

Supplementary Information for

Distance metrics for ranked evolutionary trees

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

Corresponding author. E-mail: juliapr@stanford.edu

This PDF file includes:

Supplementary text
Figs. S1 to S9
Tables S1 to S16
SI References

Supporting Information Text

1. Proof of Theorem 1

Proof. Consider a ranked tree shape T^R with n leaves sampled at m different sampling times. We denote the total number of change points in T^R by $r = n + m - 1$ and its ordered change point times by $(u_r, u_{r-1}, \dots, u_1)$, $0 = u_r < u_{r-1} < \dots < u_1$, with time increasing into the past (rootward). The internal nodes of T^R are labeled by the indices of their coalescent times, and all leaves of T^R are labeled by the indices of their sampling times. We note that for convenience, internal nodes are no longer labeled $2, \dots, n$ from the root to leaves, but they are labeled by their time-event indices (see Figure S1). Each internal node has a unique label, but the leaf nodes with the same sampling time share the same label. We define $N = \{1, \dots, r\}$ to be a set of all node labels, I to be a set of all internal node labels, and S to be a set of all leaf node labels. Note that I and S are disjoint and contain $n - 1$ and m elements, respectively.

For $i \in I$, let $o_i = (x_{i,1}, x_{i,2})$ denote the ordered pair of labels of the two immediate descendants of internal node i , such that $i < x_{i,1} \leq x_{i,2}$. We denote the set of all pairs i and $o_i = (x_{i,1}, x_{i,2})$ choices in T^R by $X = \{(i, o_i) \mid i \in I\}$. Then X completely specifies T^R : T^R is a directed graph from the root to tips and X encodes its adjacency matrix and the order of the internal node indices $i \in I$ determines internal node rankings.

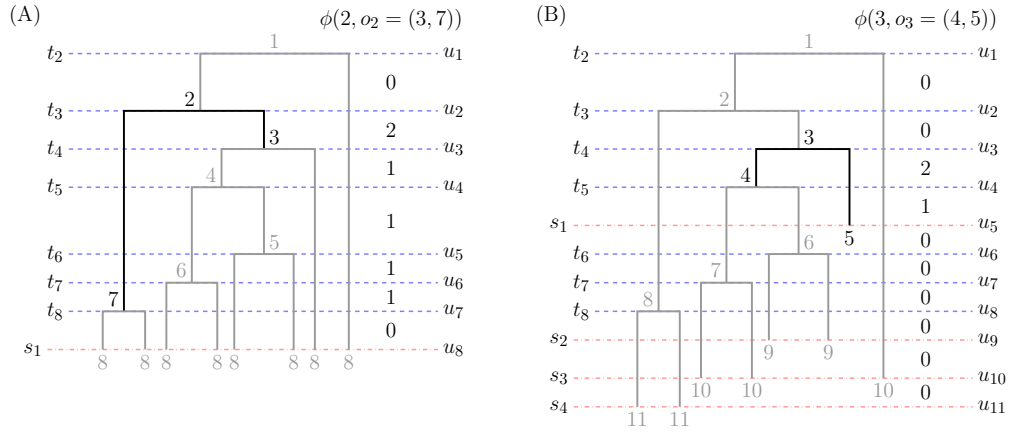


Fig. S1. Example of ϕ mapping. (A) An example isochronous ranked tree shape. The set of internal node labels is $I = \{1, 2, 3, 4, 5, 6, 7\}$ and the set of leaf node labels is $S = \{8\}$. For convenience, internal nodes are labeled by their time-event indices throughout the proof. The internal node with label 2 at time u_2 has descendant nodes 3 and 7 at time u_3 and u_7 , respectively ($o_2 = (3, 7)$). The column vector $\phi(2, o_2 = (3, 7)) = (0, 2, 1, 1, 1, 1, 0)$ indicates the number of direct descendants of node 2 at each change point time interval. (B) An example heterochronous ranked tree shape. The set of internal node labels is $I = \{1, 2, 3, 4, 6, 7, 8\}$ and the set of leaf node labels is $S = \{5, 9, 10, 11\}$. The internal node with label 3 at time u_3 has descendants node 4 and node 5 at time u_4 and u_5 , respectively ($o_3 = (4, 5)$). The column vector $\phi(3, o_3 = (4, 5)) = (0, 0, 2, 1, 0, 0, 0, 0, 0, 0)$ indicates the number of direct descendants of node 3 at each change point time interval.

