Distance metrics for ranked evolutionary trees

Jaehee Kim, Noah A. Rosenberg, and Julia A. Palacios

Genealogical tree modeling is essential for estimating evolutionary parameters in population genetics and phylogenetics. Recent mathematical results concerning ranked genealogies without leaf labels unlock opportunities in the analysis of evolutionary trees. In particular, comparisons between ranked genealogies facilitate the study of evolutionary processes of different organisms sampled at multiple time periods. We propose metrics on ranked tree shapes and ranked genealogies for lineages isochronously and heterochronously sampled. Our proposed tree metrics make it possible to conduct statistical analyses of ranked tree shapes and timed ranked tree shapes or ranked genealogies. Such analyses allow us to assess differences in tree distributions, quantify estimation uncertainty, and summarize tree distributions. We show the utility of our metrics via simulations and an application in infectious diseases.

coalescent | distance metric | phylogenetics | ranked genealogy | ranked tree shape

Genetic genealogies are rooted binary trees that describe the ancestral relationships of a sample of molecular sequences at a locus. The properties of these genealogies are influenced by the nature of the evolutionary forces experienced by the sample’s ancestry. Hence, assessing differences among gene genealogies can provide information about differences in these forces. In this article, we propose a distance on genealogies that enables biologically meaningful comparisons between genealogies of nonoverlapping samples. Our proposed distance, in its more general form, is a distance on ranked genealogies.

Ranked genealogies are rooted binary unlabeled trees with branch lengths and ordered internal nodes (Fig. 1A). These genealogies are also known as Tajima’s genealogies (1, 2), as they were examined by Tajima (3). Unlabeled ranked genealogies are coarser than labeled ranked genealogies but finer than unlabeled unranked genealogies (Fig. 1); in a labeled or unlabeled unranked genealogy, multiple bifurcations may occur simultaneously (4).

Recently, there has been increasing interest in modeling ranked genealogies for studying evolutionary dynamics (1, 5, 6). One method for inferring evolutionary parameters from molecular data is based on the Tajima coalescent for ranked genealogies (2): Modeling of ranked genealogies, as opposed to labeled ranked genealogies as in the Kingman coalescent, reduces the dimensionality of the inference problem. Thus, for example, in studying macroevolution, Maliet et al. (7) proposed a model on ranked tree shapes—ranked genealogies without branch lengths but with internal node orders retained—to investigate factors influencing nonrandom extinction and the loss of phylogenetic diversity.

Many metrics on labeled trees have been proposed, such as the Robinson–Foulds (RF) metric (8), the Billera–Holmes–Vogtmann (BHV) metric (9), and the Kendall–Colijn (KC) metric (10). These metrics have been used for summarizing posterior distributions and bootstrap distributions of trees on the same set of taxa (11, 12), for comparing estimated genealogies of the same organisms obtained with different procedures, and for quantifying uncertainty in inferring evolutionary histories (13). The Colijn–Plazzotta (CP) metric (14) on the coarser-resolution space of tree shapes—rooted binary unlabeled unranked trees without branch lengths—has recently been devised. To our knowledge, no other metrics on ranked tree shapes or ranked unlabeled genealogies have been proposed to date.

A tree metric on the space of ranked genealogies can facilitate evaluations of the quality of an estimation procedure, by enabling measurements of the distance between an estimated ranked genealogy and the true ranked genealogy. It can assist in comparing different estimators from different procedures and in comparing estimated ranked genealogies of different organisms living at different geographical locations and different time periods. Moreover, a useful tree metric not only provides a quantitative measure for ranked genealogy comparison, but can also discriminate between key aspects of different evolutionary processes (15). We show that our metrics separate samples of genealogies originating from different sampling distributions of ranked tree shapes and ranked genealogies. Our distances enable the computation of summary statistics, such as the mean and the variance, from samples of ranked genealogies. When ranked genealogy samples are obtained from posterior distributions of genealogies, such as those obtained from Bayesian Evolutionary Analysis by Sampling Trees (BEAST) (16), our tree metrics enable the construction of credible sets and Markov chain Monte Carlo (MCMC) convergence assessment.

We first define a metric on ranked tree shapes. Our metric relies on an integer-valued triangular matrix representation of ranked tree shapes. This matrix representation allows us to use metrics on matrices, such as the L1 norm and L2 (Frobenius) norm, to define metrics on ranked tree shapes. The choice of these two distances produces computational benefits, as computations of the metrics are quadratic in the number of leaves. Our metrics on ranked tree shapes retain more information than metrics based on unlabeled unranked tree shapes alone. We expand our metric definition to ranked genealogies, including branch labels. Our metrics enable quantification of evolutionary differences, assessment of tree uncertainty, and construction of statistical summaries of a tree distribution. They are computationally efficient and particularly useful for comparing phylodynamics of infectious diseases involving heterochronous samples and for comparative analyses of organisms that live in different geographic regions.

Significance

Rooted binary trees inferred from molecular sequence data provide information about the evolutionary history of populations and species. We introduce metrics on ranked tree shapes and ranked genealogies, in which the shape and temporal branching order in a tree are considered, but not the taxon labels. Our metrics enable quantification of evolutionary differences, assessment of tree uncertainty, and construction of statistical summaries of a tree distribution. They are computationally efficient and particularly useful for comparing phylodynamics of infectious diseases involving heterochronous samples and for comparative analyses of organisms that live in different geographic regions.

Author contributions: J.K. and J.A.P. designed research; J.K., N.A.R., and J.A.P. performed research; J.K. and J.A.P. analyzed data; and J.K., N.A.R., and J.A.P. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

1To whom correspondence may be addressed. Email: juliap@stanford.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1922851117/-/DCSupplemental.

First published November 2, 2020.

28876-28886 | PNAS | November 17, 2020 | vol. 117 | no. 46
www.pnas.org/cgi/doi/10.1073/pnas.1922851117
We are introducing metrics for unlabeled ranked tree shapes and unlabeled ranked genealogies. To define our metrics on ranked tree shapes, we need to introduce a unique encoding of a ranked tree shape as an integer-valued triangular matrix.

**Tree Metrics for Comparative Analysis**

**Metrics on Labeled Trees.** Many tree metrics have been proposed for phylogenetic studies (8–10, 27–30). We consider several of these. Billera et al. (9) introduced a metric on labeled unranked genealogies that is now commonly known as BHV space. Owen and Provan (31) and Chakerian and Holmes (11) provided polynomial algorithms and implementations for calculating the geodesic distance metric proposed by Billera et al. (9). Recently, Kendall and Colijn (10) proposed a metric on labeled unranked tree shapes and labeled unranked genealogies, representing each tree as a convex combination of two vectors—one encoding the number of edges from the root to the most recent common ancestor of every pair of tips in the tree and the other encoding the sum of the branch lengths of the corresponding paths. The most popular metric that can be computed in polynomial time is the symmetric difference of Robinson and Foulds (RF) (8) on labeled unranked tree shapes. This measure has an associated branch-length measure RFL on labeled genealogies (27).

