Inferring the geographic history of recombinant lineages using the full ancestral recombination graph

Puneeth Deraje[1†*], James Kitchens[2†*], Graham Coop[2‡], and Matthew M. Osmond[1‡]

[†]These authors contributed equally to this work
[‡]These authors equally supervised this work
[1]Department of Ecology & Evolutionary Biology, University of Toronto
[2]Department of Evolution & Ecology and Center for Population Biology, University of California - Davis
[*]Corresponding authors: puneeth.deraje@mail.utoronto.ca and jkitchens@ucdavis.edu

Abstract

Spatial patterns of genetic relatedness among contemporary samples reflect the past movements of their ancestors. Our ability to untangle this spatial history has the potential to improve dramatically given that we can now infer the ultimate description of genetic relatedness, an ancestral recombination graph (ARG). By extending spatial methods previously applied to trees, we generalize a model of Brownian dispersal to ARGs, thereby accounting for correlations along a chromosome when computing the likelihood-based estimates of dispersal rate and locations of genetic ancestors. We develop an efficient algorithm that allows us to apply our method to complex ARGs, scalable to thousands of samples. We evaluate our method's ability to reconstruct

spatial histories using simulations. Surprisingly, despite using the fullest information available in the data, we find that our dispersal estimates are biased, highlighting a discrepancy between the histories of recombinant lineages and Brownian dispersal models. We identify potential resolutions to this problem based on relaxing the constraints that ARGs place on the movement of lineages and show that ARG-based spatial inference can be used to effectively track the geographic history of admixed individuals. Approaches like this will be key to understanding the interplay of migration, recombination, drift, and adaptation in geographically spread populations.

Keywords: Ancestral recombination graph, spatial population genetics, genetic inference, Brownian motion, networks, genetic ancestry.

1 Introduction

Life moves - offspring disperse, populations collapse together, and species' ranges shift. While most of these events go unnoticed in the moment, they leave spatial patterns in the genetic diversity of a sample. Though often faint, we can use such signals to gain insight into the spatial history of the samples' shared genetic ancestors.

One broad set of approaches to infer the spatial history of samples divides up sample genomes into a small number of geographic regions and estimate split times and rates of gene flow between these regions, from allele frequency (e.g., Excoffier *et al.*, 2021) and gene trees (e.g., Müller *et al.*, 2018). While useful, these approaches require the *a priori* grouping of samples and can obscure the fact that there can be population structure at many geographical scales. For example, in many species genetic differentiation builds up relatively smoothly with the geographic distance between samples. Such patterns motivate modeling approaches that treat space explicitly in

so-called isolation-by-distance models. A subset of these models assume local migration between groups of individuals (demes) on a lattice (e.g. the stepping-stone Kimura and Weiss, 1964; Malécot, 1948) allowing the rate of increase in F_{ST} with geographic distance to be used to estimate dispersal rates (Rousset, 2000, 1997). The alternative is to treat space in a truly continuous manner, avoiding the need to assume well-mixed demes or a fixed layout of individuals. At its core, the classic continuum model (Malécot, 1948; Wright, 1943) assumes lineages move as independent Brownian motions. The tractability of this model makes it useful for spatial inference, for example, inferring dispersal rate and the locations of genetic ancestors from gene trees (e.g., Lemmon and Lemmon, 2008; Novembre and Slatkin, 2009). As a result, independent Brownian motion remains a staple feature of many phylogeographic methods (e.g., Dellicour *et al.*, 2021). Further research has aimed at resolving the issues that arise when assuming lineages move as independent Brownian motions, namely the clustering of individuals in forward-in-time models (Felsenstein, 1975) and sampling inconsistency in backward-in-time models (Barton *et al.*, 2010a). The spatial Λ-Fleming-Viot process (Barton *et al.*, 2010a, 2013, 2010b) is an alternative that, among other things, incorporates local density dependence to avoid these issues. However the model's mathematical complexity makes inference computationally expensive (Wirtz and Guindon, 2023) and therefore limited to small sample sizes. Thus, independent Brownian motion, despite its limitations, continues to be an analytically tractable and computationally feasible model that is often useful for spatial inference, at least when dealing with non-recombining sequences.

On the empirical front, the increasing feasibility of whole genome sequencing has led to an influx of genetic data and motivated advances in the inference of the genealogical history of a sample undergoing recombination (e.g., Deng *et al.*, 2024; Kelleher *et al.*, 2019; Rasmussen *et al.*, 2014; Schaefer *et al.*, 2021; Speidel *et al.*, 2019; Wohns *et al.*, 2022). Recombination allows different regions of the same chromosome to have

different gene trees. Although single-tree approaches may be suitable when studying non-recombining sequences (e.g., mitochondrial DNA), they do not capture the range of genetic relationships found across recombining genomes nor the correlations between these relationships. For this, we must turn to ancestral recombination graphs (ARGs).

An ARG contains the complete genetic history of a sample of recombining genomes (Griffiths and Marjoram, 1996; Hudson, 1983; Lewanski *et al.*, 2023). It is commonly displayed as 1) a sequence of trees with each tree representing the history of a continuous block of the genome or 2) a single directed acyclic graph with annotated edges corresponding to their genomic intervals (see Fig 5A for an example of each). While the two representations can be interchangeable, tree sequences often lack the recombination events that tie the trees together and are further simplified, removing this equivalency (Wong *et al.*, 2023). ARGs are an incredibly rich source of information about the history of the sample (Harris, 2019; Hejase *et al.*, 2020; Lewanski *et al.*, 2023).

The utility of ARGs for spatial inference is still in its nascent stages, but growing. A number of approaches now exist for implicit space (e.g., Fan *et al.*, 2023; Guo *et al.*, 2022). Three approaches also exist for explicit continuous space: Osmond and Coop (2021) infer dispersal rate and ancestral locations under Brownian motion independently applied to trees sparsely sampled from a simplified ARG; Wohns *et al.* (2022) infer ancestral locations by placing each node at the midpoint of its descendant nodes on a simplified ARG; and Grundler *et al.* (2024) infer dispersal rate and ancestral locations by placing each node at the location that minimizes a migration cost averaged over the trees in a simplified ARG that the node appears in. None of these continuous space approaches utilize all of the information contained in the ARG with an explicit model for spatial movement, which is what we aim to do here.

Here we extend the classic Brownian motion model for trees to describe movement

down an ARG, forcing two lineages to meet at a recombination node. Under this model we derive the full likelihood of the sample locations given an ARG, allowing us to to infer the dispersal rate and the location of every genetic ancestor in the ARG using the complete genealogical history of the sample. We highlight both the mathematical and computational challenges that are posed due to recombination loops in the ARG. We then provide a mathematically rigorous solution and a computationally fast algorithm to infer spatial histories, which we test with simulations. While using all the information in an ARG, it is not immediately clear how well our method will infer spatial history, especially given the unlikeliness of two independent Brownian motions meeting (Etheridge, 2019).

2 Methods

Our aim is to infer dispersal rates and the locations of genetic ancestors given a complete description of the genetic relatedness among a set of samples, a time-calibrated ARG with recombination nodes (often referred to as a full ARG, e.g., Baumdicker *et al.*, 2022; Lewanski *et al.*, 2023; Shipilina *et al.*, 2023).

An ARG is graphical representation of the genealogical history of a set of sample genomes that may have undergone recombination in the past (Wong *et al.*, 2023). As the history of the samples at each site can be depicted as a tree, an ARG weaves together these histories based upon their shared structure to represent the history of the samples across the entire genome. Each node in an ARG represents a haploid genome and directed edges connect these nodes to describe the line of inheritance over time. Nodes that are the product of recombination will be connected to two ancestral nodes with annotated edges referring to the specific regions of the chromosome that were inherited from each ancestor. Here, we define a path through the ARG as a sequence of edges which connects a sample node (a tip of the ARG) back in time to

a root. Time is measured backwards from the present, starting from the most recent sample node at time $t = 0$ and increasing as we go deeper into the past towards the root(s).

We model the movement of genetic material forward in time by Brownian motion with dispersal rate σ^2 (see Table 1 for a list of key symbols). In other words, we assume that the location of a node is normally distributed about its parent node with variance $\sigma^2 t$, where t is the length of the edge connecting them (in generations). Since in each generation autosomal inheritance is equally likely via the mother or father, the effective variance is the average of maternal and paternal variances (e.g., see Smith *et al.*, 2023). While the computations are shown for one dimension, they are readily extended to two dimensions by replacing the dispersal rate σ^2 with a dispersal matrix $\Sigma = \begin{bmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{bmatrix}$.

Brownian motion is a commonly used model in phylogeography (e.g., Lemmon and Lemmon, 2008), where there is a single tree relating the samples, directly analogous to earlier models of continuous trait evolution on a phylogeny (Felsenstein, 1985). More recently, this model has been applied to a sequence of sparsely sampled trees along the genome (Osmond and Coop, 2021) assuming independence between each sampled tree. As we explain below, extending this approach from a sequence of independent trees to a graph is non-trivial. We will start by building intuition through the simplest examples, such as small, handmade ARGs with a single root, before tackling more complex scenarios that arise from simulations. While all samples are contemporary in the ARGs presented in this paper ($t = 0$), our method can be readily used with non-contemporary samples, including samples that are direct ancestors of another sample.

Table 1: Description of key symbols used

Symbol	Name	Description and Notes
n_s	Number of samples	
n_p	Number of paths	Total number of paths from all samples to the root
n_r	Number of roots	Total number of roots in the ARG
σ^2	Dispersal rate	Variance in the offspring locations relative to parent location
$\mathbf{S}_p$	Path Matrix	Entries are shared times between any two paths. This could refer to either the Full Path Matrix of the Minimal Path Matrix
$\mathbf{S}$	Samples Matrix	$\sigma^2\mathbf{S}$ is covariance matrix between sample locations
$\mathbf{P}_p$	Path-Sample Matrix	$n_p \times n_s$ matrix whose entry is 1 if path i is associated with sample j
$\mathbf{R}_p$	Path-Root Matrix	$n_p \times n_r$ matrix whose entry is 1 if path i is associated with root j
$\overrightarrow{L}$	Sample location vector	Vector valued random variable of sample locations of length n_s
$\overrightarrow{L_p}$	Path location vector	Vector valued random variable of length n_p
$\vec{\mu}$	Root location vector	Vector of length n_r
L_a		Random variable denoting location of internal node a
$\vec{s_a}$		Vector valued random variable of length n_p with shared time between the path of internal node a and paths corresponding to $\mathbf{S}_p$

2.1 Problem of Loops

The first step in estimating dispersal rates and the locations of genetic ancestors given an ARG is calculating a likelihood for the sample locations. Under Brownian motion, the vector of sample locations, $\overrightarrow{L}$, has a multivariate normal distribution, specifically

$$\overrightarrow{L} \sim \mathcal{N}(\mu \mathbb{1}_n, \sigma^2 \mathbf{S}), \tag{1}$$

where μ is the location of the root of the ARG, σ^2 is the dispersal rate, and $\sigma^2\mathbf{S}$ is the covariance matrix between the sample locations. The key in computing the maximum likelihood estimates (MLEs) of the root location, μ, and the dispersal rate, σ^2, is knowing the entries of the sample matrix, $\mathbf{S}$.

The methods for calculating the sample matrix, $\mathbf{S}$, differ when working with trees versus ARGs. In the case of a tree, we can assume that the displacement along any edge of the tree foward in time is independent of the displacement along every other edge and is distributed as $D_{edge} \sim \mathcal{N}(0, \sigma^2 t_{edge})$, where t_{edge} is the length of the edge. In a tree, each sample is associated with a single path connecting it back to the root. Hence, the sample location is given by the sum of the displacements along all edges in the sample's path. The entries of the sample matrix, $\mathbf{S}$, for a tree are then simply the shared times between the paths associated with the corresponding pair of samples.

Computing the entries of $\mathbf{S}$ for an ARG is not as straightforward because of two related reasons stemming from the presence of recombination nodes. Firstly, each recombination node creates a loop within the graph. We assume that two individuals have to physically be at the same location for mating. So the two parental lineages need to meet at the recombination node. So, the displacement along the edges involved in the loop are not independent; they must satisfy the condition that the sum of the displacements around the left half of the loop must equal the sum around the right half. For instance in Figure 1, the displacement from node 7 to node 6 and the displacement from node 6 to node 4/5 must sum to the displacement from node 7 directly to node 4/5, $D_{7,6} + D_{6,4/5} = D_{7,4/5}$. Each recombination node in a graph has an associated loop, and we define η_{loops} as the combination of all of these constraints. Secondly, the presence of a loop means that the samples below the recombination node have multiple paths connecting them to the same root. For example in Figure 1, there are two paths that connect sample 0 to the root, $(0, 3, 5, 6, 7)$ and $(0, 3, 4, 7)$. It is therefore not possible to calculate a single shared time between two samples, as

it was for a tree.

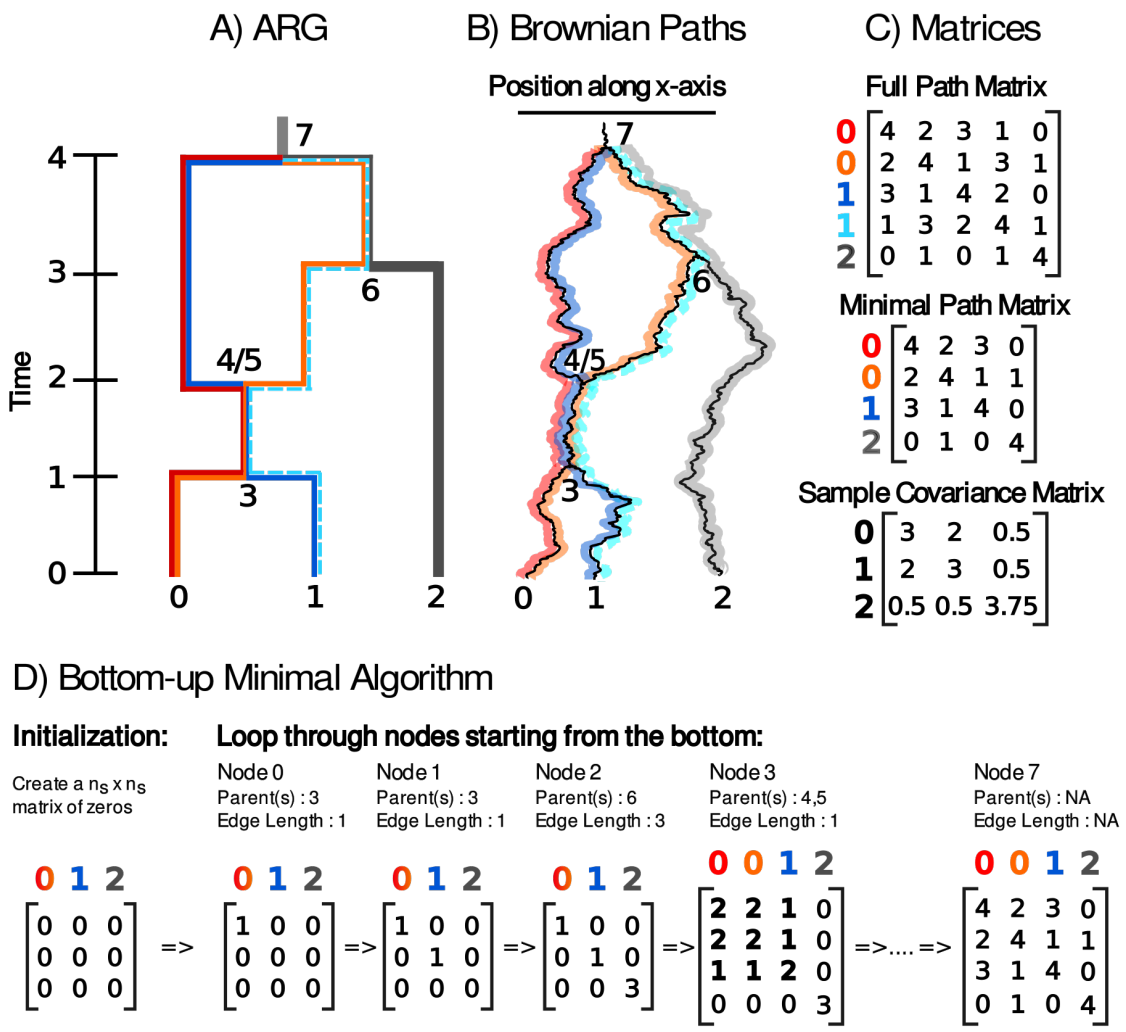

Figure 1: Overview of methods. (A) A simple example ARG of three samples with paths highlighted in different colors. The dashed blue line is the path that is excluded while calculating the minimal shared time matrix. (B) A cartoon of Brownian motion down this ARG. Note that we are implicitly assuming here that the lineages all coalesce in the recent past, otherwise their spatial locations would traverse back and forth across space instead of neatly merging to the middle. (C) The full and minimal shared time matrices between pairs of paths and the sample matrix. (D) Illustration of the bottom-up algorithm to build the minimal path matrix.

