. In empirical applications, pairwise branch lengths are estimated from inferred genealogical trees rather than the true underlying trees. The gap between these two sources requires careful evaluation, since most existing tree inference methods assume panmixia and do not incorporate a formal migration model. In addition, imputation procedures used to address demes with no samples may introduce further errors. In a sequential inference framework based on expected pairwise branch lengths , these errors can propagate, and similar propagation issues can arise in SNP-based inference and LPSC-based inference, although the sources of error can differ. Addressing this problem will be particularly important for inference spanning many epochs. One possible strategy to reduce error propagation is a block-wise sequential approach, in which parameters are inferred jointly across multiple consecutive epochs rather than strictly one epoch at a time.

Second, our proof-of-concept example does not address the issue of long-range migration. This challenge has already been noted in previous studies (Petkova et al., 2016; Marcus et al., 2021; Shen and Novembre, 2025). In reality, migration is typically local, which renders the migration-rate matrix sparse. Such sparsity can be leveraged for computational efficiency (Marcus et al., 2021) and parameter identifiability (Lundgren and Ralph, 2019). However, a strictly local network may be insufficient, as real populations may also exhibit a limited number of long-range connections. The difficulty is that these long-range edges are not known *a priori*; if they were, one could simply include them in the network. In FEEMS, two strategies are suggested: using a "greedy" algorithm akin to TreeMix (Pickrell and Pritchard, 2012), or combining the graphical lasso (Friedman et al., 2008) with graph Laplacian smoothing (Wang et al., 2016). The former has already been explored in FEEMSmix (Shastry et al., 2025), whereas the latter remains to be investigated.

Third, to further accelerate the inferences, it is important to adopt efficient algorithms to compute high-dimensional matrix exponentials—particularly for matrices with specific sparse structures. The case of pure migration rate estimation requires exponentiating a $d \times d$ matrix (e^{Lt}), while coalescence rate estimation involves a $d^2 \times d^2$ matrix (e^{St}). While directly computing e^{Lt} is still acceptable, directly computing e^{St} can be computationally expensive for large d. In this case, computation and approximation strategies that maximally exploit the structure of S can be very helpful.

As long as the inference challenges outlined earlier are properly addressed, this framework can support powerful applications across evolutionary biology, epidemiology, and conservation. In human evolution, it would facilitate reconstruction of demographic history from ancient DNA, clarifying migration patterns and their connections to cultural and environmental change. In epidemiological research, it could be applied to reconstruct detailed pathogen transmission dynamics across space and time, facilitating the monitoring of disease spread. In conservation genetics, it may reveal how gene flow among natural populations shifts through time and how these changes are shaped by environmental pressures such as habitat loss, fragmentation, and climate change. Most critically, the migration–coalescence decoupling enables inference of

migration rates with unprecedented spatial resolution. Together, these contributions provide new avenues to analyze population histories where spatial and temporal dimensions interact—from deep evolutionary timescales to contemporary ecological monitoring.

Acknowledgements

We would like to thank Ahmed Selim, Sherif Negm and Egor Lappo for helpful discussions and feedback on the manuscript. The research was supported by funding from NIH NIGMS grant R35-GM149521 to John Novembre.

Code Availability

The code to reproduce all results is available at https://github.com/ShenHaotv/When-migration-leaves-a-clean-trace.

References

Al-Asadi, H., Petkova, D., Stephens, M., and Novembre, J. (2019). Estimating recent migration and population-size surfaces. *PLOS Genetics*, 15(1):e1007908.

Baumdicker, F. et al. (2022). Efficient ancestry and mutation simulation with msprime 1.0. *Genetics*, 220:iyab229.

Beerli, P. and Felsenstein, J. (2001). Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. *Proceedings of the National Academy of Sciences of the United States of America*, 98(8):4563–4568.

Carmi, S., Palamara, P. F., Vacic, V., Lencz, T., Darvasi, A., and Pe'er, I. (2013). The variance of identity-by-descent sharing in the wright-fisher model. *Genetics*, 193(3):911–928.