**Metrics on Unlabeled Trees.** A metric on unlabeled trees is desirable because it enables comparison between trees of different samples. However, few such metrics are available. Colijn and Plazotta (14) proposed a metric on tree shapes which is the Euclidean norm of the difference between two integer vectors that uniquely describe the two trees. Poon et al. (32) developed a kernel function that measures the similarity between two genealogies by accounting for differences in branch lengths and matching the number of descendants over all nodes for both trees. Lewitus and Morlon (33) proposed the Jensen–Shannon distance between the spectral density profiles of the corresponding modified graph Laplacians of the genealogies.

We are introducing metrics for unlabeled ranked tree shapes and unlabeled ranked genealogies. To define our metrics on ranked tree shapes, we need to introduce a unique encoding of a ranked tree shape as an integer-valued triangular matrix.

**New Approaches**

**Unique Encoding of Ranked Tree Shapes and F Matrix.** Let $T^R$ be a ranked tree shape with $n$ leaves sampled at time $0 = u_n$. Let $(u_{n-1}, \ldots, u_1)$ be the $n - 1$ coalescence times. Here, time increases into the past (rootward), $u_{n-1} < \ldots < u_1$, and $u_0$ and $u_1$ correspond to the most recent sampling time and the time to the most recent common ancestor (root), respectively. An $F$ matrix that encodes $T^R$ is an $(n - 1) \times (n - 1)$ lower triangular matrix of integers with elements $F_{i,j} = 0$ for all $i < j$, and for $1 \leq j \leq i$, $F_{i,j}$ is the number of extant lineages in $(u_{j+1}, u_i)$ that do not bifurcate during the entire time interval $(u_{j+1}, u_i)$ traversed forward in time (tipward).

In Fig. 2B, we show the corresponding $F$ matrix to the ranked tree shape depicted in Fig. 2A. In the interval $(u_2, u_3)$, there are two lineages, so $F_{3,1} = 2$. Traversing tipward, one of the two lineages extant at time $(u_2, u_3)$ branches at time $u_3$, while the other lineage does not branch throughout the entire interval $(u_2, u_3)$. This gives the first column of the $F$ matrix: $F_{3,1} = F_{3,2} = 1$. For the second column, we start with three lineages in the interval $(u_3, u_2)$, $F_{2,2} = 3$. Traversing tipward, of the three lineages extant at $(u_3, u_2)$, one branch at $u_3$ ($F_{3,2} = 2$), and one branch at $u_4$ ($F_{4,2} = 1$). We construct the third column by...
AB

Unique encoding of isochronous ranked tree shapes and corresponding encodings. The ranked tree shapes, as both types of trees have unique matrix L(1) starting from four lineages in (u4, u1), there are five lineages, which gives F4,4 = 5.

If T R is a heterochronous ranked tree shape with n leaves and m sampling events, then the corresponding F-matrix representation is an (n + m - 2) × (n + m - 2) lower triangular matrix of integers, where Fij for 1 ≤ j ≤ i is the number of extant lineages in (ui+1, ui) that do not bifurcate or become extinct during the entire time interval (ui+1, ui), traversing forward in time. Here, (u0+m+1, u0+m+2, ..., u1), such that 0 = u0+m+1 < u0+m+2 < ... < u1, are the n + m - 1 ordered change points of T R, at each of which the number of branches changes either by a coalescent event or by a sampling event. We show an example of a heterochronous ranked tree shape and its F-matrix encoding in SI Appendix, Fig. S5 C and D.

Although the branch lengths and the actual values of coalescent and sampling times are irrelevant for the specification of the ranked tree shape, we rely on the ui to identify the order and type of change points: coalescence or sampling in T R and F.

Theorem 1 (Unique Encoding of Ranked Tree Shapes). The map by which ranked tree shapes with n samples and m sampling events are encoded as F-matrices of size (n + m - 2) × (n + m - 2) uniquely associates a ranked tree shape with an F matrix.

In other words, given an F matrix, if it encodes a ranked tree shape, then it encodes exactly one ranked tree shape. The proof appears in SI Appendix, section 1. In the next section, we leverage the F-matrix representation of ranked tree shapes to define a distance between ranked tree shapes. From this point onward, we assume a matrix F represents a ranked tree shape.

Metrics on Ranked Tree Shapes and Ranked Genealogies. We define two distance functions d1 and d2 on the space of ranked tree shapes with n leaves as follows. For a pair of ranked tree shapes T R 1 and T R 2 and their corresponding F-matrix representations F(1) and F(2) of size r × r, where r ≥ n - 1,

\[ d_1(T_{R1}, T_{R2}) = n \sum_{i=1}^{r} \sum_{j=1}^{i} |F_{ij}^{(1)} - F_{ij}^{(2)}|, \]

\[ d_2(T_{R1}, T_{R2}) = \sqrt{ \sum_{i=1}^{r} \sum_{j=1}^{i} \left( F_{ij}^{(1)} - F_{ij}^{(2)} \right)^2 }. \]

The two distances are metrics as they inherit L1 and L2 element-wise matrix metrics with well-understood properties (34). The definitions are valid for both isochronous and heterochronous ranked tree shapes, as both types of trees have unique matrix encodings.

In the following definition, we include branch lengths to define distances between ranked genealogies. We define two weighted distance functions d1w and d2w on the space of ranked genealogies with n samples and m sampling events. We first define the weight matrix W of a ranked genealogy g R as a lower triangular matrix of size (n + m - 2) × (n + m - 2) with entries Wij = uj - ui+1 for j ≤ i and Wij = 0 otherwise. Here, (u0+m+1, u0+m+2, ..., u1), such that 0 = u0+m+1 < u0+m+2 < ... < u1, is the vector of n - 1 coalescent times ti and m sampling times sj, ordered with time increasing into the past (rootward). For a pair of ranked genealogies g R 1 and g R 2, and their corresponding F-matrix representations F(1) and F(2),

\[ d_1^w(g_{R1}, g_{R2}) = \sum_{i=1}^{r} \sum_{j=1}^{i} \left( F_{ij}^{(1)} W_{ij}^{(1)} - F_{ij}^{(2)} W_{ij}^{(2)} \right), \]

\[ d_2^w(g_{R1}, g_{R2}) = \sum_{i=1}^{r} \sum_{j=1}^{i} \left( F_{ij}^{(1)} W_{ij}^{(1)} - F_{ij}^{(2)} W_{ij}^{(2)} \right)^2. \]

where W(1) and W(2) are the weight matrices associated with g R 1 and g R 2, respectively. SI Appendix, Fig. S6 shows an example weight matrix W, associated with the example heterochronous ranked genealogy and its F matrix in SI Appendix, Fig. S5 C and D.

Proposition 2. The weighted distances d1w and d2w are metrics.

The proof appears in SI Appendix, section 2. Our distances on ranked tree shapes and ranked genealogies are distances between trees with the same number of leaves and, additionally in the heterochronous case, the same number of sampling events. An extension to cases in which the numbers of sampling events differ but the total numbers of leaves remain the same is described in SI Appendix, section 3.

In the following section, we propose sample summary statistics based on our metrics d1 and d2 for ranked tree shapes and d1w and d2w for ranked genealogies.