2.2 Solution of Paths

To solve this problem, we start by incorrectly allowing the displacement along each edge in the ARG to be independent. This means that a sample with multiple paths to the root has multiple locations, one for each path. Let n_p be the total number of paths from all the samples to the root. Unlike trees, where the number of paths is always equal to the number of samples, the number of paths through an ARG grows with the number of recombination nodes within the graph. Let $\overrightarrow{L_p}$ be a vector of length n_p containing the sample locations of each of these paths. $\overrightarrow{L_p}$ is then multivariate normal with mean $\mu\mathbb{1}_p$ and covariance matrix $\sigma^2\mathbf{S}_p$, where $\mathbf{S}_p$ is the matrix of shared times between each pair of paths. In order to compute the correct distribution of the sample locations, $\overrightarrow{L}$, we need to condition $\overrightarrow{L_p}$ on the paths along the same recombination loop meeting, η_{loops}. While computing $\overrightarrow{L_p}|\eta_{loops}$ is possible, it is cumbersome and difficult to do algorithmically. Instead, we condition on an equivalent set of events, that all paths starting at the same sample should have the same location. We call this set of events η_{paths}. In Figure 1, this set consists of two conditions, one corresponding to paths ending at sample 0 and the other at sample 1, $\eta_{paths} = \{\mu + D_{7,4/5} + D_{4/5,3} + D_{3,0} = \mu + D_{7,6} + D_{6,4/5} + D_{4/5,3} + D_{3,0}\ ,\ \mu + D_{7,4/5} + D_{4/5,3} + D_{3,1} = \mu + D_{7,6} + D_{6,4/5} + D_{4/5,3} + D_{3,1}\}$. Note that by cancelling common terms on opposite sides of the equality, η_{paths} reduces to η_{loops}, which we can prove more generally (Appendix B). $\overrightarrow{L_p}|\eta_{paths}$ is easier to compute algorithmically and gives the same normal distribution as above (Equation 1), with the sample covariance matrix now defined as

$$\mathbf{S} = (\mathbf{P}_p^T \mathbf{S}_p^{g-} \mathbf{P}_p)^{-1}. \tag{2}$$

so the for each recombination nodes, Here, $\mathbf{S}_p^{g-}$ is the generalized inverse of $\mathbf{S}_p$ and $\mathbf{P}_p$ is a conversion matrix of size $n_p \times n_s$ that pairs each path i with its associated sample j (see Appendix A for a proof). The i, j^{th} element of $\mathbf{P}_p$ is 1 if path j starts

at sample i, otherwise it is 0.

Finally, while we could calculate the shared times between every path through the ARG, many of the paths are not linearly independent from one another and are therefore redundant in our calculations. In practice, we use only a minimal subset of paths and its corresponding shared time matrix (see Figure 1 for an example). While the total number of paths grows rapidly (faster than linearly) with the number of recombinations, the minimal number of paths grows linearly (being equal to the number of samples plus the number of recombinations; see Appendix C).

2.3 Multiple Roots

There are multiple reasons why we might want to focus on just the recent past. First, long-term movement may not be accurately captured by Brownian motion due to geographic boundaries or large-scale population movements (Ianni-Ravn *et al.*, 2023). Second, as we move further back into the past, sample lineages can become spatially well mixed (Wakeley, 1999). Therefore, it becomes difficult or impossible to extract any meaningful spatial information about the deep history of samples. Third, with ARG inference, deeper nodes in the ARG are often poorly resolved, both in timing and topology. To avoid these issues, we will often want to cutoff the ARG at a given time in the past and ignore any deeper connections.

When we chop an ARG below the grand most recent common ancestor (root), the graph no longer has a single root. Instead, there are multiple roots, each being associated with specific paths through the ARG. Let $\vec{\mu}$ be the vector of root locations (of length n_r) and $\mathbf{R}_p$ be a conversion matrix of size $n_p \times n_r$. The i, j^{th} element of $\mathbf{R}_p$ is 1 if path j terminates at root i, otherwise it is 0. Then $\mathbf{R}_p\vec{\mu}$ is the vector of root locations associated with each of the n_p paths.

The full covariance between the sample locations is the covariance created by the structure of the ARG plus the covariance between the root locations, $\sigma^2(\mathbf{S}_p +$

$\mathrm{Cov}(\mathbf{R}_p\vec{\mu}, \mathbf{R}_p\vec{\mu}))$. Here we assume that the root locations are independent of each other and each have zero variance, $\mathrm{Cov}(\mathbf{R}_p\vec{\mu}, \mathbf{R}_p\vec{\mu}) = 0$. The assumption of independence is reasonable if we cutoff the ARG at a point by which the ancestors are well mixed in a finite habitat. The assumption of zero variance in root locations can be relaxed by adding a variance to the covariance term between paths starting at the same root.

We can compute the MLEs for the root locations given the observed sample locations, $\vec{\ell^*}$, as a solution to

$$\mathbf{R}_p^{\mathrm{T}}\mathbf{S}_p^{g-}\mathbf{R}_p\widehat{\vec{\mu}} = \mathbf{R}_p\mathbf{S}_p^{g-}\mathbf{P}_p\vec{\ell^*} \tag{3}$$

(see Appendix A.2 for a proof). If $\mathbf{R}_p^{\mathrm{T}}\mathbf{S}_p^{g-}\mathbf{R}_p$ is invertible, then there exists a unique solution,

$$\widehat{\vec{\mu}} = (\mathbf{R}_p^{\mathrm{T}}\mathbf{S}_p^{g-}\mathbf{R}_p)^{-1}\mathbf{R}_p\mathbf{S}_p^{g-}\mathbf{P}_p\vec{\ell^*}. \tag{4}$$

2.4 Dispersal Rate and Location of Genetic Ancestors

Once we have estimated the locations of the roots, we can compute the MLE of the dispersal rate,

$$\widehat{\sigma^2} \quad = \quad \frac{(\mathbf{P}_p\vec{\ell^*} - \mathbf{R}_p\widehat{\vec{\mu}})^{\mathrm{T}}\mathbf{S}_p^{g-}(\mathbf{P}_p\vec{\ell^*} - \mathbf{R}_p\widehat{\vec{\mu}})}{n}. \tag{5}$$

With estimates of both the root locations and dispersal rate we can then calculate the distribution of any internal node location (genetic ancestor). Given an internal node, we choose an arbitrary path from that node to one of the roots. Conditional on the dispersal rate, root locations, and observed sample locations, the location of the

internal node, L_a, is normally distributed with mean

$$E[L_a|\widehat{\vec{\mu}}, \widehat{\sigma^2}] = \widehat{\mu}_{r_a} + \vec{s_a}^{\mathrm{T}}\mathbf{S}_p^{g-}(\mathbf{P}_p\vec{\ell^*} - \mathbf{R}_p\widehat{\vec{\mu}}), \tag{6}$$

where $\vec{s_a}$ is the vector of shared times with the minimal paths used to construct $\mathbf{S}_p$ and $\widehat{\mu}_{r_a}$ is the MLE location of the ${r_a}^{\text{th}}$ root (the ${r_a}^{\text{th}}$ element of $\vec{\mu}$). The total variance in the internal node's location is a combination of the variance due to Brownian motion and the variance due to uncertainty in the root locations,

$$\mathrm{Var}(L_a) = \widehat{\sigma^2}\Bigg(\underbrace{t_a - \vec{s_a}^{\mathrm{T}}\mathbf{S}_p^{g-}\vec{s_a}}_{\text{due to Brownian motion}} + \underbrace{(\vec{e}_{r_a} - \mathbf{R}_p^{\mathrm{T}}\mathbf{S}_p^{g-}\vec{s_a})^{\mathrm{T}}(\mathbf{R}_p^{\mathrm{T}}\mathbf{S}_p^{g-}\mathbf{R}_p)^{-1}(\vec{e}_{r_a} - \mathbf{R}_p^{\mathrm{T}}\mathbf{S}_p^{g-}\vec{s_a})}_{\text{due to uncertainity in root locations}} \Bigg), \tag{7}$$

where $\vec{e}_{r_a}$ is a length n_r vector that is zero everywhere except at the ${r_a}^{\text{th}}$ position.

2.5 Algorithm

The key object for estimating dispersal rates and ancestor locations is the matrix of shared times between the minimal set of paths through the ARG, $\mathbf{S}_p$. Though there are existing methods in `Python` to identify all paths from the samples to the roots, calculating the intersection between these paths does not scale well to larger ARGs, primarily due to repeated calculation of common edges across different paths (Figure 2). We therefore developed an algorithm that requires traversing each edge only once (Appendix C.1) and, in doing so, have greatly sped up the calculation of $\mathbf{S}_p$ (Figure 2).

Briefly, the algorithm entails a bottom-up traversal of the ARG starting at the sample nodes and updating the shared time matrix as we move upwards towards the roots (Figure 2). For each node visited, the algorithm calculates the edge length between that node and its parent. This is added to the corresponding cells in the shared time matrix. When we reach a recombination node (which has multiple parents),

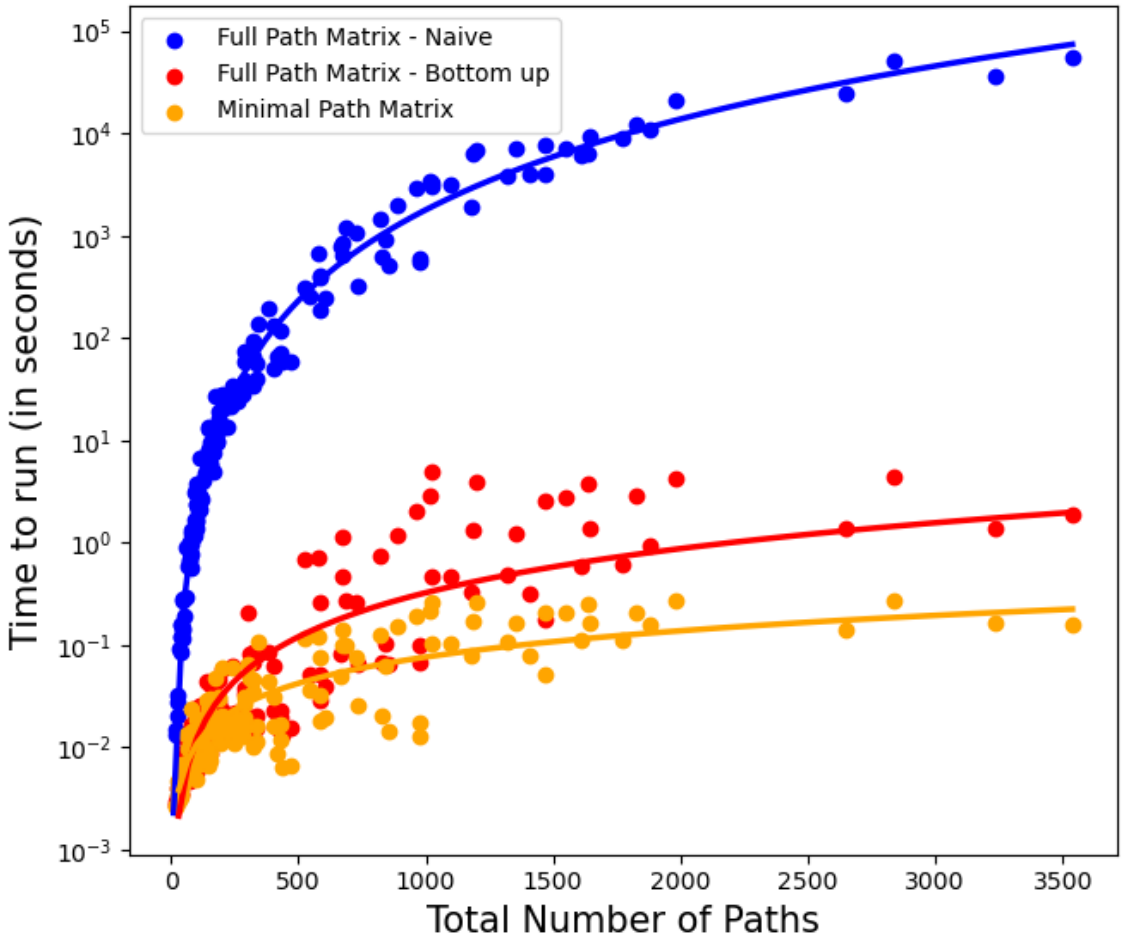

Figure 2: **Algorithm benchmarks**. Number of seconds to compute the full path and minimal path matrices using different algorithms plotted as a function of the total number of paths in the ARG. "Full Path Matrix - Naive" (blue) uses existing Python methods to compute the full path matrix. "Full Path Matrix - Bottom up" (red) instead computes the full path matrix with a single bottom-up traversal of the ARG. "Minimal Path Matrix" (orange) uses the bottom-up method to compute the covariance between the smallest set of linearly independent paths, which is sufficient for estimating parameters of interest. Random ARGs of various sizes were generated (number of samples ranged up to 500 and sequence lengths up to 5000 basepairs) using the msprime Python package. The solid lines are the best fits under a power law. The best fit exponents for the power law are 2.946 (Full Path - Naive), 1.432 (Full Path - Bottom up) and 0.853 (Minimal Path).

the relevant rows and columns are duplicated, expanding the size of the matrix and corresponding with the separation of these paths in the ARG. This keeps the size of the matrix small for as long as possible, making it more efficient. Currently, the algorithm is implemented using the `tskit` package (Kelleher *et al.*, 2018).

2.6 Windowing

While our method and algorithm are designed to scale to full chromosomes, we can decide to analyze a smaller portion (window) of the genome, e.g., a window centered on a particular locus or the recombination breakpoint of interest. Windowing may also offer a more practical approach when working with large datasets, such as biobanks (e.g., Zhang *et al.*, 2023).

There are number of ways we can choose windows. A genetic ancestor (internal node) is associated with a specific region of the chromosome. To locate this ancestor we may therefore want to center a window at the center of this region, which will generally fall within a local tree. The smallest window we can consider is simply this local tree. In contrast, we might be interested in a particular recombination event, which necessarily falls directly between two local trees. The smallest window we then want to consider is the two neighboring trees. Here, we define the size of a window as the number of neighboring trees on either side of the smallest window. This means that a window of size 0, referred to in the results as W_0, includes just the windows described above (local tree for genetic ancestor or two neighboring trees for recombination breakpoint). A window of size 1 (W_1) includes one neighboring tree on either side of the smallest interval, and so on. When a window extends beyond an edge of a chromosome there will be fewer trees than the window size suggests. Smaller windows are faster to analyze but contain less information. The size of a suitable window will likely depend on the organism and sample, and remains an open question. Note that different windows will give different locations for the same node.

2.7 Spatially-Explicit Simulations

We performed individual-based two-dimensional spatial simulations using SLiM v4.0 (Haller and Messer, 2023), extending those run by Osmond and Coop (2021). To assess the utility of our method to infer parameters under more realistic dynamics than Brownian motion, we simulate density-dependent reproduction as well as finite boundary habitats. Simulations started with 10,000 individuals uniformly randomly distributed in a 100×100 unit area, with each individual being diploid for a 1 megabase chromosome. All individuals are hermaphrodites. In each generation an individual acts once as a mother, choosing its mate randomly based on distance (we assume a Gaussian mating kernel with variance σ_m^2). The number of offspring for each mating pair is a Poisson random variable with mean $\lambda = \frac{2}{1+C}$, where C is the sum of the interaction strengths with neighbors. Here, interaction strengths are Gaussian with variance σ_c^2. It is possible that there are no mates within the interaction distance, in which case no offspring are produced. Offspring are placed relative to their mother's position with a normal random variable offset in each dimension with variance σ_d^2. If the offset would place the offspring outside of the area, the offset is reflected off of the boundary wall and back into the area. The locations and relationships between individuals are recorded in a tree sequence, which is saved at the end of the simulation (Haller *et al.*, 2019).