De Maio, N., Wu, C.-H., O'Reilly, K. M., and Wilson, D. (2015). New routes to phylogeography: a Bayesian structured coalescent approximation. *PLoS Genetics*, 11(8):e1005421.

Drummond, A. J., Nicholls, G. K., Rodrigo, A. G., and Solomon, W. (2002). Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. *Genetics*, 161(3):1307–1320.

Drummond, A. J., Rambaut, A., Shapiro, B., and Pybus, O. G. (2005). Bayesian coalescent inference of past population dynamics from molecular sequences. *Molecular Biology and Evolution*, 22(5):1185–1192.

Ewing, G., Nicholls, G., and Rodrigo, A. (2004). Using temporally spaced sequences to simultaneously estimate migration rates, mutation rate and population sizes in measurably evolving populations. *Genetics*, 168(4):2407–2420.

Ewing, G. and Rodrigo, A. (2007). Estimating population parameters using the structured serial coalescent with bayesian MCMC inference when some demes are hidden. *Evolutionary Bioinformatics*, 2(2):227–235.

Friedman, J., Hastie, T., and Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441.

Kingman, J. F. C. (1982a). The coalescent. *Stochastic Processes and their Applications*, 13(3):235–248.

Kingman, J. F. C. (1982b). On the genealogy of large populations. *Journal of Applied Probability*, 19:27–43.

Lundgren, E. and Ralph, P. L. (2019). Are populations like a circuit? comparing isolation by resistance to a new coalescent-based method. *Mol. Ecol. Resour.*, 19(6):1388–1406.

Marcus, J., Ha, W., Barber, R. F., and Novembre, J. (2021). Fast and flexible estimation of effective migration surfaces. *eLife*, 10:e61927.

McVean, G. (2009). A genealogical interpretation of principal components analysis. *PLoS Genetics*, 5:e1000686.

Müller, N. F., Rasmussen, D., and Stadler, T. (2018). MASCOT: Parameter and state inference under the marginal structured coalescent approximation. *Bioinformatics*, 34(22):3843–3848.

Müller, N. F., Rasmussen, D. A., and Stadler, T. (2017). The structured coalescent and its approximations. *Molecular Biology and Evolution*, 34(11):2970–2981.

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

Palamara, P. F., Lencz, T., Darvasi, A., and Pe'er, I. (2012). Length distributions of identity by descent reveal fine-scale demographic history. *The American Journal of Human Genetics*, 91(5):809–822.

Petkova, D., Novembre, J., and Stephens, M. (2016). Visualizing spatial population structure with estimated effective migration surfaces. *Nature Genetics*, 48:94–100.

Pickrell, J. K. and Pritchard, J. K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genetics*, 8(11):e1002967.

Rodrigo, A. G. and Felsenstein, J. (1999). Coalescent approaches to hiv population genetics. In Crandall, K. A., editor, *The Evolution of HIV*, pages 233–272. Johns Hopkins University Press.

Shastry, V., Musiani, M., and Novembre, J. (2025). Jointly representing long-range genetic similarity and spatially heterogeneous isolation-by-distance. bioRxiv. `https://doi.org/10.1101/2025.02.10.637386`.

Shen, H. and Novembre, J. (2025). Fine-resolution asymmetric migration estimation. bioRxiv. `https://doi.org/10.1101/2025.05.29.656894`.

Strobeck, C. (1987). Average number of nucleotide differences in a sample from a single subpopulation: a test for population subdivision. *Genetics*, 117:149–153.

Takahata, N. (1991). Genealogy of neutral genes and spreading of selected mutations in a geographically structured population. *Genetics*, 29:585–595.

Vaughan, T. G., Kühnert, D., Popinga, A., Welch, D., and Drummond, A. J. (2014). Efficient Bayesian inference under the structured coalescent. *Bioinformatics*, 30(16):2272–2279.

Volz, E. M. (2012). Complex population dynamics and the coalescent under neutrality. *Genetics*, 190(1):187–201.

Wang, Y.-X., Sharpnack, J., Smola, A. J., and Tibshirani, R. J. (2016). Trend filtering on graphs. *Journal of Machine Learning Research*, 17(105).