 Ranked Tree Shape and Ranked Genealogy Summary Statistics. We first use our proposed distances to define a notion of mean value and dispersion value from a finite sample \{T_{R1}, T_{R2}, ..., T_{Rn}\} of ranked tree shapes with n leaves. Our proposed measures of centrality are the L2-medorid sets defined as

\[ T_{i} := \arg \min_{T \in \{T_{R1}, T_{R2}, ..., T_{Rn}\}} \sum_{j=1}^{n} d_i^2(T, T_{Rj}), \]

for i = 1, 2. We note that our definition of the L2-medorid set corresponds to the ranked tree shape(s) in the sample minimizing the sum of squared distances as opposed to the sum of distances. In addition, when the sample is replaced by the complete population of ranked tree shapes with n leaves or when we allow the L2 medoid to belong to the population of trees but do not require it to be a sampled tree, Eq. 5 corresponds to the Fréchet mean or barycenter under uniform sampling probabilities (35).

We use the following as a measure of dispersion around the medoid for i = 1, 2:

\[ \sigma_i^2 := \frac{1}{n} \sum_{j=1}^{n} d_i^2(T_{ij}, T_{Rj}). \]

When the L2 medoid is not unique, the dispersion is defined with respect to any chosen L2-medorid ranked tree shape from the set of L2 medoids.
Similarly, given a finite sample of ranked genealogies \( \mathbf{g}_1, \ldots, \mathbf{g}_n \) with \( n \) leaves, the \( L_2 \)-medoid set is defined as

\[
g_i := \arg \min_{\mathbf{g} \in \{\mathbf{g}_1, \ldots, \mathbf{g}_n\}} \sum_{j=1}^{n} \left[ d_{i}^{w}(\mathbf{g}, \mathbf{g}_j) \right]^2
\]

for \( i = 1, 2, \) and our empirical measure of dispersion for ranked genealogies is the mean sum of squared distances to the medoid.

### Statistical Comparison of Ranked Tree Shape and Ranked Genealogy Sampling Distributions

To assess the utility of a metric in distinguishing between two different ranked tree shape (or ranked genealogy) sampling distributions, we propose a nonparametric test of equality of tree sampling distributions as follows. Given two sets of independent samples of ranked tree shapes (or ranked genealogies) from distributions \( P \) and \( Q \), we are interested in testing the null hypothesis \( H_0 : P = Q \).

Assume \( X_1, \ldots, X_N \) are independent and identically distributed (i.i.d.) ranked tree shapes (or ranked genealogies) with \( n \) leaves according to a probability distribution \( P \), and independently, \( Y_1, \ldots, Y_N \) are i.i.d. ranked tree shapes (or ranked genealogies) with \( n \) leaves according to \( Q \). For a given distance function \( d \) defined on the space of ranked tree shapes (or ranked genealogies), let \( \overline{X} \) be the \( L_2 \) medoid of the \( X_1, \ldots, X_N \) samples and let \( \overline{Y} \) be the \( L_2 \) medoid of the \( Y_1, \ldots, Y_N \) samples, as defined in Eq. 5. The mean confusion is defined as

\[
C^{x,y} = \frac{1}{2(N - 1)} \sum_{j=1}^{N} \left[ d(X_j, \overline{X}) \leq d(X_j, \overline{Y}) + d(Y_j, \overline{X}) \leq d(Y_j, \overline{Y}) \right].
\]

The closer the value of the mean confusion statistic is to 0, the more easily the two tree distributions can be distinguished from each other. The mean confusion as defined in Eq. 8 assumes the two \( L_2 \) medoids are unique, each observed only once, and ignores ties (trees equidistant to \( \overline{X} \) and \( \overline{Y} \)). When the sample space of ranked tree shapes is large or when the sample space is the space of ranked genealogies, the expected value of the mean confusion statistic under the null hypothesis is close to 0.5. In SI Appendix, section 4, we define a more general mean confusion statistic that accounts for nonunique \( L_2 \) medoids and ties. We note that for our simulations and real data analyses, Eq. 8 is used since we observe a unique \( L_2 \) medoid in each ranked tree shape or ranked genealogy distribution, and no ties are present in pairwise comparisons.

An \( \alpha \)-level test can be constructed for testing the null hypothesis using \( N_{\text{perm}} \) random permutations. The permutation \( P \) value is computed as (36)

\[
P = \frac{1 + \sum_{k=1}^{N_{\text{perm}}} \mathbf{1}_{C^{x,y} \leq C^{*}}}{1 + N_{\text{perm}}},
\]

where \( C^{x,y}_k \) is a confusion statistic from the \( k \)th replicate and \( C^{x,y} \) is the observed value from data.

For comparing more than two sampling distributions, we compute the confusion matrix \( C \), where entry \( C_{ij} \) corresponds to the percentage of trees in the \( i \)th sampling distribution that are closest to the \( L_2 \) medoid of the \( j \)th sampling distribution. Each row of the confusion matrix sums to 100\%, and the diagonal elements represent the percentage of trees that are closer to the \( L_2 \) medoid of their originating distribution than \( L_2 \) medoids of the other sampling distributions. The average diagonal of the confusion matrix, which we term mean separation, provides a measure of overall separation between sampling distributions based on the distance used. It ranges from 0 to 100, with a larger value indicating better separation between sampling distributions.

### Table 1. Summary of distances in the spaces of ranked tree shapes and ranked genealogies

| Name | Tree space | Original tree space | Reference |
|------|------------|---------------------|-----------|
| \( d_1, d_2 \) | Ranked tree shape | Ranked tree shape | This paper |
| \( d^*_1, d^*_2 \) | Ranked | Ranked | This paper |
| \( d_{\text{BHV-RTS}} \) | Ranked tree genealogy | Labeled genealogy | Billera, Holmes, and Vogtmann (9) |
| \( d_{\text{BHV-RG}} \) | Ranked tree genealogy | Labeled genealogy | Billera, Holmes, and Vogtmann (9) |
| \( d_{\text{KC-RTS}} \) | Ranked tree genealogy | Labeled genealogy | Kendall and Colijn (10) |
| \( d_{\text{KC-RG}} \) | Ranked tree genealogy | Labeled genealogy | Kendall and Colijn (10) |
| \( d_{\text{CP-RTS}} \) | Ranked tree shape | Tree shape | Colijn and Pizzotta (14) |

A brief description of existing distances that are originally defined on the space of tree shapes and the space of labeled genealogies, and their adaptations to the spaces of ranked tree shapes and ranked genealogies, can be found in SI Appendix, sections 5 and 6.

### Adapting Other Tree Metrics to Ranked Tree Shapes and Ranked Genealogies

Although there are no other metrics designed specifically for ranked tree shapes and ranked genealogies, we can adapt other tree distances defined on finer or coarser tree spaces to the space of ranked tree shapes and ranked genealogies, and we can compare them to our metrics. We adapted metrics that are originally defined on the space of labeled genealogies—the BHV distance and the KC distance—to the space of ranked genealogies by artificially assigning uniquely defined leaf labels to ranked genealogies. The unique assignment of the leaf labels on a ranked tree shape consists of assigning labels in increasing index order starting with leaves subtended by an internal node closest to the tips and ending with leaves subtended closest to the root. For adapting the BHV distance and the KC distance to the space of ranked tree shapes, we assigned a unit length to each change point time interval and unique leaf labels to tips of a ranked tree shape. We also compared our distances on ranked tree shapes to the CP distance defined on the space of tree shapes by ignoring internal node labels. Details of these adaptations can be found in SI Appendix, sections 5 and 6.