We emphasize that, due to both local density-dependence and habitat boundaries, the movements of genetic lineages will not be Brownian. We are interested in how well we can infer the dispersal rate, effectively $\sigma_d^2 + \sigma_m^2/2$, and the locations of ancestors under the assumption of Brownian motion.

2.8 Data Availability

Our method is available as a Python package at https://github.com/osmond-lab/sparg. The code for all of our analyses in this paper is available in the "manuscript" branch of the GitHub repository (https://github.com/osmond-lab/sparg/tree/manuscript).

3 Results

3.1 Dispersal Rate

We first used our method to estimate dispersal rates (Equation 5) from simulated ARGs and compare these to dispersal estimates obtained from marginal trees (using the composite likelihood-based method of Osmond and Coop, 2021). Unexpectedly, we found that our ARG-based dispersal estimates were systematically biased. While disappointing, this highlights modeling issues arising from using Brownian motion models of dispersal that are worth understanding given that Brownian motion is a commonly used model. Despite the bias in the ARG-based estimate, we show that they more fully account for the uncertainty in our dispersal estimates compared to the composite likelihood estimate. This showcases an advantage of using the full ARG as opposed to a sequence of trees.

3.1.1 Biased Estimates

Consistent with previous work (e.g., Ianni-Ravn *et al.*, 2023; Kalkauskas *et al.*, 2021), the dispersal estimate from the composite likelihood over trees underestimates the true simulated value (blue curve in Figure 3A), due to habitat boundaries. As we increase the number of trees this estimate, the average dispersal estimate over all trees, asymptotes. In contrast, the dispersal estimate from the full ARG likelihood

systematically increases as we include more trees, starting as an underestimate but eventually leading to an overestimate of the true dispersal rate (black curve in Figure 3A).

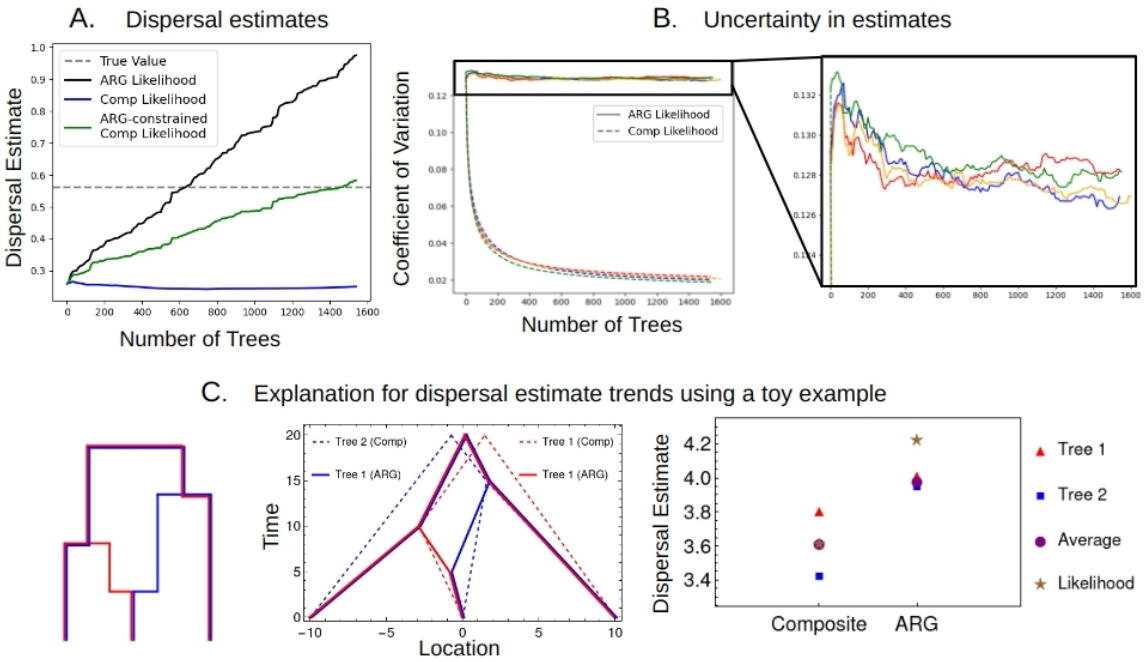

Figure 3: **Dispersal rate**. (A) Dispersal rate estimates as a function of the number of trees used. We compare the maximum composite likelihood estimate over trees (blue), the maximum composite likelihood estimate over trees after fixing the node locations according to the ARG (green), and the full ARG maximum likelihood estimate (black). (B) The coefficient of variation in the dispersal rate when computed using the composite likelihood over trees (dashed) and the ARG likelihood (solid) as a function of the number of trees used. Each color is an independent replicate. (C) A toy 2-tree ARG (left). The most likely locations of ancestors given the sample locations (middle) for the two trees (red and blue) when computed independently (dashed) and when computed using the ARG (solid). Dispersal rates estimates (right) assuming node locations in the two trees are independent (composite) or a compromise (ARG).

There are two main reasons for the systematic increase in the ARG dispersal rate, both of which arise, mathematically speaking, because of how the MLE dispersal rate under Brownian motion is calculated, by first choosing internal node locations to minimize the displacements that descendent nodes need to disperse. The MLE

dispersal rate is then the average squared displacement over edges,

$$\hat{\sigma^2} = \min_{\text{node locations}} \frac{1}{n} \sum_{\text{edge in ARG}} \frac{D_{edge}^2}{t_{edge}}, \tag{8}$$

where n is the number of edges and D_{edge} and t_{edge} are the displacement and length of an edge (Maddison, 1991). We can then understand the ARG dispersal bias as coming from two factors:

1. **Compromising ancestral locations across trees**. When trees are treated independently, e.g., under the composite likelihood estimate of dispersal, the node locations in each tree can be independently chosen to minimize the displacements (the set of D_{edge} for each tree). However, the nodes within an ARG are often shared across multiple trees, and so the choices of node locations must minimize the dispersal distances across all of the trees together rather than any tree individually. This ultimately leads to higher dispersal rates for each tree. To demonstrate this we recalculated the composite likelihood dispersal rate but now with each individual tree's node locations computed from the ARG ("constrained composite likelihood") and found it to be higher than when each tree's node locations are minimized independently (see toy example in Figure 3C). Further, the more trees represented in the ARG, the greater the constraint, leading to higher dispersal rates for each tree and consequently a higher average (green curve in Figure 3A) compared to the (unconstrained) composite likelihood (blue curve in Figure 3A). Thus, the compromise over node locations within the ARG partly explains why the ARG dispersal estimate is higher than the composite likelihood estimate over trees and why it increases with the number of trees (more trees lead to more constraint), but it does not fully explain the issue, as the constrained composite likelihood dispersal rate is lower than the ARG dispersal rate (black curve in Figure 3A).

2. **Averaging dispersal rates over trees**. To understand why the ARG dispersal estimate is greater than even the constrained composite likelihood estimate, note that constrained composite likelihood estimate is the average dispersal estimate over all trees,

$$\hat{\sigma^2}_{constrained} = \frac{1}{\#\text{trees}} \sum_{\text{tree}} \frac{1}{n} \sum_{\text{edge in tree}} \frac{D^2_{\text{edge}}}{t_{\text{edge}}}, \tag{9}$$

where the displacement over the edge, D_{edge}, is calculated from node locations inferred from the full ARG (Equation 8). Now, an edge that exists in every tree would contribute the same amount to the constrained composite estimate (Equation 9) as it would to the ARG estimate (Equation 8), but an edge that belongs to fewer trees will contribute less to the constrained estimate (Equation 9). Consequently, the more trees that we include in the ARG, the greater the number of edges that are not shared by all trees. Hence, the constrained composite likelihood estimate does not increase as quickly as the ARG estimate. That said, the averaging in the composite likelihood (Equation 9) is not the right approach as it ignores the non-independence of the trees. Intuitively, each edge in the ARG should be given equal weight while computing the dispersal rate, justifying the ARG estimate (Equation 8) under the assumption of Brownian motion. While the ARG estimate may be better justified from a statistical perspective, the mis-specification of Brownian motion as a model of dispersal generates significant biases.

We next propose two methods that reduce the compromise across the trees. Both methods have better properties with respect to dispersal estimates.

1. **Windowing**. Using a subset of adjacent trees from the ARG reduces the compromise in ancestral locations. This preserves the key benefits of our method,

that is, using the full likelihood over multiple trees and the computational efficiency of the minimal paths algorithm. To calculate the dispersal rate we can split the genome into a series of non-overlapping windows and take the composite likelihood over these. The downside of this approach is that we ignore the correlations between trees in different windows (but see, for e.g., Larribe and Fearnhead (2011) and Meligkotsidou and Fearnhead (2007) for previous uses of composite likelihood in genetics). The resulting dispersal estimate asymptotes with the number of trees at a larger value than the composite likelihood estimate, as there are still some compromises in node locations (red curve in Figure S1).

2. **Relaxed Meeting**. Instead of forcing parental lineages to precisely meet at a recombination node, we explore the other extreme and allow the two parents of a recombinant offspring to be any distance from each other, placing the offspring at their midpoint (Appendix D). Doing this allows adjacent trees to adjust the locations of their non-shared nodes more freely. The resulting dispersal estimate increases more slowly than the constrained composite likelihood with the number of trees (Figure S1) and sometimes decreases. Unfortunately, the dispersal estimate still increases on average and is more computationally expensive since we need to use the Full Paths Matrix instead of the Minimal Paths Matrix.

3.1.2 More Appropriate Uncertainty

Despite the biases of the ARG dispersal rate estimate, the ARG method better captures the amount of information that each additional tree contributes. Specifically, the structure of marginal trees along a sequence are correlated as nodes and branches are shared along the sequence. This implies that the spatial histories contained in these trees is also correlated. The composite likelihood ignores this correlation and

therefore every new tree reduces the uncertainty in the dispersal rate much more than it should. The ARG likelihood accounts for the correlation and, therefore, reduces the uncertainty in accordance to the new information being contributed by each tree (Figure 3B).

3.2 The Locations of Genetic Ancestors

3.2.1 Using More Information

Our method uses the structure and edge lengths from the full ARG and in doing so captures more information than existing methods. To illustrate the implications of this for locating genetic ancestors, we apply our method to a simple ARG (Figure 4A) and compare our estimate of the most likely location of node 8 to 1) the estimate from each marginal tree (as a heuristic for the likelihood method used in Osmond and Coop, 2021) and 2) the estimate from a simple averaging method, where a node's location is the average of its child nodes' locations (as a heuristic for the method used in Wohns *et al.*, 2022).

Our method captures the effect of varying the time of any node (black curves in Figure 4B). In contrast, the averaging method is insensitive to the time of any node as it does not use edge lengths (green lines in Figure 4B). Meanwhile, the tree method (red and blue curves in Figure 4B) captures changes in the time of a node that affects the edge lengths in a tree (e.g., node 4) but not of a node that leaves the trees unchanged (e.g., node 5/6). When the time of a node affects both the trees and is in a loop (e.g., node 7), the ARG and tree methods can show opposite trends. Note that, under our model, increasing the time of node 7 should move the location of node 8 towards the locations of nodes 0 and 1 and away from the location of node 2 because the amount of time the path above 2 spends in the loop is reduced, increasing

the probability node 2 has moved further away. Only our method accurately captures this.

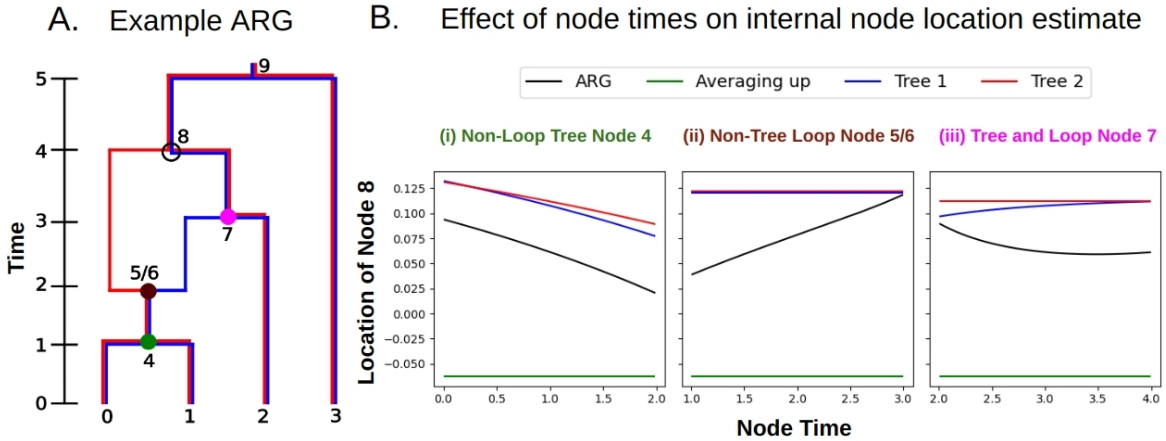

Figure 4: **Ancestor locations**. (A) Toy 2-tree ARG. (B) Most likely location of node 8 computed using the ARG (black), individual trees (red and blue), and the averaging-up method (green), plotted as a function of the time of (i) a node that alters individual trees and the ARG but is not part of the ARG loop, (ii) a node that doesn't alter individual trees but is part of the ARG loop and (iii) a node which is part of the ARG loop that alters a tree and the ARG.

3.2.2 Reduced Uncertainty

The ARG provides more information about the geographic locations of genetic ancestors than any single tree. Nodes that are shared across multiple trees provide useful information about historical connections. We can also estimate node locations with more certainty. We demonstrate this with a simple 4-tree ARG (Figure 5A), tracking the location of ancestors along a particular path (Figure 5B). We compare the uncertainty (variance) in these locations with the uncertainty we get when using only a single tree (Figure 5C).

In the absence of any other information, the variance in the location of a genetic ancestor increases linearly with time under Brownian motion (at rate 1 in Figure 5C). In a tree, new information enters via coalescence events, which reduces the rate of increase in variance along edges below coalescence nodes (here the edge from 0 to 7).

In an ARG, we generally see less uncertainty along a lineage because we get more coalescence events (here nodes 10 and 11) and we also get some information coming in at recombination events (here nodes 2/3 and 5/6). Only along edges that only exist in a single tree (e.g.,here from node 6 to 7) are the uncertainties from the ARG and the single tree potentially equal in some places. In practice, inferred ARGs will have many more samples than our toy example and so nearly every edge will exist in more than one tree. Our method would then estimate ancestral locations with much more certainty than any single tree.

Note that when enough information comes in it is possible for the uncertainty in ancestor location to decline as we look back in time, using either trees or ARGs. Here we see such a decline at node 7 when using the ARG. We do not see this decline in uncertainty with the tree because less information comes in; the loop in the ARG reduces the uncertainty in the locations of ancestors along it.

3.2.3 Accuracy

Having shown that our method of locating ancestors captures more information from the ARG and has lower uncertainty with hand-built ARGs, we now quantify the accuracy of our location estimates using individual-based spatially-explicit simulations (see details in Methods). We simulated a single ARG (1 megabase chromosome, containing 1538 trees), subset it to 1000 samples, and chopped it off at 2000 generations from the present (as we are particularly interested in the accuracy of estimates in the recent past). We then chose 1000 random genetic ancestors to locate by selecting a random sample, a random time, and a random position along the chromosome. We estimated the locations of these ancestors using our method (applied using six different window sizes) and compared to the averaging-up approach (Wohns *et al.*, 2022). For the averaging-up method, we first simplified our ARG to its corresponding succinct tree sequence (Kelleher *et al.*, 2018; Wong *et al.*, 2023).

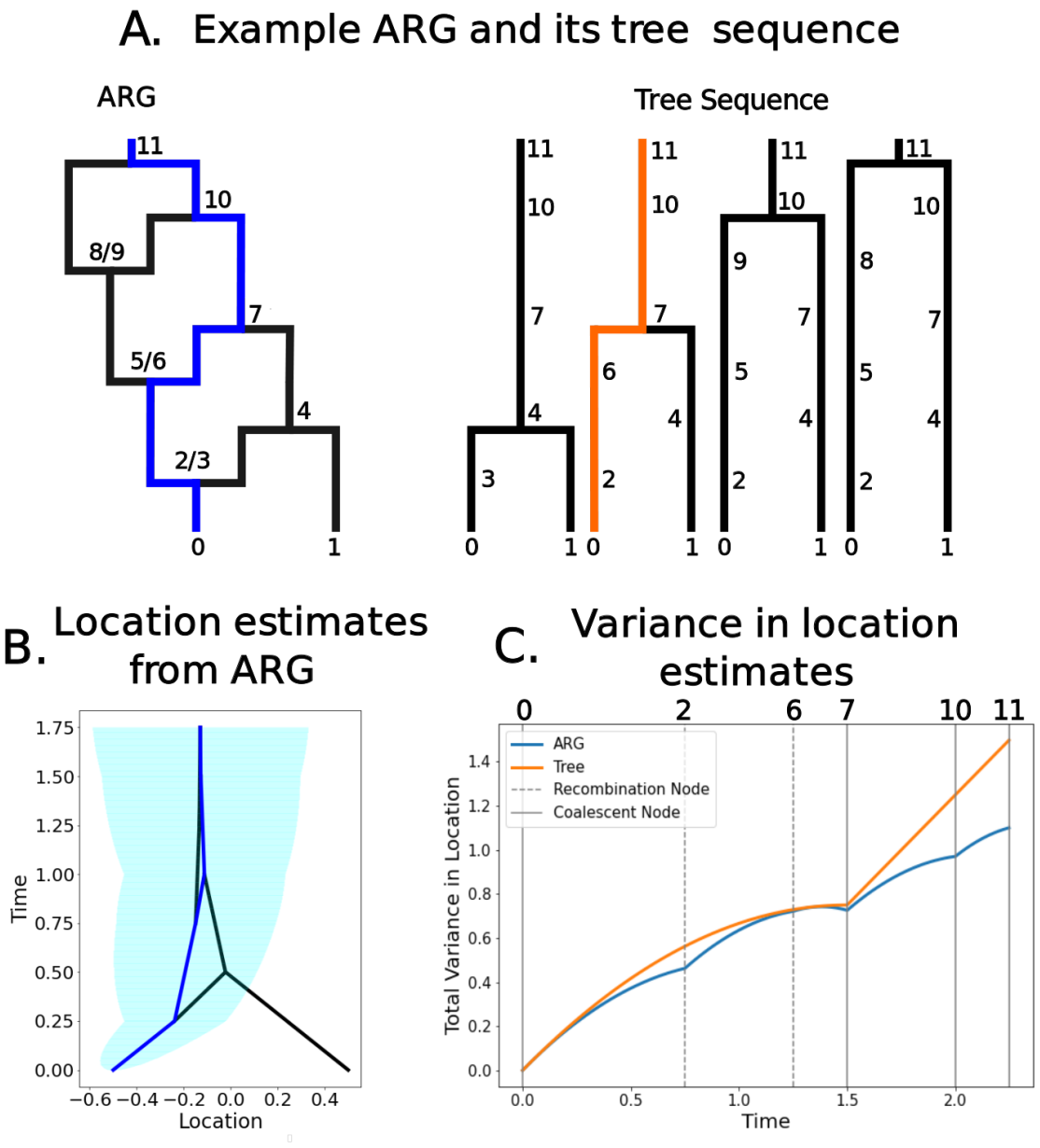

Figure 5: **Uncertainty in ancestor locations**. (A) Example ARG and the corresponding tree sequence with the same path highlighted. (B) The most likely location estimates (lines) and the 95% confidence interval of the highlighted path (shading) using the ARG. (C) The variance in location estimates along the highlighted path using the full ARG (blue) and using just the single tree (orange).

We first calculated the error between estimates of the most likely location of an ancestor and its true location (Figure 6A). Interestingly, we see that the average error of our method increases as we use more trees (increasing window size). This turns out to be caused by an increasing bias in ancestral locations estimates towards the average sample location (Figure S2). As we explained in the dispersal section above, including more trees leads to more compromises in internal node locations, generally pushing them faster towards the average sample location. While the internal node locations should generally converge to the average sample location eventually, the compromises imposed by our model are too strong and as a result the window size with the lowest average error is a single tree. Perhaps surprisingly, we see that the relatively simple averaging-up method, which uses a simplified ARG and ignores edge lengths, performs just as well on average.

We next estimated the variance in ancestor locations, using the true dispersal rate (since our estimated dispersal rates are biased). As described above for toy ARGs, using more information (trees) reduces our predicted uncertainty in ancestor locations (Figure 6B). Comparing the expected and observed coverage of confidence intervals, the higher variance from small window sizes overestimates uncertainty while the lower variance from large window sizes underestimates uncertainty (Figure 6C). Because our model is an approximation of the simulation, no window size perfectly estimates uncertainty. The averaging-up method does not provide a metric of uncertainty.

3.2.4 Visualizing a Recombination Event

One particularly promising application of our ARG-based inference is to locate recombination events, as well as the lineages involved in these events. This would, for example, allow us to more fully visualize the geographic history of admixed individuals. To demonstrate this, we set up a simulation that starts with two geographically isolated subpopulations. Over 1000 generations, individuals in the two subpopulations

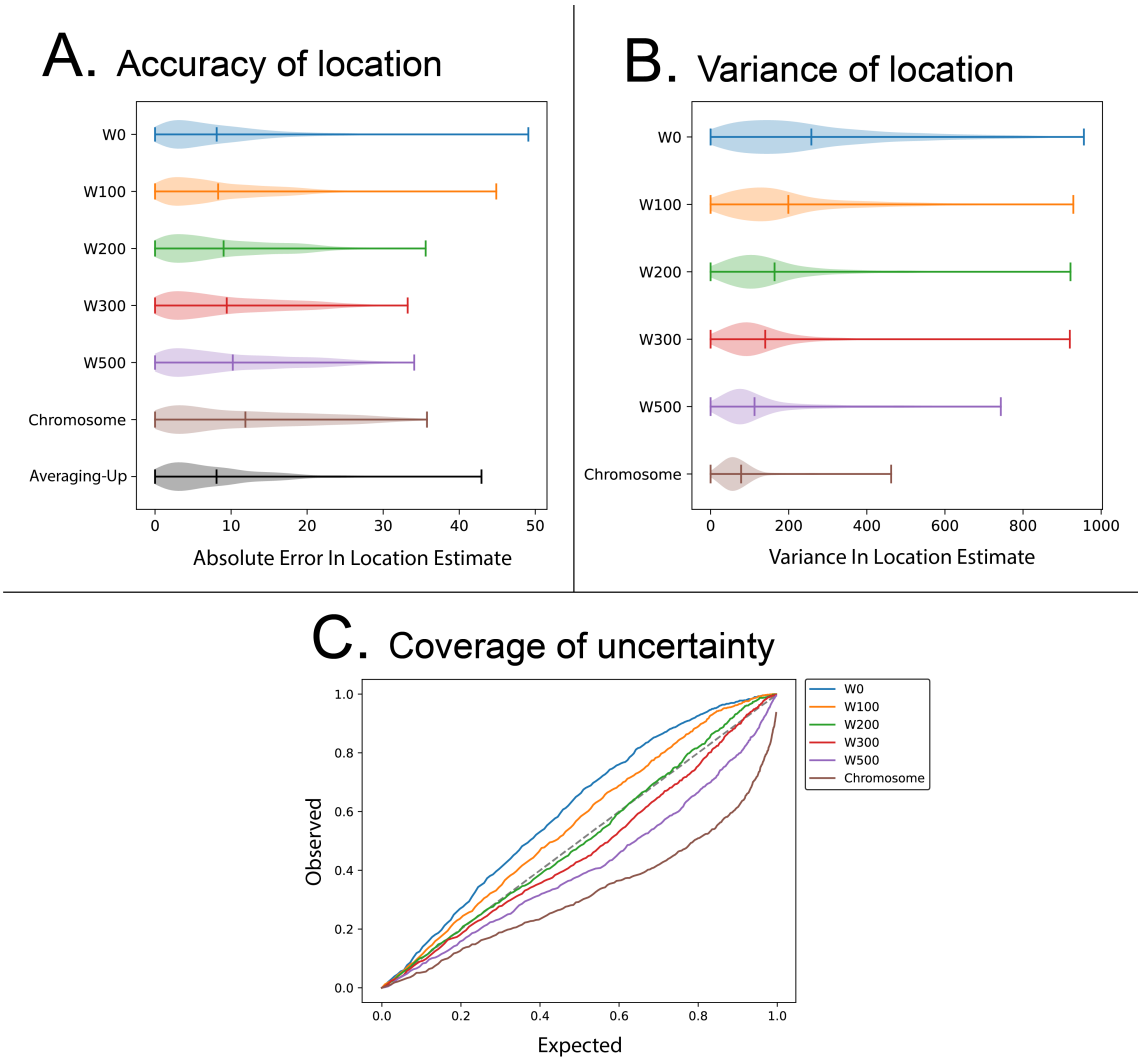

Figure 6: **Accuracy of ancestral location estimates**. We simulated an ARG and chose 1000 random ancestors. (A) Error in the most likely ancestor location estimates. Horizontal lines show the extremes and mean of each distribution. (B) Estimated variance in ancestor locations. (C) Expected versus observed coverage of confidence intervals. The dashed line is the 1:1 ratio. As our dispersal estimates are biased the variance, and therefore also the confidence intervals, were calculated using the true dispersal rate.

disperse and meet. At the end of the simulation there is still a clear cline in genetic ancestry along the axis of separation between the original two sub-populations (Fig 7A). This cline is required for us to reconstruct the locations of genetic ancestors back to their original subpopulations. If this cline did not exist, meaning that the two sub-populations have completely mixed, then there would not be a signal of the original subpopulation separation. We then reconstructed the spatial history of a sample on either side of a recombination event that splits the subpopulation it inherited genetic material from. Using our approach, we estimate the ancestral locations reasonably well and capture the fact that the two lineages were in the same location below the recombination event (Figure 7B). In contrast, when we locate the same ancestors using only a local tree for each lineage, we fail to capture the recent intertwined history or explicitly locate the recombination event.

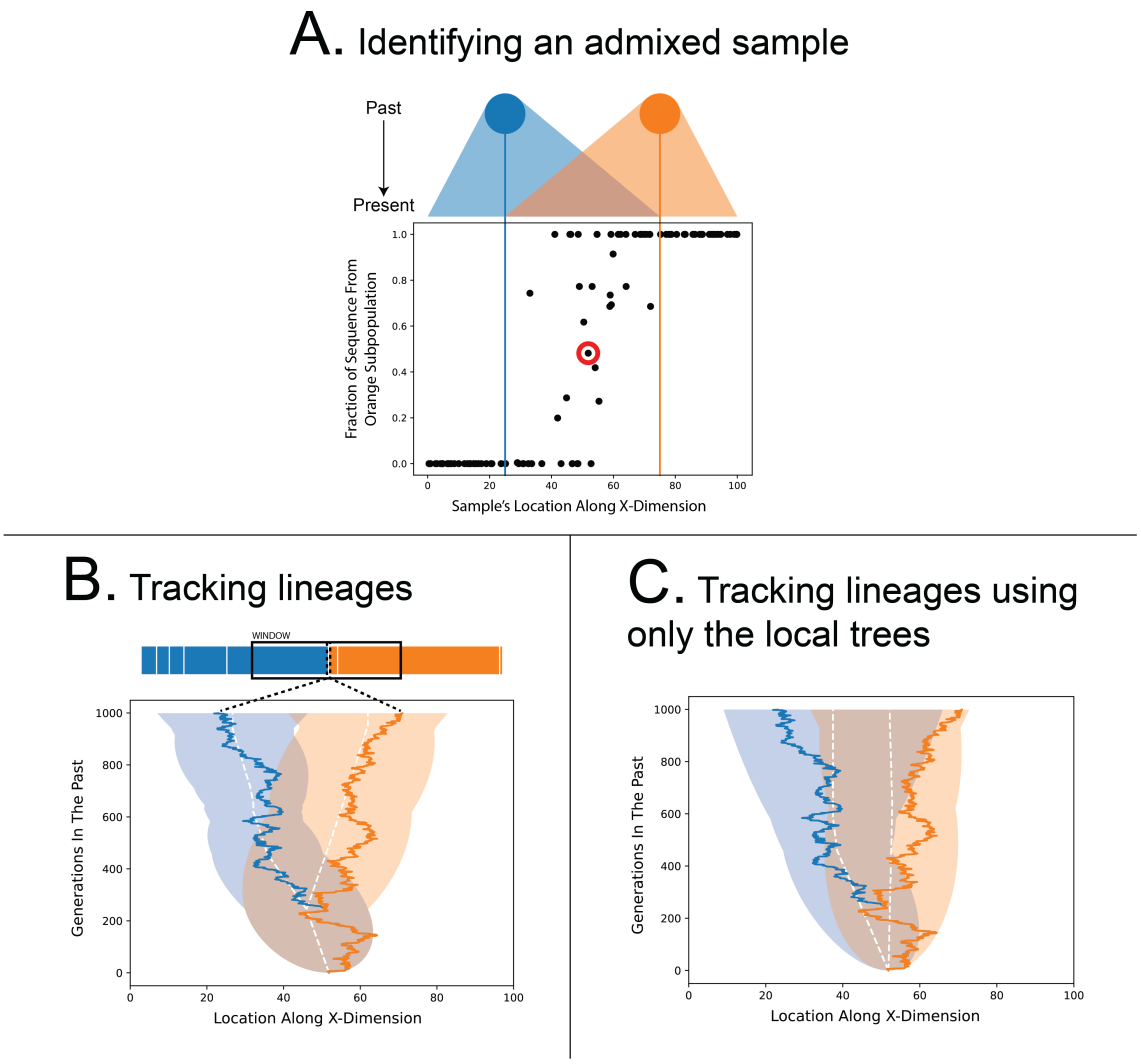

Figure 7: **Visualizing the geographic history of admixture**. We simulated two geographically isolated subpopulations dispersing into one another, leading to samples with genetic ancestors from both subpopulations. (A) At the top of subfigure is a cartoon of the dispersion of the subpopulations over time. The colored vertical lines mark the average starting position of each subpopulation along the X-dimension. As you look at samples positioned from left to right, the fraction of their sequence associated with the orange subpopulation increases. The circled sample is used as the example in subfigures B and C. (B) Regions of the circled sample's chromosome are painted according to their associated ancestral subpopulation. We mark a window of 202 trees centered on the recombination breakpoint of interest (W_{100}). Below the chromosome is a spatial reconstruction of the ancestral lineages on either side of the breakpoint. The solid curves mark the true location of the lineages, the dashed white lines are the estimated most likely ancestral locations, and the shading is the 95% confidence interval around these estimates (using the true dispersal rate). (C) Another ancestral reconstruction this time using only the local tree for each lineage. The local trees lack the necessary information to locate the recombination event.

4 Discussion

We have developed a method for estimating dispersal rates and locating genetic ancestors from ancestral recombination graphs. By using the full ARG for inference, our method can make use of the complete genetic history of the samples. After extending a model of Brownian motion from trees to graphs, we developed an efficient algorithm that scales to large ARGs (Figures 1 and 2). We then showed that our method, while appropriately handling non-independence across trees, results in biased dispersal estimates, in part due to excess constraint on internal node locations (Figure 3). When locating ancestors, we show that our method uses more information than previous methods (Figure 4) and therefore locates with less uncertainty (Figure 5). However, the excess constraint on internal node locations again leads to a bias, but this bias can be reduced by analyzing smaller genomic windows (Figure 6). An important application of our method is that we can geographically track a sample's genetic ancestry back in time as it splits and merges via recombination and coalescence. This allows us to locate recombination events of interest and visualize admixed geographic ancestries (Figure 7).

Using information from multiple trees, our method estimates the location of ancestors with more confidence. This may help improve our ability to infer historical migrations, e.g., in humans. Further, appropriately modeling the splitting of recombinant lineages and thereby locating recombination events has multiple potential applications. As we have shown, it can provide a more complete picture of the geographic ancestry of admixed individuals (Fig 7). Recombination can also bring advantageous alleles together on the same background, and so identifying where recombinant haplotypes arose can be crucial to understanding the geography of adaptation. For example, new strains of circulating viruses and pathogens are often a consequence of recombination between existing strains, so being able to identify where those recombinant strains

first arose is an important step in understanding the spread of pathogens (Ignatieva *et al.*, 2022; Tamura *et al.*, 2023).