We term the adapted BHV, KC, and CP distances on the space of ranked tree shapes \( d_{\text{BHV-RTS}}, d_{\text{KC-RTS}}, \) and \( d_{\text{CP-RTS}} \), respectively. We denote the adapted BHV and KC distances on ranked genealogies by \( d_{\text{BHV-RG}} \) and \( d_{\text{KC-RG}} \), respectively. Table 1 shows a summary of the distances defined on the spaces of ranked tree shapes and ranked genealogies, as used in this paper.

### Results

Having introduced the metrics on the spaces of ranked tree shapes and ranked genealogies, we first examine the behavior of our metrics in an illustrative example for \( n = 5 \). We then show their utility in simulated data under various evolutionary models. Finally, we apply them to analyze real human influenza A/H3N2 virus data.

### Interpreting Proposed Distances between Ranked Tree Shapes of \( n = 5 \) Leaves

Before demonstrating the utility of our metrics in distinguishing different sampling distributions by simulations and a real data application, as a simple illustrative example, we first explore our distances on ranked tree shapes of \( n = 5 \).
leaves—the smallest number of taxa giving a nontrivial space of ranked tree shapes—and compare them with the adaptations of other metrics.

In Fig. 3, we show all five possible ranked tree shapes and their corresponding pairwise $d_i$ distances. Trees $T_1^R$, $T_2^R$, and $T_3^R$ are with the same tree shape but different ranked tree shapes. The trees in each of the pairs $(T_1^R, T_2^R)$ and $(T_1^R, T_4^R)$ differ by one rank switch, and their pairwise distance is $d_i = 1$. The $d_i$ distance between the pair $(T_2^R, T_3^R)$ is 2; indeed, to go from $T_2^R$ to $T_3^R$, we need two rank switches. All pairwise distances are shown in SI Appendix, Fig. S7.

The pairs $(T_1^R, T_3^R)$ and $(T_2^R, T_5^R)$ have the maximum $d_i$ distance. In both cases, we need a rank switch and a change of tree shapes to move from one tree to the other. The pairs of trees with the maximum $d_{\text{BHVT-RST}}$ distance are the same two pairs with the maximum $d_i$ distance. However, an additional pair, $(T_3^R, T_4^R)$, also has the same $d_{\text{BHVT-RST}}$ maximum, even though $T_3^R$ and $T_4^R$ differ by two rank change events but have the same tree shape.

Qualitative behaviors of distances $d_{\text{KC-RST}}$ among the trees $T_1^R$, $T_2^R$, $T_3^R$, and $T_4^R$ mirror $d_{\text{CP-RST}}$; however, differs in that the trees $T_2^R$, $T_3^R$, and $T_4^R$ are equidistant from $T_5^R$ but not from the caterpillar tree $T_1^R$, whereas, using $d_1$ or $d_2$, all pairwise distances to the trees $T_2^R$, $T_3^R$, and $T_4^R$ from $T_1^R$ or $T_5^R$ are distinct. We also note that because $d_{\text{CP-RST}}$ is a distance originally defined on tree shapes, all pairwise $d_{\text{CP-RST}}$ distances between the trees $T_2^R$, $T_3^R$, and $T_4^R$ are 0, and all $d_{\text{CP-RST}}$ distances to the trees $T_2^R$, $T_3^R$, and $T_4^R$ from $T_1^R$ or $T_5^R$ have the same value of 2.

Separation between Ranked Tree Shapes with Different Distributions. Having demonstrated that our proposed metrics, $d_1$ and $d_2$, effectively differentiate both tree topologies and internal node rankings in the $n = 5$ case, we now evaluate by simulation how $d_1$ and $d_2$ can distinguish different ranked tree shape distributions from models with biological relevance.

We consider isorhizous ranked tree shapes from a two-parameter branching model to which we refer as the alpha–beta-splitting model (7). These parameters capture two phenomena that have been widely used for phylogenetic studies. The parameter $\beta \in [-2, \infty)$, as in the beta-splitting model (37), determines the degree of balance, an important tree attribute that carries information on underlying evolutionary or epidemiological processes, including speciation and extinction (38–41), natural selection (42–45), and infectious disease transmission (46–48). The other parameter $\alpha \in (-\infty, \infty)$ regulates the relationship between clade family size and proximity to the root (clade size–age relation), which is fundamental to understanding the effects of ecological, evolutionary, geographic, and other factors on biodiversity (7, 49–52). More details about these models can be found in SI Appendix, sections 7 and 8.

We evaluate and compare our metrics, $d_1$ and $d_2$, with the adaptations of other metrics to the space of ranked tree shapes (Table 1)—$d_{\text{BHVT-RST}}$, $d_{\text{KC-RST}}$, and $d_{\text{CP-RST}}$—by the following methods. We summarize each sampling distribution by the summary statistics, the $L_2$ medoid (Eq. 5) and the dispersion (Eq. 6). We then assess the performance of the metrics in distinguishing different ranked tree shape distributions by several measures: the confusion matrix, mean separation, and mean confusion (Eq. 8) and its permutation $P$ value (Eq. 9) for testing equality between pairs of sampling distributions (107 permutations). Additionally, for ease of visualization and interpretation, we use multidimensional scaling (MDS) (53) to embed our distance metrics into Euclidean spaces of two dimensions.

Distinguishing Different Tree Balance Distributions. To show how our proposed metrics can be used to differentiate ranked tree shapes sampled from distributions with different degrees of tree balance, we simulated ranked tree shapes under the alpha–beta-splitting model with different values of the tree balance parameter ($\beta$) while keeping the clade size–age relation parameter ($\alpha$) fixed. We considered $\beta \in [-1.9, -1.5, -1.0, 0, 100]$, representing a sequence from unbalanced to balanced ranked tree shapes. We refer to the ranked tree shape distributions with their corresponding well-known models of speciation: the Yule model (54, 55) ($\beta = 0$), the proportional-to-distinguishable-arrangements (PDA) model (41) ($\beta = -1.5$), and the Aldous branching (AB) model (40) ($\beta = -1$). Additionally, we refer to $\beta = -1.9$ and $\beta = 100$ as “unbalanced” and “balanced”, respectively. We simulated 1,000 ranked tree shapes with $n = 100$ leaves for each $\beta$ value, generating 5,000 simulated ranked tree shapes. We then computed the pairwise distance matrices of size $5,000 \times 5,000$ with $d_1$, $d_2$, $d_{\text{BHVT-RST}}$, $d_{\text{KC-RST}}$, and $d_{\text{CP-RST}}$.

The confusion matrices displayed in SI Appendix, Table S2 show the greatest mean separation across the five distributions for $d_1$ and $d_2$: About 83% of the trees are closest to the $L_2$ medoid of their originating distribution with $d_1$ and $d_2$ distances, 70.5% with $d_{\text{KC-RST}}$, 75.0% with $d_{\text{CP-RST}}$, and only 20.3% with $d_{\text{BHVT-RST}}$. Our $d_1$ and $d_2$ distances discriminate tree distributions with different tree balance parameters to a greater extent than $d_{\text{BHVT-RST}}$, $d_{\text{CP-RST}}$, and $d_{\text{KC-RST}}$ show more similar performance.