Brownian motion is a convenient but rough approximation for the movement of genetic ancestors down an ARG. For the case of trees, it is well known that the assumptions of constant global population size and independent branching Brownian motions lead to a clustering of individuals (Barton *et al.*, 2010a; Felsenstein, 1975), not the uniform distribution we see in our simulations and expect in nature as a result of local density-dependence. Further, our simple model of Brownian motion assumes an unbounded space. As a result of the latter, dispersal estimates from trees generated in finite space show a systematic downward bias (Figure 3A and Ianni-Ravn *et al.*, 2023; Kalkauskas *et al.*, 2021). Modeling Brownian motion down ARGs includes an additional approximation, the meeting of lineages at recombination nodes. Here we assume the two lineages precisely meet, which is an unlikely event (Etheridge, 2019), especially if they were ever far apart. This implies that the two lineages must not have dispersed very far from one another since their common ancestor, pulling all nodes in and below a recombination loop closer together (Figure 3C). This is another way of describing an increased constraint on internal node locations (see Results). Therefore, given the same sample locations, the addition of a recombination loop increases our dispersal estimate. As every additional tree comes with an additional recombination loop, our dispersal estimate increases with the number of trees (Figure 3A). This pulling in of node locations below recombination loops also explains our bias in ancestor locations (Figures 6 and S2).

One way to get less biased parameter estimates is to use a more accurate model. An obvious choice is the spatial Λ-Fleming-Viot process (Barton *et al.*, 2010a,b). This approach models local density dependence and a finite habitat, leading to more accurate inferences than simple Brownian motion from trees (Kalkauskas *et al.*, 2021).

The spatial Λ-Fleming-Viot process also naturally incorporates recombination (Barton *et al.*, 2010b; Etheridge and Véber, 2012) between lineages that are not at the exact same location. However, the complexity of this model makes it computationally intensive for inference and therefore limited to small sample sizes, even in the case of no recombination (Wirtz and Guindon, 2023). It may then be helpful to think of ways to improve the more tractable and scalable model of Brownian motion. Here we focus on the main hurdle in improving estimates from ARGs, the excess constraint we see in internal node locations.

Assuming that randomly moving lineages must precisely meet at recombination nodes is excessive, unlikely, and generates a lot of constraint. In our simulations individuals mate with others nearby according to a mating kernel. This could describe, for example, pollen/gamete dispersal or non-random movements between nearby individuals that come together to mate. We therefore explored the alternative assumption, that lineages do not have to precisely meet at recombination nodes (Appendix D). To make this alternative tractable we assumed the opposite extreme, that the two parents of a recombination node can be any distance apart, and placed the recombination node at their midpoint (an approach used for phenotypic evolution in phylogenetic networks Bastide *et al.*, 2018). Not forcing the parental lineages to meet greatly decreases the constraint on internal node locations across trees, leading to lower dispersal estimates (Figure S1). However, the dispersal estimate still increases with the number of trees, indicating that some excess constraint remains.

The other alternative we explored to reduce excess constraint on internal node locations was to use only a portion (window) of the available sequence. This approach strikes a compromise between using the full ARG from the entire sequence, which exhibits too much constraint, vs. assuming each local tree is spatially independent, which loses information. We show that windowing consequently produces dispersal estimates (Figure 3) and distributions of ancestor locations (Figure 6) that fall between

those estimated using the full ARG and those assuming independent trees. Given that local trees locate ancestors with less bias than the full ARG (in our simulations) but ignore recombination events, a key advantage of the windowing approach is to locate recombination events and visualize the geographic history of admixed samples (Figure 7). However, there are some caveats to this approach. Most importantly, it is not possible in practice to know a suitable window size for the parameter you want to estimate. There is also additional computation burden as our algorithm needs to run separately for every window considered (e.g., for ancestors centered on different local trees).

Here we have developed a model and algorithm for inferring spatial histories from ARGs and tested our method on simulated data but additional issues arise in the application to empirical data. One is that our method is designed to be applied on the full ARG, a single graph with recombination loops. `ARGweaver` (Rasmussen *et al.*, 2014) infers full ARGs on which our method can be applied. However, the methods that scale to larger sample sizes and sequence lengths, `tsinfer`+`tsdate` (Kelleher *et al.*, 2019; Wohns *et al.*, 2022) and `Relate` (Speidel *et al.*, 2019), lack some information contained in the full ARG, e.g., they both lack marked recombination nodes. While our method could be immediately applied to a simplified ARG, it is not clear if our model of motion is appropriate (e.g., we would then force lineages to meet not at a recombination node but at the coalescent node below it) or if it could be modified to suit. Scaling-up the inference of full ARGs (e.g., Deng *et al.*, 2024) will allow our method to be applied to larger datasets.

The main hurdle to applying our method to real datasets, even small ones, is that we have assumed that we know the ARG with complete certainty. This is not possible in practice. To incorporate uncertainty in ARG inference we could use importance sampling (Osmond and Coop, 2021), which would require knowing the probability of an ARG under both the assumptions used to infer it and the assumptions of our spatial

model. Both `ARGweaver` and `SINGER` provide the probability of each sampled ARG under their panmictic assumptions. The remaining hurdle is to derive the probability of an ARG under our spatial model.

In summary, we have developed a mathematically rigorous and computationally efficient method that uses the complete genealogical history of a set of samples to reconstruct the spatial history of their genetic ancestors. While there is some bias in our estimates due to model mis-specification, we can usefully estimate the locations of recombination nodes, allowing us to locate important recombination events or visualize the geographic history of admixed genomes.

5 Acknowledgements

We thank Mete Yuksel, Chris Carlson, and the Coop lab for helpful conversations. This work was supported by the Natural Sciences and Engineering Research Council of Canada (RGPIN–2021-03207 and DGECR-2021-00114 to MO), the National Institutes of Health (NIH R35 GM136290 awarded to GC), and the National Science Foundation (NSF DISES 2307175 to GC). Computations were performed on the Niagara supercomputer at the SciNet HPC Consortium. SciNet is funded by: the Canada Foundation for Innovation; the Government of Ontario; Ontario Research Fund – Research Excellence; and the University of Toronto.

References

Barton N, Etheridge A, Véber A. 2010a. A new model for evolution in a spatial continuum. Electronic Journal of Probability. .

Barton NH, Etheridge AM, Véber A. 2013. Modelling evolution in a spatial continuum. Journal of Statistical Mechanics: Theory and Experiment. 2013:P01002.

Barton NH, Kelleher J, Etheridge AM. 2010b. A new model for extinction and recolonization in two dimensions: quantifying phylogeography. Evolution. 64:2701–2715.

Bastide P, Solís-Lemus C, Kriebel R, William Sparks K, Ané C. 2018. Phylogenetic Comparative Methods on Phylogenetic Networks with Reticulations. Systematic Biology. 67:800–820.

Baumdicker F, Bisschop G, Goldstein D, Gower G, Ragsdale AP, Tsambos G, Zhu S, Eldon B, Ellerman EC, Galloway JG *et al*. 2022. Efficient ancestry and mutation simulation with msprime 1.0. Genetics. 220:iyab229.

Dellicour S, Gill MS, Faria NR, Rambaut A, Pybus OG, Suchard MA, Lemey P. 2021. Relax, keep walking—a practical guide to continuous phylogeographic inference with beast. Molecular Biology and Evolution. 38:3486–3493.

Deng Y, Nielsen R, Song YS. 2024. Robust and Accurate Bayesian Inference of Genome-Wide Genealogies for Large Samples. bioRxiv. .

Etheridge A. 2019. Spatial population models. Ensaios Matematicos. 34:75–186.

Etheridge A, Véber A. 2012. The spatial λ-fleming–viot process on a large torus: genealogies in the presence of recombination. .

Excoffier L, Marchi N, Marques DA, Matthey-Doret R, Gouy A, Sousa VC. 2021. fastsimcoal2: demographic inference under complex evolutionary scenarios. Bioinformatics. 37:4882–4885.

Fan C, Cahoon JL, Dinh BL, Ortega-Del Vecchyo D, Huber C, Edge MD, Mancuso N, Chiang CW. 2023. A likelihood-based framework for demographic inference from genealogical trees. bioRxiv. .

Felsenstein J. 1975. A Pain in the Torus: Some Difficulties with Models of Isolation by Distance. The American Naturalist. 109:359–368.

Felsenstein J. 1985. Phylogenies and the Comparative Method. The American Naturalist. 125:1–15.

Griffiths R, Marjoram P. 1996. Ancestral Inference from Samples of DNA Sequences with Recombination. Journal of Computational Biology. 3:479–502.

Grundler M, Terhorst J, Bradburd G. 2024. A geographic history of human genetic ancestry. unpulished. .

Guo F, Carbone I, Rasmussen DA. 2022. Recombination-aware phylogeographic inference using the structured coalescent with ancestral recombination. PLOS Computational Biology. 18:e1010422.

Haller BC, Galloway J, Kelleher J, Messer PW, Ralph PL. 2019. Tree-sequence recording in slim opens new horizons for forward-time simulation of whole genomes. Molecular ecology resources. 19:552–566.

Haller BC, Messer PW. 2023. Slim 4: multispecies eco-evolutionary modeling. The American Naturalist. 201:E127–E139.

Harris K. 2019. From a database of genomes to a forest of evolutionary trees. Nature genetics. 51:1306–1307.

Hejase HA, Dukler N, Siepel A. 2020. From summary statistics to gene trees: methods for inferring positive selection. Trends in Genetics. 36:243–258.

Hudson RR. 1983. Properties of a neutral allele model with intragenic recombination. Theoretical Population Biology. 23:183–201.

Ianni-Ravn MK, Petr M, Racimo F. 2023. Exploring the effects of ecological parameters on the spatial structure of genetic tree sequences. bioRxiv. .

Ignatieva A, Hein J, Jenkins PA. 2022. Ongoing Recombination in SARS-CoV-2 Revealed through Genealogical Reconstruction. Molecular Biology and Evolution. 39:msac028.

Kalkauskas A, Perron U, Sun Y, Goldman N, Baele G, Guindon S, De Maio N. 2021. Sampling bias and model choice in continuous phylogeography: Getting lost on a random walk. PLOS Computational Biology. 17:e1008561.

Kelleher J, Thornton KR, Ashander J, Ralph PL. 2018. Efficient pedigree recording for fast population genetics simulation. PLOS Computational Biology. 14:e1006581.

Kelleher J, Wong Y, Wohns AW, Fadil C, Albers PK, McVean G. 2019. Inferring whole-genome histories in large population datasets. Nature genetics. 51:1330–1338.

Kimura M, Weiss GH. 1964. The stepping stone model of population structure and the decrease of genetic correlation with distance. Genetics. 49:561.

Larribe F, Fearnhead P. 2011. On Composite Likelihoods in Statistical Genetics. Technical Report 1.

Lemmon AR, Lemmon EM. 2008. A likelihood framework for estimating phylogeographic history on a continuous landscape. Systematic biology. 57:544–561.

Lewanski AL, Grundler MC, Bradburd GS. 2023. The era of the arg: an empiricist's guide to ancestral recombination graphs. arXiv preprint arXiv:2310.12070. .

Maddison WP. 1991. Squared-change parsimony reconstructions of ancestral states for continuous-valued characters on a phylogenetic tree. Systematic Biology. 40:304–314.

Malécot G. 1948. The mathematics of heredity. The mathematics of heredity.. .

Meligkotsidou L, Fearnhead P. 2007. Postprocessing of Genealogical Trees. Genetics. 177:347–358.

Müller NF, Rasmussen D, Stadler T. 2018. Mascot: parameter and state inference under the marginal structured coalescent approximation. Bioinformatics. 34:3843–3848.

Novembre J, Slatkin M. 2009. Likelihood-based inference in isolation-by-distance models using the spatial distribution of low-frequency alleles. Evolution. 63:2914–2925.

Osmond MM, Coop G. 2021. Estimating dispersal rates and locating genetic ancestors with genome-wide genealogies. bioRxiv. .

Rasmussen MD, Hubisz MJ, Gronau I, Siepel A. 2014. Genome-Wide Inference of Ancestral Recombination Graphs. PLoS Genetics. 10:e1004342.

Rousset. 2000. Genetic differentiation between individuals. Journal of Evolutionary Biology. 13:58–62.

Rousset F. 1997. Genetic differentiation and estimation of gene flow from f-statistics under isolation by distance. Genetics. 145:1219–1228.

Schaefer NK, Shapiro B, Green RE. 2021. An ancestral recombination graph of human, neanderthal, and denisovan genomes. Science advances. 7:eabc0776.

Shipilina D, Pal A, Stankowski S, Chan YF, Barton NH. 2023. On the origin and structure of haplotype blocks. Molecular Ecology. 32:1441–1457.

Smith CC, Tittes S, Ralph PL, Kern AD. 2023. Dispersal inference from population genetic variation using a convolutional neural network. Genetics. 224:iyad068.

Speidel L, Forest M, Shi S, Myers SR. 2019. A method for genome-wide genealogy estimation for thousands of samples. Nature genetics. 51:1321–1329.

Tamura T, Ito J, Uriu K, Zahradnik J, Kida I, Anraku Y, Nasser H, Shofa M, Oda Y, Lytras S *et al*. 2023. Virological characteristics of the SARS-CoV-2 XBB variant derived from recombination of two Omicron subvariants. Nature Communications. 14:2800. Publisher: Nature Publishing Group.

Wakeley J. 1999. Nonequilibrium Migration in Human History. Genetics. 153:1863–1871.

Wirtz J, Guindon S. 2023. On the connections between the spatial Lambda-Fleming-Viot model and other processes for analysing geo-referenced genetic data. arXiv preprint arXiv:2307.02431. .

Wohns AW, Wong Y, Jeffery B, Akbari A, Mallick S, Pinhasi R, Patterson N, Reich D, Kelleher J, McVean G. 2022. A unified genealogy of modern and ancient genomes. Science. 375.

Wong Y, Ignatieva A, Koskela J, Gorjanc G, Kelleher J. 2023. A general and efficient representation of Ancestral Recombination Graphs. bioRxiv. .

Wright S. 1943. Isolation by distance. Genetics. 28:114–138.

Zhang BC, Biddanda A, Gunnarsson ÁF, Cooper F, Palamara PF. 2023. Biobank-scale inference of ancestral recombination graphs enables genealogical analysis of complex traits. Nature Genetics. 55:768–776.

A Likelihood of sample locations

Notation: For locations, we denote random variables by italicized uppercase letters, the values they take by the corresponding lowercase letters, and use asterisks to denote the observed values. Further, matrices are denoted by bold capital letters and parameters are denoted by greek letters.

We start with assuming that the displacement along each edge of the ARG, B_{edge}, is independent. Hence each path, from a sample to the root, has an unique distribution for the location of the sample, even if they end at the same sample. We denote this set of locations by the n_p length random vector, $\overrightarrow{L_p}$. In order to get the actual distribution of sample locations we condition on η_{loops}, i.e., we need to get the conditional probability density

$$f_{\overrightarrow{L_p}|\eta_{loops}}(\vec{\ell_p}) = \frac{f_{\overrightarrow{L_p}}(\vec{\ell_p} \cap \eta_{loops})}{f_{\overrightarrow{L_p}}(\eta_{loops})}. \tag{10}$$

However, we have that $\eta_{loops} = \eta_{paths}$ (appendix C.1). And since η_{paths} is the condition that paths that end at the same samples have identical locations, the numerator becomes $f_{\overrightarrow{L_p}}(\vec{\ell_p} \cap \eta_{loops}) = f_{\overrightarrow{L_p}}(\vec{\ell_p} \cap \eta_{paths}) = f_{\overrightarrow{L_p}}(\mathbf{P}_p\vec{\ell}\,)$, where $\mathbf{P}_p$ is the path-sample matrix which is a $n_p \times n_s$ matrix whose ij^{th} entry is 1 if the ith path ends at sample j and $\vec{\ell}$ is a vector of length n_s that has the location of each sample. Therefore, we have

$$f_{\overrightarrow{L_p}|\eta_{loops}}(\vec{\ell_p}) = \mathbb{1}[\vec{\ell_p} = \mathbf{P}_p\vec{\ell}\,]\frac{f_{\overrightarrow{L_p}}(\mathbf{P}_p\vec{\ell}\,)}{f_{\overrightarrow{L_p}}(\eta_{paths})}, \tag{11}$$

where $\mathbb{1}$ is the indicator function which is 1 if $\vec{\ell_p} = \mathbf{P}_p\vec{\ell}$ and 0 otherwise. This basically ensures the probability density is 0 whenever any two paths ending at the same sample have different locations. We will now compute this distribution for the case of single root and then do the more general multiple root case.

A.1 Single Root

For the case where the ARG has a single root, located at μ, we have that the location of path ends, $\overrightarrow{L_p}$, is distributed multivariate normally with mean $\mu\mathbb{1}_{n_p}$ and covariance

matrix $\sigma^2\mathbf{S}_p$. Therefore, the numerator of Eq 11,

$$f_{\overrightarrow{L_p}}(\mathbf{P}_p\vec{\ell}\,) = \int \frac{1}{\sqrt{(2\pi\sigma^2)^{\text{rnk }\mathbf{S}_p}|\mathbf{S}_p|}}\exp\left(-\frac{(\vec{\ell}\,-\mu\mathbb{1}_{n_s})^T\mathbf{P}_p^T\mathbf{S}_p^{g-}\mathbf{P}_p(\vec{\ell}\,-\mu\mathbb{1}_{n_s})}{2\sigma^2}\right)d\vec{\ell} \quad (12)$$

We also evaluate the denominator in Eq 11,

$$\begin{aligned} f_{\overrightarrow{L_p}}(\eta_{paths}) &= \int f_{\overrightarrow{L_p}}(\vec{\ell}_p \cap \eta_{paths})d\vec{\ell}_p && (13)\\ &= \int f_{\overrightarrow{L_p}}(\mathbf{P}_p\vec{\ell}\,)d\vec{\ell} && (14)\\ &= \int \frac{1}{\sqrt{(2\pi\sigma^2)^{\text{rnk }\mathbf{S}_p}|\mathbf{S}_p|}}\exp\left(-\frac{(\vec{\ell}\,-\mu\mathbb{1}_{n_s})^T\mathbf{P}_p^T\mathbf{S}_p^{g-}\mathbf{P}_p(\vec{\ell}\,-\mu\mathbb{1}_{n_s})}{2\sigma^2}\right)d\vec{\ell} && (15)\\ &= \frac{\sqrt{(2\pi\sigma^2)^{n_s}|\mathbf{S}|}}{\sqrt{(2\pi\sigma^2)^{\text{rnk }\mathbf{S}_p}|\mathbf{S}_p|}}, && (16) \end{aligned}$$

where $\mathbf{S} = (\mathbf{P}_p^T\mathbf{S}_p^{g-}\mathbf{P}_p)^{-1}$ is the sample covariance matrix.

The probability density of the path locations, conditional on the loops (paths) meeting, is then

$$f_{\overrightarrow{L_p}|\eta_{loops}}(\vec{\ell_p}) = \mathbb{1}[\vec{\ell_p} = \mathbf{P}_p\vec{\ell}\,]\frac{1}{\sqrt{(2\pi\sigma^2)^{n_s}|\mathbf{S}|}}\exp\left(-\frac{(\vec{\ell}\,-\mu\mathbb{1}_{n_s})^T\mathbf{S}^{-1}(\vec{\ell}\,-\mu\mathbb{1}_{n_s})}{2\sigma^2}\right), \quad (17)$$

which is the probability density of a multivariate normal random variable with mean $\mu\mathbb{1}_{n_s}$ and covariance matrix $\sigma^2\mathbf{S}$. This is the likelihood of sample locations, $\overrightarrow{L}$, given Brownian motion down the ARG. The MLE of the dispersal rate and root location is then given by

$$\begin{aligned} \widehat{\mu} &= (\mathbb{1}_{n_p}\mathbf{S}_p^{g-}\mathbb{1}_{n_p})^{-1}\mathbb{1}_{n_p}\mathbf{S}_p^{g-}\vec{\ell^*} && (18)\\ \widehat{\sigma^2} &= \frac{(\vec{\ell^*}-\mu\mathbb{1}_{n_s})^T\mathbf{P}_p^T\mathbf{S}_p^{g-}\mathbf{P}_p(\vec{\ell^*}-\mu\mathbb{1}_{n_s})}{n_s} && (19) \end{aligned}$$

A.2 Multiple Roots

We next want to generalize this to multiple roots, which occurs when we chop off an ARG at some time in the past. Let n_r be the number of roots, $\vec{\mu}$ be $n_r \times 1$ vector of root locations, and $\mathbf{R}_p$ be the $n_p \times n_r$ path-root matrix defined in table 1. Then the probability distribution of the path locations is multivariate normal with mean $\mathbf{R}_p\vec{\mu}$ and covariance $\sigma^2\mathbf{S}_p$. Now, as in the single root case we want to find the distribution of the path locations conditioned on the paths meeting at the samples (equation 10).

As with a single root, the numerator can be written $f_{\overrightarrow{L_p}}(\vec{\ell_p} \cap \eta_{loops}) = f_{\overrightarrow{L_p}}(\vec{\ell_p} \cap \eta_{paths}) = f_{\overrightarrow{L_p}}(\mathbf{P}_p\vec{\ell}\,)$, but now this is

$$f_{\overrightarrow{L_p}}(\mathbf{P}_p\vec{\ell}\,) = \frac{\exp[-\frac{1}{2\sigma^2}(\mathbf{P}_p\vec{\ell}\,-\mathbf{R}_p\vec{\mu})^{\mathrm{T}}\mathbf{S}_p^{g-}(\mathbf{P}_p\vec{\ell}\,-\mathbf{R}_p\vec{\mu})]}{\sqrt{(2\pi\sigma^2)^p|\mathbf{S}_p|}}. \tag{20}$$

Similarly, the denominator becomes

$$\begin{aligned} f_{\overrightarrow{L_p}}(\eta_{paths}) &= \int_{\mathbb{R}^n} f_{\overrightarrow{L_p}}(\mathbf{P}_p\vec{\ell}\,)d\vec{\ell} && (21)\\ &= \int_{\mathbb{R}^n} \frac{\exp[-\frac{1}{2\sigma^2}(\mathbf{P}_p\vec{\ell}\,-\mathbf{R}_p\vec{\mu})^{\mathrm{T}}\mathbf{S}_p^{g-}(\mathbf{P}_p\vec{\ell}\,-\mathbf{R}_p\vec{\mu})]}{\sqrt{(2\pi\sigma^2)^p|\mathbf{S}_p|}}d\vec{\ell}\,. && (22)\end{aligned}$$

To find this integral we will multiply and divide by a constant to make the integrand a probability density for $\vec{\ell}$. In order to do that, note that the term in the exponent

can be expanded like

$$
\begin{aligned}
& (\mathbf{P}_p\vec{\ell} - \mathbf{R}_p\vec{\mu})^{\mathrm{T}}\mathbf{S}_p^{g-}(\mathbf{P}_p\vec{\ell} - \mathbf{R}_p\vec{\mu}) && (23)\\
= \; & \vec{\ell}^{\,\mathrm{T}}\mathbf{P}_p^{\mathrm{T}}\mathbf{S}_p^{g-}\mathbf{P}_p\vec{\ell} - 2\vec{\mu}^{\mathrm{T}}\mathbf{R}_p^{\mathrm{T}}\mathbf{S}_p^{g-}\mathbf{P}_p\vec{\ell} + \vec{\mu}^{\mathrm{T}}\mathbf{R}_p^{\mathrm{T}}\mathbf{S}_p^{g-}\mathbf{R}_p\vec{\mu} && (24)\\
= \; & \vec{\ell}^{\,\mathrm{T}}[\mathbf{P}_p,\mathbf{P}_p]\vec{\ell} - 2\vec{\mu}^{\mathrm{T}}[\mathbf{R}_p,\mathbf{P}_p]\vec{\ell} + \vec{\mu}^{\mathrm{T}}[\mathbf{R}_p,\mathbf{R}_p]\vec{\mu} && (25)\\
= \; & \vec{\ell}^{\,\mathrm{T}}[\mathbf{P}_p,\mathbf{P}_p]\vec{\ell} - 2\vec{\mu}^{\mathrm{T}}[\mathbf{R}_p,\mathbf{P}_p][\mathbf{P}_p,\mathbf{P}_p]^{-1}[\mathbf{P}_p,\mathbf{P}_p]\vec{\ell} + \vec{\mu}^{\mathrm{T}}[\mathbf{R}_p,\mathbf{R}_p]\vec{\mu} && (26)\\
= \; & \vec{\ell}^{\,\mathrm{T}}[\mathbf{P}_p,\mathbf{P}_p]\vec{\ell} - 2\vec{\mu}_\ell^{\mathrm{T}}[\mathbf{P}_p,\mathbf{P}_p]\vec{\ell} + \vec{\mu}^{\mathrm{T}}[\mathbf{R}_p,\mathbf{R}_p]\vec{\mu} && (27)\\
= \; & \vec{\ell}^{\,\mathrm{T}}[\mathbf{P}_p,\mathbf{P}_p]\vec{\ell} - 2\vec{\mu}_\ell^{\mathrm{T}}[\mathbf{P}_p,\mathbf{P}_p]\vec{\ell} + \vec{\mu}_\ell^{\mathrm{T}}[\mathbf{P}_p,\mathbf{P}_p]\vec{\mu}_\ell - \vec{\mu}_\ell^{\mathrm{T}}[\mathbf{P}_p,\mathbf{P}_p]\vec{\mu}_\ell + ... && (28)\\
& ... \; \vec{\mu}^{\mathrm{T}}[\mathbf{R}_p,\mathbf{R}_p]\vec{\mu} && (29)\\
= \; & (\vec{\ell} - \vec{\mu}_\ell)^{\mathrm{T}}[\mathbf{P}_p,\mathbf{P}_p](\vec{\ell} - \vec{\mu}_\ell) - \vec{\mu}_\ell^{\mathrm{T}}[\mathbf{P}_p,\mathbf{P}_p]\vec{\mu}_\ell + \vec{\mu}^{\mathrm{T}}[\mathbf{R}_p,\mathbf{R}_p]\vec{\mu} && (30)\\
= \; & (\vec{\ell} - \vec{\mu}_\ell)^{\mathrm{T}}[\mathbf{P}_p,\mathbf{P}_p](\vec{\ell} - \vec{\mu}_\ell) + ... && (31)\\
& ... \; \vec{\mu}^{\mathrm{T}}\big([\mathbf{R}_p,\mathbf{R}_p] - [\mathbf{R}_p,\mathbf{P}_p][\mathbf{P}_p,\mathbf{P}_p]^{-1}[\mathbf{P}_p,\mathbf{R}_p]\big)\vec{\mu}. && (32)
\end{aligned}
$$

In step 1 above we have used $\vec{\ell}^{\,\mathrm{T}}\mathbf{P}_p^{\mathrm{T}}\mathbf{S}_p^{g-}\mathbf{R}_p\vec{\mu} = \vec{\mu}^{\mathrm{T}}\mathbf{R}_p^{\mathrm{T}}\mathbf{S}_p^{g-}\mathbf{P}_p\vec{\ell}$, as these are 1×1 matrices and therefore are the transpose of each other. We have also used the shorthand $[\mathbf{A},\mathbf{B}] = \mathbf{A}^{\mathrm{T}}\mathbf{S}_p^{g-}\mathbf{B}$ and introduced $\vec{\mu}_\ell = [\mathbf{P}_p,\mathbf{P}_p]^{-1}[\mathbf{P}_p,\mathbf{R}_p]\vec{\mu}$, a $n\times1$ vector which will correspond to the expectation of the sample locations (as shown below). Using this expansion, we have

$$
\begin{aligned}
f_{\overrightarrow{L_p}}(\eta_{paths}) &= N_{old}\int_{\mathbb{R}^n}\exp\left[-\frac{1}{2\sigma^2}(\vec{\ell} - \vec{\mu}_\ell)^{\mathrm{T}}[\mathbf{P}_p,\mathbf{P}_p](\vec{\ell} - \vec{\mu}_\ell)\right]d\vec{\ell} && (33)\\
&= N_{old}N_{new}\int_{\mathbb{R}^n}\frac{\exp\left[\frac{-1}{2\sigma^2}(\vec{\ell} - \vec{\mu}_\ell)^{\mathrm{T}}[\mathbf{P}_p,\mathbf{P}_p](\vec{\ell} - \vec{\mu}_\ell)\right]}{\sqrt{(2\pi\sigma^2)^n|[\mathbf{P}_p,\mathbf{P}_p]^{-1}|}}d\vec{\ell} && (34)\\
&= N_{old}N_{new} && (35)
\end{aligned}
$$

where $N_{old} = \dfrac{\exp\left[-\dfrac{1}{2\sigma^2}\vec{\mu}^{\mathrm{T}}\big([\mathbf{R}_p,\mathbf{R}_p] - [\mathbf{R}_p,\mathbf{P}_p][\mathbf{P}_p,\mathbf{P}_p]^{-1}[\mathbf{P}_p,\mathbf{R}_p]\big)\vec{\mu}\right]}{\sqrt{(2\pi\sigma^2)^p|\mathbf{S}_p|}}$ and $N_{new} = \sqrt{(2\pi\sigma^2)^n|[\mathbf{P}_p,\mathbf{P}_p]^{-1}|}$.

Using this notations, we can also rewrite Eq 20 as

$$f_{\overrightarrow{L_p}}(\mathbf{P}_p\vec{\ell}\,) = N_{old}\exp\left[-\frac{1}{2\sigma^2}(\vec{\ell}\,-\vec{\mu}_\ell)^{\mathrm{T}}[\mathbf{P}_p,\mathbf{P}_p](\vec{\ell}\,-\vec{\mu}_\ell)\right] \tag{36}$$

Dividing numerator by denominator, the distribution of the path locations after conditioning on the loops (paths) meeting is

$$\begin{aligned} f_{\overrightarrow{L_p}|\eta_{loops}}(\vec{\ell_p}) &= \frac{f_{\overrightarrow{L_p}}(\mathbf{P}_p\vec{\ell}\,\cap\eta_{paths})}{f_{\overrightarrow{L_p}}(\eta_{paths})} && (37)\\ &= \frac{N_{old}\exp\left[-\frac{1}{2\sigma^2}(\vec{\ell}\,-\vec{\mu}_\ell)^{\mathrm{T}}[\mathbf{P}_p,\mathbf{P}_p](\vec{\ell}\,-\vec{\mu}_\ell)\right]}{N_{old}N_{new}} && (38)\\ &= \frac{\exp\left[-\frac{1}{2\sigma^2}(\vec{\ell}\,-\vec{\mu}_\ell)^{\mathrm{T}}[\mathbf{P}_p,\mathbf{P}_p](\vec{\ell}\,-\vec{\mu}_\ell)\right]}{N_{new}} && (39)\end{aligned}$$

This is a multivariate normal distribution with mean $\vec{\mu}_\ell = [\mathbf{P}_p,\mathbf{P}_p]^{-1}[\mathbf{P}_p,\mathbf{R}_p]\vec{\mu}$ and covariance $\sigma^2\mathbf{S} = \sigma^2[\mathbf{P}_p,\mathbf{P}_p]^{-1}$. This is the likelihood of sample locations, $\overrightarrow{L}$, given Brownian motion down the ARG.