To test for equality in sampling distributions, we computed the pairwise mean confusions and associated $P$ values. SI Appendix, Table S3 A and B shows mean confusions computed using our metrics, $d_1$ and $d_2$. All off-diagonal values are smaller than 0.21 and statistically significant ($P < 0.0001$). In contrast, comparisons with $d_{\text{BHVT-RST}}$ and $d_{\text{KC-RST}}$ (SI Appendix, Table S3 C and D) produce higher pairwise mean confusion values across all pairs—with off-diagonal values larger than 0.45 with $d_{\text{BHVT-RST}}$ and values between 0.04 and 0.28 with $d_{\text{KC-RST}}$—and thus, lower discrimination. Mean confusion values closer to 0 are indicative of good discrimination. Comparisons with $d_{\text{CP-RST}}$ (SI Appendix, Table S3E) display similar pairwise mean confusion values and significance to our distances (all $P$ values are significant at $P < 0.0001$).

SI Appendix, Table S4A shows the dispersion statistic for each tree distribution and for each distance function. The unbalanced sample shows the smallest dispersion value (all sampled
trees are closer to each other) with our distances and \( d_{\text{BH-V-RTS}} \), whereas, with \( d_{\text{KC-RTS}} \) and \( d_{\text{CP-RTS}} \), it displays the largest dispersion value (more separation among trees within the same distribution).

Our findings are confirmed visually through the MDS plots displayed in Fig. 4. In the MDS plots, each dot corresponds to a tree, and each color corresponds to the sampling distribution for a specified \( \beta \) value. Fig. 4C shows the \( L_2 \)-medoid trees using our \( d_1 \) distance. The total distances explained by the first two MDS dimensions are 91.4\% using \( d_1 \) (Fig. 4A) and 94.4\% using \( d_2 \) (Fig. 4B). The two-dimensional MDS mappings using the other distances explain less than 80\%.

**Distinguishing Different Internal Node Ranking Distributions.** To show how our proposed metrics can be used to differentiate ranked tree shapes sampled from distributions with different internal node rankings, we generated 1,000 random trees from the alpha-beta-splitting model with \( n = 100 \) tips, fixed \( \beta = 0 \), and varying values of \( \alpha \) in \( \{-2, -1, 0, 1, 2\} \), producing 5,000 total trees. By varying the value of \( \alpha \) while keeping the balance parameter \( \beta \) fixed, we test the performance of our metrics in distinguishing between ranked tree shapes with small family sizes closer to the root (\( \alpha < 0 \)) and ranked tree shapes with small family sizes closer to the tips (\( \alpha > 0 \)).

The confusion matrices displayed in SI Appendix, Table S4 show greater mean separation for \( d_1 \) and \( d_2 \) than for any of the other distances. More than 75\% of the trees are closest to the \( L_2 \)-medoid of their originating distributions with \( d_1 \) and \( d_2 \), and less than 35\% have this property with the other distances.

The pairwise mean confusions computed using \( d_1 \) and \( d_2 \) (SI Appendix, Table S5A and B) have values strictly smaller than the mean confusions computed with the other distances. All comparisons with our distances are statistically significant (\( P < 0.0001 \)), indicating good distinguishability of distributions with different \( \alpha \) values.

The dispersion statistics in SI Appendix, Table S1B indicate that, according to our \( d_1 \) and \( d_2 \) metrics, trees from the alpha-beta-splitting model with \( \alpha = 0 \) are the most dispersed group among the five \( \alpha \) values considered. This, however, is not the case with the other distances. In particular, no variations in dispersion across different distributions are observed using \( d_{\text{CP-RTS}} \), and a positive correlation between \( \alpha \) and dispersion is seen with \( d_{\text{BH-V-RTS}} \).

Our two-dimensional MDS representations confirm the findings observed in the confusion matrix and mean confusions: \( d_1 \) and \( d_2 \) (Fig. 5A and B), when compared to the other three metrics (Fig. 5D-F), distinguish to a greater extent between trees with positive and negative \( \alpha \) parameters. In addition, our \( d_1 \) and \( d_2 \) distances show tighter embeddings in two-dimensional MDS, with a high proportion of the total distance explained. Although the MDS visualizations in Fig. 5A and B show that most clusters of trees are well separated according to their sampling distributions with \( d_1 \) and \( d_2 \), a large overlap exists between groups of trees with \( \alpha = 1 \) and \( \alpha = -2 \). The observed similarity is evident from the \( L_2 \)-medoid trees (Fig. 5C).
Our results confirm that our metrics capture both tree balance and clade size-age relation more effectively than the other adapted metrics on ranked tree shapes, in that simulation models differing in these features are more easily discriminated with our distances than with the others. To further demonstrate this point, we simulated 1,000 trees in each of the varying tree balance and internal node rankings: \((\alpha, \beta) = (-1, 0), (1, -1.5), (1, 0), (1, 0.5)\). The groups with the same \(\beta\) value share the same tree balance distribution, and the groups with the same \(\alpha\) value share the same internal node ranking distribution. In SI Appendix, Table S6 and Fig. S8 show that \(d_1\) and \(d_2\) effectively separate all four distributions, while \(d_{KC-RTS}\) and \(d_{CP-RTS}\), designed to measure differences in tree shapes only, separate tree distributions with the same tree balance but with different internal node rankings less effectively.

**Separation between Ranked Genealogies with Different Distributions.** Having investigated the utility of our metrics in separating different ranked tree shape distributions, we now demonstrate our metrics on the space of ranked genealogies. We consider two scenarios with evolutionary and epidemiological importance: isochronous ranked genealogy distributions resulting from different demographic histories and heterochronous ranked genealogy distributions inferred from human influenza A/H3N2 virus data from two geographical regions. We evaluate and compare our metrics, \(d_w^1\) and \(d_w^2\), with the adaptations of other metrics to the space of ranked genealogies (Table 1)—\(d_{BHVRG}\) and \(d_{KC-RG}\)—using the same measures we employed in the previous section. We note that there is no adaptation of the CP metric to the space of ranked genealogies, as the CP metric is defined on the space of tree shapes and, thus, does not incorporate tree branch lengths.

**Distinguishing Different Demographic Histories.** We assess how our proposed weighted metrics can differentiate genealogies with different branch length distributions. In coalescent models, the rate of coalescence (or branching) depends on the effective population size \(\lambda(t)\), as shown in Eq. 10. Under the neutral coalescent, the tree topology distribution (SI Appendix, Eq. S2) is independent of the branch lengths (see Materials and Methods for details). To investigate the performance of our weighted metrics \(d_w^1\) or \(d_w^2\) on genealogies in separating trees according to their branch length distribution, we generated 1,000 random ranked genealogies with \(n = 100\) leaves from the Tajima coalescent with different parameter values chosen for the seasonal logistic trajectory, the seasonal epidemic trajectory, and the constant population size model \(N_0 = 10^4\). The functional form and parameter values chosen for the seasonal logistic trajectory mimic estimated trajectories of human influenza A virus in temperate regions (56). We removed the leaf labels and retained branch lengths to produce isochronous ranked genealogies. We produced pairwise distance matrices of 3,000 \(\times\) 3,000 using the weighted metrics \(d_w^1\) and \(d_w^2\).

The dispersion statistics in SI Appendix, Table S9A show that the sampled genealogies under the constant population size trajectory have the greatest dispersion, and the sampled genealogies under the exponential growth trajectory have the least dispersion.