A.2.1 Maximum likelihood parameter estimates

The log likelihood function for the parameters is then given by

$$\begin{aligned} \log L(\vec{\mu},\sigma^2) &= -\frac{1}{2\sigma^2}(\vec{\ell}\,-\vec{\mu}_\ell)^{\mathrm{T}}[\mathbf{P}_p,\mathbf{P}_p](\vec{\ell}\,-\vec{\mu}_\ell) - n\log\sigma + const && (40)\\ &= -\frac{1}{2\sigma^2}\left[\vec{\ell}^{\,\mathrm{T}}[\mathbf{P}_p,\mathbf{P}_p]\vec{\ell}\,-2\vec{\mu}_\ell^{\mathrm{T}}[\mathbf{P}_p,\mathbf{P}_p]\vec{\ell}\,+\vec{\mu}_\ell^{\mathrm{T}}[\mathbf{P}_p,\mathbf{P}_p]\vec{\mu}_\ell\right]... && (41)\\ &\quad ... - n\log\sigma + const. && (42)\end{aligned}$$

We can use this to find the MLE root locations by first differentiating the log likelihood function with respect to each root location

$$
\begin{aligned}
-2\sigma^2 \frac{\partial \log L}{\partial \mu_i} &= -2\frac{\partial \vec{\mu}_\ell^{\mathrm{T}}}{\partial \mu_i}[\mathbf{P}_p, \mathbf{P}_p]\vec{\ell} + \frac{\partial \vec{\mu}_\ell^{\mathrm{T}}[\mathbf{P}_p, \mathbf{P}_p]\vec{\mu}_\ell}{\partial \mu_i} && (43)\\
&= -2\frac{\partial \vec{\mu}_\ell^{\mathrm{T}}}{\partial \mu_i}[\mathbf{P}_p, \mathbf{P}_p]\vec{\ell} + \frac{\partial \vec{\mu}_\ell^{\mathrm{T}}}{\partial \mu_i}[\mathbf{P}_p, \mathbf{P}_p]\vec{\mu}_\ell + \vec{\mu}_\ell^{\mathrm{T}}[\mathbf{P}_p, \mathbf{P}_p]\frac{\partial \vec{\mu}_\ell}{\partial \mu_i} && (44)\\
&= -2\frac{\partial \vec{\mu}_\ell^{\mathrm{T}}}{\partial \mu_i}[\mathbf{P}_p, \mathbf{P}_p]\vec{\ell} + 2\frac{\partial \vec{\mu}_\ell^{\mathrm{T}}}{\partial \mu_i}[\mathbf{P}_p, \mathbf{P}_p]\vec{\mu}_\ell. && (45)
\end{aligned}
$$

Now, note that since $\vec{\mu}_\ell = [\mathbf{P}_p, \mathbf{P}_p]^{-1}[\mathbf{P}_p, \mathbf{R}_p]\vec{\mu}$ then $\dfrac{\partial \vec{\mu}_\ell}{\partial \mu_i} = [\mathbf{P}_p, \mathbf{P}_p]^{-1}[\mathbf{P}_p, \mathbf{R}_p]\vec{e_i}$, where $\vec{e_i}$ is the unit vector of length n with 1 in the ith position. Therefore, we have

$$
-2\sigma^2 \frac{\partial \log L}{\partial \mu_i} = -2\vec{e_i}^{\mathrm{T}}[\mathbf{R}_p, \mathbf{P}_p]\vec{\ell} + 2\vec{e_i}^{\mathrm{T}}[\mathbf{R}_p, \mathbf{P}_p][\mathbf{P}_p, \mathbf{P}_p]^{-1}[\mathbf{P}_p, \mathbf{R}_p]\vec{\mu} \qquad (46)
$$

Equating this to zero for all $i \in \{1, 2, ..., r\}$ we get the MLE root locations, $\widehat{\vec{\mu}}$, as the solutions to the following system of linear equations

$$
\begin{aligned}
[\mathbf{R}_p, \mathbf{P}_p][\mathbf{P}_p, \mathbf{P}_p]^{-1}[\mathbf{P}_p, \mathbf{R}_p]\vec{\mu} &= [\mathbf{R}_p, \mathbf{P}_p]\vec{\ell} && (47)\\
\Rightarrow \widehat{\vec{\mu}} &= \left([\mathbf{R}_p, \mathbf{P}_p][\mathbf{P}_p, \mathbf{P}_p]^{-1}[\mathbf{P}_p, \mathbf{R}_p]\right)^{-1}[\mathbf{R}_p, \mathbf{P}_p]\vec{\ell}^*. && (48)
\end{aligned}
$$

Note that using this equation to get the root locations leads to unexpected behaviour. Specifically, ancestor locations rapidly move away from each other as we go back in time (see Fig 8). This is probably because of the low probability of two Brownian motions meeting, forwards in time, when they start at different locations. The highest probability is for them to meet in the middle, which forces them to diverge backwards in time. To avoid this issue, we use the unconditional distribution of $\overrightarrow{L_p}$ (i.e., not conditioning on the paths meeting at the samples) to compute the MLE root

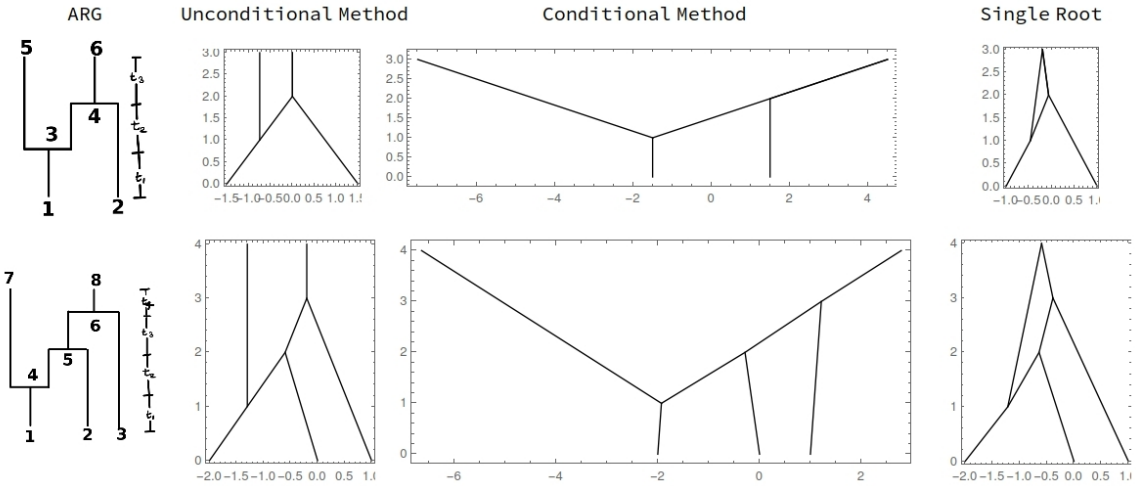

Figure 8: Root locations for two different ARG computed under the conditional and unconditional distributions.

locations, which gives

$$\widehat{\vec{\mu}} = (\mathbf{R}_p^{\mathrm{T}}\mathbf{S}_p^{g-}\mathbf{R}_p)^{-1}\mathbf{R}_p\mathbf{S}_p^{g-}\mathbf{P}_p\vec{\ell^*}.$$

This behaves as we expected, with ancestor locations that do not diverge (another panel of the figure we will put here). We leave a more complete understanding of why the conditioned distribution behaves unexpectedly to future work. Note that when there is a single root the conditional and unconditional MLE root locations are equal and collapse to the well-known MLE (Eq 18).

Differentiating the log likelihood function with respect to σ^2 and setting to zero, the MLE dispersal rate is

$$\widehat{\sigma^2} = \frac{(\vec{\ell^*} - \widehat{\vec{\mu}}_\ell)^{\mathrm{T}}\mathbf{S}^{-1}(\vec{\ell^*} - \widehat{\vec{\mu}}_\ell)}{n} \tag{49}$$

$$= \frac{(\vec{\ell^*} - \widehat{\vec{\mu}}_\ell)^{\mathrm{T}}\mathbf{P}_p\mathbf{S}_p^{g-}\mathbf{P}_p(\vec{\ell^*} - \widehat{\vec{\mu}}_\ell)}{n}. \tag{50}$$

B Equivalence of Loops and Paths Methods

Here we prove $\eta_{loops} = \eta_{paths}$ for any arbitrary ARG. Before providing a formal proof which requires more detailed notations, we outline the idea here.

1. In order to prove the equivalence, we need to show that for each condition in η_{loops} there exists an equivalent condition or set of conditions in η_{paths} and vice versa. In other words, for each loop we need to find a pair of paths that only differ inside that loop. And conversely, for every pair of paths from the same sample, we need to find a set of loops such that any difference in the paths belongs to one of the loops.

2. Given a loop, we find a pair of paths as follows

   (a) Find the bottom and top of the loop. This is where the loop begins and ends.

   (b) Find a path from the bottom to one of the samples. This exists because every node is connected to at least one sample.

   (c) Find a path from the top to one of the roots. This exists because every node is connected to at least one root.

   (d) To get the two paths, insert a different path from the loop in between the two paths found in the above two steps.

   (e) These two paths now only differ in the edges contained in the loop.

3. Given a pair of paths starting at the same sample, here is how you find the set of loops that are equivalent.

   (a) Start from the sample and move up one node at a time until you hit a node that is different in the two paths. This is the beginning of the first loop.

(b) Now, find the first node after the beginning of the first loop that is common to the two paths. This is the end of the first loop.

(c) If the two paths are identical after the end of the first loop, then we have found the equivalent loop condition.

(d) If not, repeat the above steps starting from the end of the first loop to find the next loop and so on. Since all paths end at roots, this process will stop and we will have a set of loops that are equivalent to the pair of paths.

B.1 Formal Notations

Definition B.1 (Directed Graphs). *A directed Graph G_d is a two-tuple, (V, E_d) where V is the finite set of vertices/nodes and $E_d \subseteq V \times V$ is the set of edges.*

NOTE B.1. *G_d is a directed graph so $(v, w) \in E_d$ does not necessarily imply $(w, v) \in E_d$. The edges are directed from the child node to the parent node.*

Definition B.2 (Parents). *Given a directed Graph $G_d = (V, E_d)$ and $v \in V$, par(v) $= \{w \in V : (v, w) \in E_d\}$ is the set of parent nodes of v and ch(v) $= \{w \in V : (w, v) \in E_d\}$ is the set of child nodes of v. Further $|par(v)|$ and $|ch(v)|$ denote the number of parent nodes and child nodes of v*

Definition B.3 (Paths). *Given a directed Graph $G_d = (V, E_d)$ and $v, w \in V$, a path from v to w is a sequence of vertices $p = (v_0, v_1, ..., v_n)$ such that $(v_i, v_{i+1}) \in E_d$ $\forall\ i \in \{0, 1, .., n-1\}$ where $v_0 = v$ and $v_n = w$. We will say v is connected to w if there exists a path from v to w and is denoted by $v \to w$. Further, we define for any $0 < l < m < n$, $p|_{v_l}^{v_m} = (v_l, v_{l+1}, ..., v_{m-1}, v_m)$, the restriction of path p between v_l and v_m.*

Definition B.4 (Loops). *Given a directed graph G_d and $v, w \in V$, we say there is a loop between v and w if there exists two paths, $\lambda_1 = (\lambda_{10}, \lambda_{11}, .., \lambda_{1n_1})$ and $\lambda_2 =$*

$(\lambda_{20}, \lambda_{21}, .., \lambda_{2n_2})$ *from* v *to* w *such that* $\lambda_{1i} \neq \lambda_{2j}$ $\forall$ $i \in \{1, 2, .., n-1\}$ *and* $j \in \{1, 2, ..., n_2 - 1\}$, *where* $\lambda_{10} = \lambda_{20} = v$ *and* $\lambda_{1n_1} = \lambda_{2n} = w$. *We will denote a loop by* $\lambda = (\lambda_1, \lambda_2)$.

Definition B.5 (Ancestral Recombination Graph)**.** *An Ancestral Recombination Graph (ARG) is a 4-tuple* (G_d, S, T, t), *where* $G_d = (V, E_d)$ *is a directed graph,* $S \subsetneq V$ *is the set of samples,* T *is the set of times, and* $t : V \to T$ *is a bijective function associating each node with its time such that*

1. $(v, w) \in E_d \Rightarrow t(v) < t(w)$
2. $ch(s) = \varnothing$ $\forall$ $s \in S$ *and* $|ch(v)| > 0$ $\forall$ $v \notin S$
3. $\exists$ $w \in V$ *such that* $v \to w$ $\forall$ $v \in V$ *and* $t(v) < t(w)$. *Such a* w *is called a Grand Common Ancestor (GCA). Further the GCA with the least time is called the Grand Most Recent Common Ancestor (GMRCA).*

Further, we say $v \in V$ *is a recombination node if* $|par(v)| = 2$ *and* v *is said to be a coalescence node if* $|ch(v)| > 1$. $4(a)$ *is necessary since an individual cannot have more than 2 parents.* $4(b)$ *ensures that there are no multiple merger nodes but is not necessary.* $4(c)$ *ensures that a node is not both a recombination node and a coalescence node.*

B.2 Spatial Ancestral Recombination Graphs (SpARGs)

Definition B.6 (SpARG)**.** *A d-dimensional Spatial Ancestral Recombination Graph SpARG is an ARG with a set of locations* $L \subset \mathbb{R}^d$ *and a bijective function* $l : V \to L$ *which maps each vertex to its spatial location.*

$\to$

We are interested in estimating the dispersal rate given a particular SpARG under a

model of Brownian motion. Therefore, we assume that displacement along any given edge of an ARG is determined by an independent Brownian motion. However, in order to get the ARG, these independent motions need to satisfy certain conditions. Essentially, we need to condition on these independent Brownian motions forming the loops present in the ARG. For this we build a few more notations and definitions.

For any edge $(v, w) \in E_d$, let $B_{vw} \sim \mathcal{N}(0, \sigma^2 t_{vw})$ be the distribution of the displacement along that edge under Brownian motion where $t_{vw} = t_w - t_v$. We then define the displacement function that takes a path as an input and returns the displacement

$$\begin{aligned} D : P &\rightarrow \mathbb{R} \\ p = (v_i)_{i=0}^{k} &\mapsto \sum_{i=0}^{k-1} B_{v_i v_{i+1}} \end{aligned}$$

In order for these independent Brownian motions to be consistent with the ARG, they need to form the loops present. Let the conditions required for this be

$$\eta_{loops} = \{\sum_{i=0}^{n-1} B_{v_i v_{i+1}} = \sum_{i=0}^{m-1} B_{w_i w_{i+1}} : \big((v_i)_{i=0}^{n}, (w_i)_{i=0}^{m}\big) \text{ is a loop between } v \text{ and } w \text{ in } ARG_{sp} \}$$

This ensures that the Brownian motions meet in a way that they form the loops in the ARG. Now, let X_i denote the displacement of the ith sample in S, s_i, relative to the GMRCA. Then the probability distribution of $\mathbf{X} = \{X_i\}_{i=1}^{n}$, where $n = |S|$ is the number of samples, is given by

$$p_{\mathbf{X}}(x_1, x_2, .., x_n) = p_{\mathbf{B}}(x_1, x_2, .., x_n | \eta_{loops}) \tag{51}$$

where $B_i = D(p_i)$ is the displacement along a path $p_i = (s_{ij})_{j=0}^{n_i}$ from s_i to the GMRCA. Note that for Eq 51 to be well-defined, that is for it have a unique value for a given set of input positions, the value should not depend on the choice of the

path from a sample to the GMRCA. To define this more formally, let $P_i = \{(s_{ij})_{j=0}^{n_i} : s_{i0} = s_i, s_{in_i} = v_{GM}, (s_{ij}, v_{i,j+1}) \in E_d \ \forall \ 0 \leq j < n_i\}$ be the set of paths from s_i to the GMRCA. Now,

$$\begin{aligned} \eta_{i,paths} &= \{D(p_{i1}) = D(p_{i2}) : p_{i1}, p_{i2} \in P_i\} \\ \eta_{paths} &= \bigcup_{i=1}^{n} \eta_{i,paths} \end{aligned}$$

Now, as long as η_{paths} is true Eq 51 is well-defined. We will show that $\eta_{loops} = \eta_{paths}$ which would do the trick.

Lemma B.1. $\eta_{paths} = \eta_{loops}$

Proof : [$\Rightarrow$] We will first show that $\eta_{paths} \subseteq \eta_{loops}$. Therefore, we need to show that given any two paths $p_{i1} = (s_{ij}^{(1)})_{j=0}^{n_i^{(1)}}$ and $p_{i2} = (s_{ij}^{(2)})_{j=0}^{n_i^{(2)}}$ between a sample, s_i, and the GMRCA, there exists loops $\lambda^{(1)}, \lambda^{(2)}, ..., \lambda^{(m)}$ such that

$$\sum_{j=0}^{n_1^{(l)}-1} B_{\lambda_{1j}^{(l)}\lambda_{1,j+1}^{(l)}} = \sum_{j=0}^{n_2^{(l)}-1} B_{\lambda_{2j}^{(l)}\lambda_{2,j+1}^{(l)}} \ \forall \ 1 \leq l \leq m \Leftrightarrow \sum_{j=0}^{n_i^{(1)}-1} B_{s_{ij}^{(1)}s_{i,j+1}^{(k_1)}} = \sum_{j=0}^{n_i^{(2)}-1} B_{s_{ij}^{(2)}s_{i,j+1}^{(k_2)}}$$

Here is how to find the loops starting from the two distinct paths. Let $n_{min} = \min\{n_i^{(1)}, n_i^{(2)}\}$ and $n_{max} = \max\{n_i^{(1)}, n_i^{(2)}\}$. Then define $J := \{j \in \{0, 1, ..., n_{min}\} : s_{il}^{(k_1)} = s_{il}^{(k_2)} \ \forall \ l \leq j\}$ and $j_{st} := \max J$. Now, $j_{st} < n_{max}$, otherwise we will have that $s_{il}^{(k_1)} = s_{il}^{(k_2)}$ for all l, which would mean the two paths are identical leading to a contradiction since we started with two distinct paths. Therefore,

$$j_{st} < n_{max}.$$

Let $v_{st}^{(1)} = s_{ij_{st}}^{(1)} = s_{ij_{st}}^{(2)}$. This is the start of the first loop. To find the end of this loop, let $V_{end} := \{u \in p_{i1} \cap p_{i2} : t_u > t_{v_{st}^{(1)}}\}$. Then, $v_{end}^{(1)} = \min_{t(u)} V_{end}$. Note that $\lambda_1 = (p_{i1}|_{v_{st}^{(1)}}^{v_{end}^{(1)}}, p_{i2}|_{v_{st}^{(1)}}^{v_{end}^{(1)}})$ is a loop.