Our metrics’ ability to distinguish different demographic scenarios is evident in the confusion matrix in SI Appendix, Table S10, where they display good performance. Across all three distributions, 99.8% of the trees are closest to the \(L_2\) medoid of their originating distribution with \(d_w^1\) and \(d_w^2\); corresponding values are 33.4% with \(d_{BHVRG}\) and 74.0% with \(d_{KC-RG}\).

To test for equality in sampling distributions, we computed the mean confusion statistics and \(P\) values. In SI Appendix, Table S11 \(A\) and \(B\) display all off-diagonal mean confusion statistics that are less than 0.002 and are statistically significant (\(P < 0.00001\)) with \(d_w^1\) and \(d_w^2\), indicating a high level of discrimination between ranked genealogy distributions with different demographic histories. The mean confusion computed with \(d_{BHVRG}\) and \(d_{KC-RG}\) have far higher values than the values with \(d_w^1\) and \(d_w^2\) (SI Appendix, Table S11 \(C\) and \(D\)).

The two-dimensional MDS plots, Fig. 6 \(A\) and \(B\), display that our weighted metrics distinguish the three simulated ranked genealogy distributions, visually confirming the same finding using the confusion matrix and the mean confusion statistic. The first two dimensions of MDS account for 72.1% of the total distance, with the first MDS coordinate separating the constant trajectory from the others and the second coordinate...
distinguishing the exponential trajectory from the logistic trajectory. A comparison of panels within Fig. 6 shows that $d_{\text{BHV-RG}}$ is far less informative than either of our metrics.

**Analysis of Human Influenza A/H3N2 Virus from Different Geographical Regions.** We next apply our metrics to ranked genealogies sampled from the posterior distributions of human influenza A/H3N2 gene sequences from two geographic regions—New York and Southeast Asia. In temperate climates, such as New York, influenza epidemics display seasonality, with peaks occurring during the winter months, whereas in regions with tropical or subtropical climates, such as Southeast Asia, influenza activity persists throughout the year (57). We reanalyzed a subset of the sequences used in a previous study of Bahl et al. (58) that showed patterns of seasonal dynamics of influenza in New York and stable dynamics in Southeast Asia. We use our metrics to assess the differences in genalogical posterior distributions between the two regions.

The frequency of the collected samples by collection date and the estimated effective population size trajectories (Fig. 7) show distinct patterns consistent with those in Bahl et al. (58): a seasonal effective population size trajectory in New York that peaks during the winter seasons and a flat effective population size trajectory in Southeast Asia.

For each geographical region, we sampled 1,000 ranked genealogies per region from posterior distributions obtained from BEAST. We refer to the resulting samples from New York and Southeast Asia as NY–NY and SEA–SEA, respectively.

To investigate the effect of the different sampling schedules, we simulated two additional groups of 1,000 trees: one with fixed New York sampling schedule (Fig. 7A) and Southeast Asia estimated effective population size trajectory (Fig. 7D) and the other with fixed Southeast Asia sampling schedule (Fig. 7B) and New York estimated effective population size trajectory (Fig. 7C). We denote the former by NY–SEA and the latter by SEA–NY.

We then aggregated the 4,000 ranked genealogies—1,000 from each of NY–NY, SEA–SEA, NY–SEA, and SEA–NY—and computed pairwise distance matrices of $4,000 \times 4,000$ with $d_w^n$, $d_w^\text{BHV-RG}$, and $d_w^\text{KC-RG}$. The corresponding confusion matrices (SI Appendix, Table S12) show that $d_w^n$ and $d_w^\text{BHV-RG}$ have good performance in discriminating all four distributions, both with mean separation of 99.9%, while $d_w^\text{BHV-RG}$ and $d_w^\text{KC-RG}$ display less discrimination, with mean separations of 28.9 and 65.6%, respectively.

As shown in SI Appendix, Table S13 A and B, the observed mean confusion statistics between all sampling distribution pairs with our distances are near zero and statistically significant ($P < 0.0001$). While all observed off-diagonal mean confusion values with $d_w^\text{KC-RG}$ (SI Appendix, Table S13D) are larger than the values with $d_w^n$ or $d_w^\text{BHV-RG}$, all $P$ values are significant, showing somewhat comparable performance with our distances. The test of equality in sampling distributions is not rejected at the significance level of 5% with $d_w^\text{BHV-RG}$ between the pair (NY–SEA, NY–NY) and between the pair (SEA–NY, NY–NY) (SI Appendix, Table S13C).

Fig. 8 A and B shows that all four distributions with different sampling and evolutionary histories are also well separated in the two-dimensional MDS plots using $d_w^n$ and $d_w^\text{BHV-RG}$. However, we note that the first two dimensions of MDS account for only 51.7% ($d_w^n$) and 50.5% ($d_w^\text{BHV-RG}$) of the total distance.

**Fig. 7.** Human influenza A/H3N2 virus collection dates and inferred effective population size trajectories. (A) Collection date histogram, New York. (B) Collection date histogram, Southeast Asia. (C) Inferred effective population size trajectory with BEAST, New York. (D) Inferred effective population size trajectory with BEAST, Southeast Asia.
In the previous section, we compared posterior samples of ranked genealogies derived from MCMC procedures that introduce correlation in the samples. Although we thinned our posterior samples, it is important to test the mixing of the Markov chains. We used our metrics to assess convergence of each of the three MCMC chains for the two geographical regions. We computed the distance between the running posterior $L_2$-medoid ranked genealogy and the global posterior $L_2$-medoid ranked genealogy after every $10^5$ iterations for each chain. That is, in each chain, we sampled 1,000 ranked genealogies from the posterior distribution so that the $k$th sample $\mathbf{g}_k$ corresponds to iteration $k \times 10^5$ of the chain. For each $k = 1, \ldots, 1,000$, we computed a running $L_2$-medoid $\overline{\mathbf{g}}_k$ of all samples up to and including the $k$th ranked genealogy using our metric. We then computed distances from the running $L_2$ medoids $\mathbf{g}_k (k = 1, \ldots, 1,000)$ to the overall $L_2$ medoid of all of the 1,000 sampled ranked genealogies $\overline{\mathbf{g}}$ ($= \mathbf{g}_{1,000}$).

Fig. 9 shows $d^{w}_i (\mathbf{g}_k, \mathbf{g})$ ($i = 1, 2$) as a function of the chain iteration index for the three independent runs for each geographical region. The initial variability of the running-mean plot is relatively high, but, after an initial burn-in period ($\approx 10\%$), the distance from the running mean to the overall mean stabilizes quickly with an increasing number of states indicating convergence.

**Discussion**

Ranked tree shapes and ranked genealogies are used to model dependencies between samples of molecular data in the fields of population genetics and phylogenetics. These tree structures have a particular value in studying heterogeneous data and in providing a resolution that is fine enough to be informative but is computationally more feasible than finer resolutions.