Now, if $p_{i1}|_{v_{end}^{(1)}}^{v_{GM}} = p_{i2}|_{v_{end}^{(1)}}^{v_{GM}}$, then we are done. Since everything before $v_{st}^{(1)}$ and after $v_{end}^{(1)}$ are identical in the two paths, the displacement along the two paths being equal is the same as the displacements forming the loop λ_1.

If $p_{i1}|_{v_{end}^{(1)}}^{v_{GM}} \neq p_{i2}|_{v_{end}^{(1)}}^{v_{GM}}$, then we can repeat the above steps on $p_{i1}|_{v_{end}^{(1)}}^{v_{GM}}$ and $p_{i2}|_{v_{end}^{(1)}}^{v_{GM}}$, to find $v_{st}^{(2)}$ and $v_{end}^{(2)}$ such that $\lambda_2 = (p_{i1}|_{v_{st}^{(2)}}^{v_{end}^{(2)}}, p_{i2}|_{v_{st}^{(2)}}^{v_{end}^{(2)}})$ is a loop and $p_{i1}|_{v_{end}^{(1)}}^{v_{st}^{(2)}} = p_{i2}|_{v_{end}^{(1)}}^{v_{st}^{(2)}}$. Keep repeating this until we have $(v_{st}^{(k)}, v_{end}^{(k)})_{k=1}^{m}$ such that $\lambda_l = (p_{i1}|_{v_{st}^{(l)}}^{v_{end}^{(l)}}, p_{i2}|_{v_{st}^{(l)}}^{v_{end}^{(l)}})$ are loops and $p_{i1}|_{v_{end}^{(l)}}^{v_{st}^{(l+1)}} = p_{i2}|_{v_{end}^{(l)}}^{v_{st}^{(l+1)}}$ $\forall\ 1 \leq l \leq m$ where $v_{st}^{(m+1)} = v_{GM}$. We can do this because it is a finite graph.

Therefore, the displacement along the two parts being equal is equivalent to the displacements forming the loops $\lambda_1, \lambda_2, .., \lambda_m$.

Thus, $\eta_{paths} \subseteq \eta_{loops}$.

[$\Rightarrow$] Now we will show that $\eta_{loops} \subseteq \eta_{paths}$. That is we need to show that given a loop λ, there exists a pair of paths p_{i1} and p_{i2} from a sample to the GMRCA such that

$$\sum_{j=0}^{n_i^{(k_1)}-1} B_{s_{ij}^{(1)} s_{i,j+1}^{(1)}} = \sum_{j=0}^{n_i^{(2)}-1} B_{s_{ij}^{(2)} s_{i,j+1}^{(2)}} \Rightarrow \sum_{j=0}^{n_1-1} B_{\lambda_{1j}\lambda_{1,j+1}} = \sum_{j=0}^{n_2-1} B_{\lambda_{2j}\lambda_{2,j+1}}$$

Here is how you find two distinct paths given a loop λ. Suppose λ is a loop from v to w. By definition of an ARG, $w \to v_{GM}$. Let this path be $p_{wv_{GM}}$

Now, if $v \in S$, then $p_{i1} = \lambda_1 \cup p_{wv_{GM}}$ and $p_{i2} = \lambda_2 \cup p_{wv_{GM}}$ are two distinct paths from the sample v to the GMRCA v_{GM}. Therefore, the condition for the Brownian motions to form the loop λ_1 is the same as the displacement along p_{i1} being equal to p_{i2}.

Now, if $v \notin S$, then we claim that there exists a sample $s_i \in S$ such that $s_i \to v$. Suppose not, i.e., $s_i \nrightarrow v \; \forall \; 1 \leq i \leq n$. Since $v \notin S$, therefore $\exists v^{(1)}$ such that $(v^{(1)}, v) \in E_d$ by definition of an ARG. Now $v^{(1)}$ also does not belong to S, otherwise we will have a vertex in S that is connected to v. Similarly, by induction we can construct $\{v^{(k)}\}_{k \in \mathbb{N}}$ such that $(v^{(k)}, v^{(k+1)}) \in E_d$ and $v^{(k)} \notin S \; \forall \; k \in \mathbb{N}$. Therefore we have infinite vertices in the *ARG* which is a contradiction. Therefore our claim has to be true. Let the path between s_i and v be p_{iv}, then $p_{iv} \cup \lambda_1 \cup p_{wv_{GM}}$ and $p_{iv} \cup \lambda_2 \cup p_{wv_{GM}}$ are the two required paths. Therfore, $\eta_{loops} \subseteq \eta_{paths}$

Therefore, we have that $\eta_{loops} = \eta_{paths}$. Q.E.D

C Minimal Path Matrix

Note that η_{loops} will have exactly as many conditions as the number of recombination nodes, say k, in the ARG. However, the size of the full paths matrix is n_p, the total number of paths, which is strictly greater than $2 * k + n_s$ and is bounded above by $2^k n_s$. Consequently the number of conditions in η_{paths} is bounded by $\binom{n_p}{2}$, which increases at least quadratically in k and potentially exponentially. Therefore, η_{paths} has multiple redundant conditions. But we only need one pair of paths for each condition η_{loops}, which only differ in one loop. We further need atleast one path each sample. Therefore, if chosen correctly, we only need $n_s + k$ paths to get the correct estimates. We call the matrix of shared times of these $n_s + k$ paths as the minimal path matrix. Below we provide the algorithm to build this matrix.

C.1 Algorithm (step by step)

1. Initialization

- The shared time matrix $\mathbf{S} \leftarrow [0]_{n \times n}$, a zero square matrix of size n, the number of samples. The entry of the ith row and j^{th} column is denoted by s_{ij}
- The list of paths PL $\leftarrow [[1], [2], ..., [n]]$ with one path for each sample node.

2. Loop through every node in the ARG in time ascending order and repeat this step for each node. In each iteration, let u be the focal node. Let I_u be the set of indices of the paths in PL that end in u. Let $V_{parents}$ be the set of parents of u and $k_u = |V_{parents}|$ be the the number of parent nodes.

   (a) If $k_u = 0$, skip this step.

   (b) If $k_u = 1$, with parent node v, then do the following

   - $s_{ij} \leftarrow s_{ij} + t_{uv}$ for all i, j in I_u, where t_{uv} is length of edge uv (time between nodes u and v).
   - $PL[i] \leftarrow PL[i] + [v]$ for all $i \in I_u$, i.e., append x to all the paths that currently end at u.

   This step is extending the paths that up to this point ended at u; they now end at v. Since they all will share the edge uv, we add its lengths to the corresponding covariance terms.

   (c) If $k_u = 2$ (i.e. parent is a recombination node), with parent nodes v_1 and v_2, do the following

   - Pick one index from I_u, say l.
   - $PL \leftarrow PL + [PL[l]]$. Duplicate the lth path. Don't update I_u
   - $PL[i] \leftarrow PL[i] + [v_1]$ for all i in I_u. Extend all existing paths that end at u to v_1.
   - $PL[-1] \leftarrow PL[-1] + [v_2]$. Extend the new path formed in this step to v_2.

- $\mathbf{S} \leftarrow \begin{bmatrix} \mathbf{S} & \mathbf{S}[:][l] \end{bmatrix}$, i.e., Duplicate the lth column of $\mathbf{S}$
- $\mathbf{S} \leftarrow \begin{bmatrix} \mathbf{S} \\ \mathbf{S}[:][l] \end{bmatrix}$, i.e., Duplicate the lth row of $\mathbf{S}$
- $s_{ij} \leftarrow s_{ij} + t_{uv_1}$ for all i, j in I_u.
- $s_{ll} \leftarrow s_{ll} + t_{uv_2}$ for all i, j in I_u.

3. The end result is $\mathbf{S}$, the minimal paths matrix.

D Relaxed Meeting

In order to relax the meeting condition at recombination nodes, we allow the parents of a recombination node to be different from each other (NOTE: By parents we mean the actual parents of the recombination node, not the parent nodes of the recombination node in the network). The location of the recombination node is the the average of its parent locations. The covariance matrix of the sample locations under this model has been computed in Bastide *et al.* (2018) (with $\gamma_e = 1/2$ in their model), which we refer to for the details. Here, we show how that matrix can be computed from the Full Paths Shared Time Matrix, $\mathbf{S}_p$ that we have.

Let P_i be the set of paths from the sample i to any of the roots. Then the covariance between two samples i and j is given by

$$\sigma^2 \sum_{p_i \in P_i} \sum_{p_j \in P_j} \frac{1}{2^{|r_i + r_j|}} \sum_{e \in p_i \cap p_j} t_e \tag{52}$$

where $p_i \cap p_j$ is the set of common edges between the two paths and r_i is the number of recombination nodes along the path p_i. Let $\vec{W}$ be a $n_p \times 1$ vector which encodes the weights associated with each path, i.e., the kth entry of $\vec{W}$ is $\frac{1}{2^{r_k}}$ where r_k is the number of recombination nodes in the kth path. Then the covariance matrix of sample

locations is given by

$$\sigma^2 \mathbf{S}_\infty = \sigma^2 \mathbf{P}_p^T (\mathbf{S}_p \circ (\vec{W}\vec{W}^T))\mathbf{P}_p \tag{53}$$

where $\circ$ is the elementwise multiplication (Hadamard product) of the two matrices. Further, with multiple roots, the mean of the sample locations is given by the vector. The mean of the samples locations in case of the multiple roots is given by

$$\mathbf{R}_\infty \vec{\mu} = \mathbf{P}_p^T (R \circ (\mathbb{1}_{n_r}^T \otimes \vec{W}))\vec{\mu} \tag{54}$$

Then the MLE estimates for root locations and dispersal rate is given by

$$\widehat{\vec{\mu}} = (\mathbf{R}_\infty^T \mathbf{S}_\infty^{-1} \mathbf{R}_\infty)^{-1} \mathbf{R}_\infty^T \mathbf{S}_\infty^{-1} \vec{\ell^*} \tag{55}$$

$$\widehat{\sigma^2} = \frac{(\vec{\ell^*} - \mathbf{R}_\infty \widehat{\vec{\mu}})^T \mathbf{S}_\infty^{-1} (\vec{\ell^*} - \mathbf{R}_\infty \widehat{\vec{\mu}})}{n_s} \tag{56}$$

E Supplemental Figures

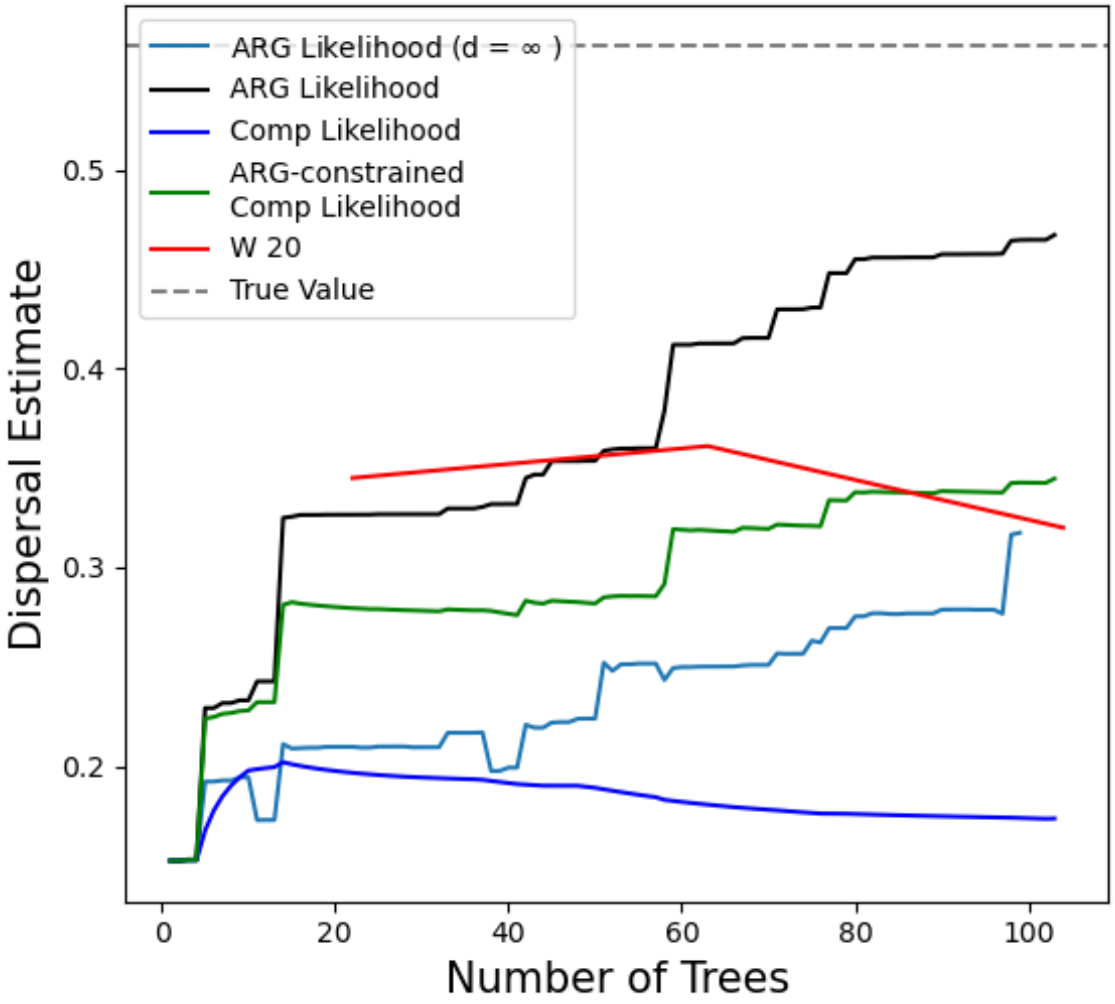

Figure S1: Dispersal rate computed under different methods as a function of the number of trees. "ARG Likelihood (d = ∞)" is the dispersal rate computed from the full ARG using the relaxed meeting condition. All other methods are as in Figure 3.

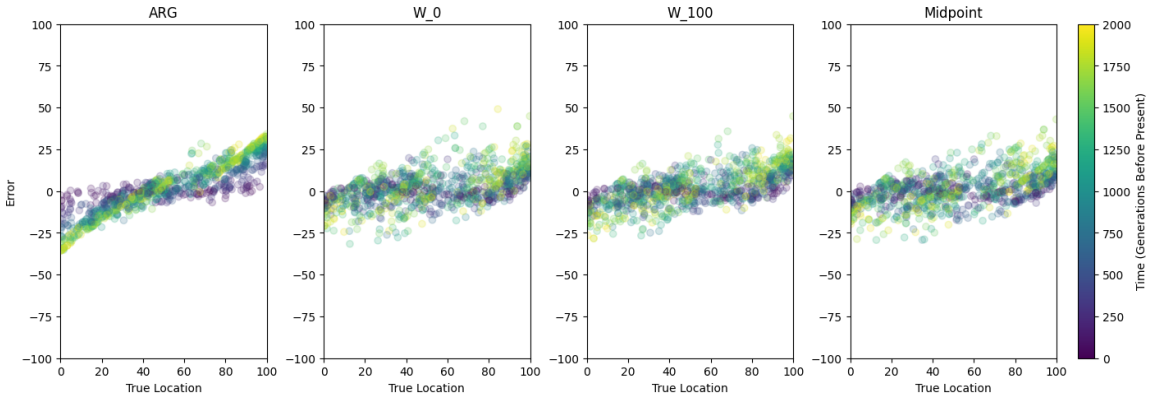

Figure S2: Error in ancestor location estimates as a function of the true location and colored by time.