In this article, we equipped the spaces of ranked tree shapes and ranked genealogies with a metric. By defining a bijection between ranked tree shapes and a triangular matrix of integers, we defined distance functions between ranked tree shapes as the $L_1$ and $L_2$ norms of the difference between two matrices. Our distances on ranked genealogies are the $L_1$ and $L_2$ norms of the difference between two matrices weighted by the genealogy branch lengths. To apply our distances to real data for comparing ranked genealogies or ranked tree shapes with different sampling events, we proposed an augmentation scheme, increasing the dimension of the matrix representations, and we defined our distances as the $L_1$ and $L_2$ norms of the difference between extended matrices of the same size. We used our distances to summarize the distribution of trees and proposed using the $L_2$-medoid tree set, a sample version of the Fréchet mean, as the central set of a sample of trees. We proposed a sample version of the Fréchet variance to quantify uncertainty. We demonstrated the utility of our metrics using simulated data from models with biological relevance and human influenza A/H3N2 virus between temperate and tropical regions.

The performance of the proposed metrics was evaluated through their ability to distinguish different tree distributions quantified by the confusion matrix, mean separation, and mean confusion statistic. We statistically assessed equality in distributions with a nonparametric test. This feature is particularly relevant where analytic expressions for underlying probability distributions are unknown. We note that our nonparametric test
assumes samples are independent, and a future direction is to account for the nonindependence of samples. Another important future direction is to develop statistical tests based on our metrics for uncovering specific evolutionary events, for example, inferring mode and strength of natural selection and identifying specific genomic regions undergoing selective pressure, beyond detecting differences in evolutionary histories from tree distributions.

Our metrics provide the basis for a decision-theoretic statistical inference that can be constructed by finding the best-ranked genealogy that minimizes the expected error or loss function, which, in turn, is a function of the tree distance. Further, our proposed metric provides a tool for evaluating convergence and the mixing of MCMC procedures on ranked genealogies. We demonstrate the applicability of our metrics as an MCMC convergence diagnostic with influenza data.

We propose our metrics be constructed by finding the best-ranked genealogy with the following parameters. In distinguishing different sampling events were generated using phylodyn::coalsim (63) were kept at their default values. Heterochronous ranked tree shapes with different branch length distributions.

Materials and Methods

Random Ranked Tree Shape Generation. To generate random isochronous ranked tree shapes, we used the simulate function in the R package apTreeshape (7) with the following parameters. In distinguishing different internal node ranking distributions, we simulated 1,000 ranked tree shapes with \( n = 100 \) leaves for each \( \beta \) value in \([-1.9, -1.5, -1.0, 100]\) with fixed \( \alpha = 1 \). In distinguishing different internal node ranking distributions, we simulated 1,000 ranked tree shapes with \( n = 100 \) leaves for each \( \alpha \) value in \([-2, -1.0, 1.2]\) with fixed \( \beta = 0 \). All other parameters were kept at their default values. Heterochronous ranked tree shapes with different sampling events were generated using phylodyn::coalsim (63) in R.

Random Ranked Genealogy Generation. For all our simulations, we assume neutral evolution. Hence, the distribution of branching or coalescence waiting times depends on the number of extant branches and is independent of the particular topology. In the isochronous coalescent model, when there are \( k \) lineages, the coalescent time \( t_k \) is time in from tips to root (Fig. 2A) has conditional density

\[
f(t_k | t_{k+1}) = \frac{(t_k)}{\lambda(t_k)} e^{-\frac{(t_k)}{\lambda(t_k)}} dt_k, \quad \text{for } k = n, \ldots, 2,
\]

where \( \lambda(t) > 0 \) is the effective population size at time \( t \) and \( t_{n+1} = 0 \). In the heterochronous coalescent, sampling times are deterministic and the rate of coalescence depends on the number of extant lineages and the effective population size \( \lambda(t) \). More details about coalescent models can be found in SI Appendix, sections 9 and 10.

We used phylodyn::coalsim (63) in R to generate random isochronous ranked genealogies with different branch length distributions and random heterochronous ranked genealogies with different sampling events and different branch length distributions.

Comparison with Other Metrics. For comparing our metrics to adaptations of other tree distances, we used the Geodesic Treepath Problem (GTP) software implemented in Java (31) for the BHV distance, multiDist implemented in the R package treespace (10) for the KC distance, and the vecMultiDistUnlab in the R package treespace (14) for the CP distance.

Human Influenza A/H3N2 Virus Analysis. We selected 122 complete HA sequences at random with overlapping collection dates between New York and Southeast Asia—January 2002 to September 2005. The National Center for Biotechnology Information (NCBI) Genbank accession numbers for the 122 HA sequences from New York and Southeast Asia used in this article are provided in SI Appendix, Table S14. We aligned the sequences using the FFT-NS-i option of Multiple Alignment using Fast Fourier Transform (MAFFT) v7.427 software (64, 65).

We used BEAST v1.10.4 (16) to sample from the posterior distribution of model parameters with the following prior settings: Shapiro–Rambaut–Drummond codon-partition substitution model (66) to accommodate rate heterogeneity among sites, the uncorrelated log-normal relaxed molecular clock, and a time-aware Bayesian Skyride prior model (67) on evolutionary histories. For each geographical region, we ran three independent MCMC chains with 100 million steps, with 10% of burn-in, and thinned to obtain a total of 1,000 samples per region. We assessed convergences and effective sample sizes of the estimates using Tracer v1.7.1 (68) and with our distances.

Code and Data Availability. An implementation of our metrics is available at https://github.com/JuliaPalacios/phylodyn. The NCBI Genbank accession numbers for the human influenza A/H3N2 HA sequences are provided in SI Appendix, Table S14.

ACKNOWLEDGMENTS. We acknowledge Nina Miolane, Susan Holmes, and Emmanuel Candès for useful discussions. We thank reviewers for helpful comments and Julia Olivieri for her initial feasibility study of one of the proposed distances. J.A.P. and N.A.R. acknowledge support from National Institutes of Health Grant R01-GM-131404. J.A.P. acknowledges support from the Alfred P. Sloan Foundation.

1. R. Sainudin, T. Stadler, A. Véber, Finding the best resolution for the Kingman-Tajima coalescent: Theory and applications. J. Math. Biol. 70, 1207–1247 (2015).
2. J. A. Palacios et al., Bayesian estimation of population size changes by sampling Tajima's trees. Genetics 213, 967–986 (2019).
3. F. Tajima, Evolutionary relationship of DNA sequences in finite populations. Genetics 105, 437–460 (1983).
4. J. Schweinsberg, Coalescents with simultaneous multiple collisions. Electron. J. Probab. 5, 1–50 (2000).
5. D. Ford, F. A. Matsen, T. Stadler, A method for investigating relative timing information on phylogenetic trees. Syst. Biol. 58, 167–183 (2009).
6. A. Lambert, T. Stadler, Birth-death models and coalescent point processes: The shape and probability of reconstructed phylogenies. Theor. Popul. Biol. 90, 113–128 (2013).
7. O. Mallet, F. Gascuel, A. Lambert, Ranked tree shapes, nonrandom extinctions, and the loss of phylogenetic diversity. Syst. Biol. 67, 1025–1040 (2018).
8. D. Robinson, L. Foulds, Comparison of phylogenetic trees. Math. Biosci. 53, 131–147 (1981).
9. L. J. Billera, S. Holmes, K. Vogtmann, Geometry of the space of phylogenetic trees. Adv. Appl. Math. 27, 733–767 (2001).
10. M. Kendall, C. Collin, Mapping phylogenetic trees to reveal distinct patterns of evolution. Mol. Biol. Evol. 33, 2785–2743 (2016).
11. J. Chakerian, S. Holmes, Computational tools for evaluating phylogenetic and hierarchical clustering trees. J. Comput. Graph. Stat. 21, 581–599 (2012).
12. D. G. Brown, M. Owen, Mean and variance of phylogenetic trees. Syst. Biol. 69, 139–154 (2020).
13. A. Willis, R. Bell, Confidence sets for phylogenetic trees. J. Comput. Graph Stat. 27, 542–552 (2018).
14. C. Colin, G. Piazzotta, A metric on phylogenetic tree shapes. Syst. Biol. 67, 113–126 (2018).
15. S. Holmes, Statistics for phylogenetic trees. Theor. Popul. Biol. 63, 17–32 (2003).
16. M. A. Suchard et al., Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. Virus Evol. 4, vey016 (2018).
17. E. M. Volz, K. Koelle, T. Bedford, Viral phydynamics. PLoS Comput. Biol. 9, 1–12 (2013).
18. E. C. Holmes, G. Dudas, A. Rambaut, K. G. Andersen, The evolution of Ebola virus: Insights from the 2013-2016 epidemic. Nature 538, 193–200 (2016).
19. M. Worobey et al., 1970s and 'Patient 0' HIV-1 genomes illuminate early HIV/AIDS history in North America. Nature 539, 98–101 (2016).
20. X. Li et al., Evolutionary history, potential intermediate animal host, and cross-species analyses of SARS-CoV-2. J. Med. Virol. 92, 602–611 (2020).
21. M. Giovanetti, D. Benvenuto, S. Angeletti, M. Cicozzi, The first two cases of 2019-nCoV in Italy: Where they came from?. J. Med. Virol. 92, 518–521 (2020).
46. G. E. Leventhal et al., Using ancient DNA and coalescent-based methods to infer extinction. Biol. Lett. 12, 20150822 (2016).

45. L. Ferretti, A. Ledda, T. Wiehe, G. Achaz, S. E. Ramos-Onsins, Decomposing the site frequency spectrum: The impact of tree topology on neutrality tests. Mol. Biol. Evol. 37, 44–77 (2011).

44. M. Pokorny, The signature of positive selection at randomly chosen loci. Genetics 160, 1179–1189 (2002).

43. M. Przeworski, The signature of positive selection at randomly chosen loci. Genetics 160, 1179–1189 (2002).

42. D. Aldous, Stochastic models and descriptive statistics for phylogenetic trees, from Yule to today. Stat. Sci. 16, 23–34 (2001).

41. M. G. B. Blum, O. Francois. Which random processes describe the tree of life? A large-scale study of phylogenetic tree imbalance. Syst. Biol. 55, 685–691 (2006).

40. D. Aldous, “Probability distributions on cladograms” in Combinatorial Mathematics VI, A. F. Horadam, W. D. Wallis, Eds. (Springer, Berlin/Heidelberg, Germany, 1979), pp. 119–126.

39. M. Li, L. Zhang, Twist–rotation transformations of binary trees and arithmetic expressions. J. Algorithm 32, 155–166 (1999).

38. E. Lehmann, J. P. Romano, Testing Statistical Hypotheses. (Springer, New York, NY, ed. 3, 2005).

37. N. Galter, F. Depauwls, N. H. Barton, Detecting bottlenecks and selective sweeps from DNA sequence polymorphism. Genetics 155, 981–987 (2000).

36. M. Przeworski, The signature of positive selection at randomly chosen loci. Genetics 160, 1179–1189 (2002).

35. M. Bordewich, C. Semple, On the computational complexity of the rooted subtree prune and regraft distance. Ann. Combinator. 8, 409–423 (2005).

34. M. Owen, J. S. Provan, A fast algorithm for computing geodesic distances in tree space. IEEE ACM Trans. Comput. Biol. Bioinf. 8, 2–13 (2011).

33. E. Lewitus, H. Morlon, Characterizing and comparing phylogenies from their Laplacian spectrum. Syst. Biol. 65, 495–507 (2015).

32. M. Li, G. E. Leventhal, G. E. Leventhal et al., Rise and fall of the Beringian steppe bison. Science 306, 1561–1565 (2004).

31. M. Owen, J. S. Provan, A fast algorithm for computing geodesic distances in tree space. IEEE ACM Trans. Comput. Biol. Bioinf. 8, 2–13 (2011).

30. M. Bordewich, C. Semple, On the computational complexity of the rooted subtree prune and regraft distance. Ann. Combinator. 8, 409–423 (2005).

29. M. Owen, J. S. Provan, A fast algorithm for computing geodesic distances in tree space. IEEE ACM Trans. Comput. Biol. Bioinf. 8, 2–13 (2011).

28. M. Li, L. Zhang, Twist–rotation transformations of binary trees and arithmetic expressions. J. Algorithm 32, 155–166 (1999).

27. D. F. Robinson, L. R. Foulds, “Comparison of weighted labelled trees” in Combinatorial Mathematics VI, A. F. Horadam, W. D. Wallis, Eds. (Springer, Berlin/Heidelberg, Germany, 1979), pp. 119–126.

26. S. Peyrégne, Computing medians and means in Hadamard spaces. IEEE Trans. Comput. Syst. Biol. 4, 64, 432–440 (2015).

25. K. Mardia, Some properties of classical multi-dimensional scaling. Commun. Stat. Theor. Methods 7, 1233–1241 (1978).

24. D. Aldous, R. Pemantle, Eds. (Springer, New York, NY, ed. 3, 2005).

23. D. Chang, B. Shapiro, Using ancient DNA and coalescent-based methods to infer extinction. Biol. Lett. 12, 20150822 (2016).

22. B. Shapiro, J. G. Milligan. Modeling the shapes of phylogenetic trees from human and zoonotic RNA viruses. PLoS One 8, 1–11 (2013).

21. B. Shapiro, J. G. Milligan, Modeling the shapes of phylogenetic trees from human and zoonotic RNA viruses. PLoS One 8, 1–11 (2013).

20. M. Li, L. Zhang, Twist–rotation transformations of binary trees and arithmetic expressions. J. Algorithm 32, 155–166 (1999).

19. D. Aldous, Stochastic models and descriptive statistics for phylogenetic trees, from Yule to today. Stat. Sci. 16, 23–34 (2001).

18. M. G. B. Blum, O. Francois, Which random processes describe the tree of life? A large-scale study of phylogenetic tree imbalance. Syst. Biol. 55, 685–691 (2006).

17. N. Galter, F. Depauwls, N. H. Barton, Detecting bottlenecks and selective sweeps from DNA sequence polymorphism. Genetics 155, 981–987 (2000).

16. M. Przeworski, The signature of positive selection at randomly chosen loci. Genetics 160, 1179–1189 (2002).

15. H. Li, T. Wiehe, Coalescent tree imbalance and a simple test for selective sweeps based on microsatellite variation. PLoS Comput. Biol. 9, e1003060 (2013).

14. L. Ferretti, A. Ledda, T. Wiehe, G. Achaz, S. E. Ramos-Onsins, Decomposing the site frequency spectrum: The impact of tree topology on neutrality tests. Genetics 207, 229–240 (2017).

13. G. E. Leventhal et al., Inferring epidemic contact structure from phylogenetic trees. PLoS Comput. Biol. 8, 1–10 (2012).