We define a function $\phi : X \rightarrow \{0, 1, 2\}^{r-1}$ as follows:

$$\phi_k(i, o_i) = \begin{cases} 0 & \text{if } 1 \leq k < i \\ 2 & \text{if } i \leq k < x_{i,1} \\ 1 & \text{if } x_{i,1} \leq k < x_{i,2} \\ 0 & \text{if } x_{i,2} \leq k < r. \end{cases}$$

The k th element of $\phi(i, o_i)$ is the number of immediate descendants of an internal node i present at the time interval (u_{k+1}, u_k) . ϕ is an injective map. To prove this, let (s, o_s) and (t, o_t) be two elements in X and let $(s, o_s) \neq (t, o_t)$. Because internal nodes of T^R are ranked, $(s, o_s) \neq (t, o_t)$ implies $s \neq t$; without loss of generality, assume $s < t$. By the definition of the map ϕ , the s th element of $\phi(s, o_s)$ is $\phi_s(s, o_s) = 2$, while the s th element of $\phi(t, o_t)$ is $\phi_s(t, o_t) = 0$ because $s < t$ and $t < x_{t,1} \leq x_{t,2}$. Thus, $\phi(s, o_s) \neq \phi(t, o_t)$ and ϕ is injective.

Let $\eta : \{1, \dots, r-1\} \times \{0, 1, 2\}^{r-1} \rightarrow \{0, 1, 2\}^{r-1}$ such that for $\mathbf{y} \in \{0, 1, 2\}^{r-1}$, the j th element of η is

$$\eta(k, \mathbf{y})_j = \begin{cases} 0 & \text{if } 1 \leq j < k \\ y_j & \text{if } k \leq j < r. \end{cases}$$

That is, $\eta(k, \mathbf{y})$ sets all the first $k-1$ entry values of \mathbf{y} to 0. Note that the first $i-1$ elements of $\phi(i, o_i)$ are 0 by definition and thus, $\eta(i, \phi(i, o_i)) = \phi(i, o_i)$.

Finally, for $T^R \in \mathcal{T}_n^R$, define $\psi : \mathcal{T}_n^R \rightarrow M_{r-1, r-1}(\mathbb{R})$, a function that maps a ranked tree shape with n leaves to a real valued square matrix of size $r-1$, with k th column:

$$\psi(T^R)_{\cdot, k} = \sum_{\substack{i \in I, \\ i \leq k}} \eta(k, \phi(i, o_i)),$$

where $\psi(T^R)_{\cdot,k}$ indicates the k th column of $\psi(T^R)$ and I is the set of all internal node labels as defined at the beginning of this section. By the definition of η , the first $k-1$ values of the k th column $\psi(T^R)_{\cdot,k}$ are 0, i.e., $\psi(T^R)$ is a lower triangular matrix. Because ϕ records the number of immediate descendants of a single internal node present at each time interval, $\psi(T^R)_{\cdot,k}$ tracks the sum of all surviving immediate descendants of internal nodes with labels $i \leq k$ starting from time interval (u_{k+1}, u_k) ; thus, $\psi(T^R)_{s,k}$, with $k \leq s$, represents the number of lineages of T^R in (u_{k+1}, u_k) that are still present at the time interval (u_{s+1}, u_s) .

We prove that ψ is an injective map. Let $T_1^R, T_2^R \in \mathcal{T}_n^R$ and $T_1^R \neq T_2^R$. Because $X = \{(i, o_i) \mid i \in I\}$ completely specifies T^R , $T_1^R \neq T_2^R$ implies that there exists an index $\ell \in \{1, \dots, n-1\}$ such that $(i_\ell^{(1)}, o_{i_\ell}^{(1)}) \neq (i_\ell^{(2)}, o_{i_\ell}^{(2)})$. Here, i_ℓ indicates the ℓ th element of I sorted in increasing order. If there is more than one such index, choose ℓ to be the smallest of them. Without loss of generality, let $i_\ell^{(1)} \leq i_\ell^{(2)}$. Then $\psi(T_1^R)_{\cdot, i_\ell^{(1)}} \neq \psi(T_2^R)_{\cdot, i_\ell^{(1)}}$ and thus $\psi(T_1^R) \neq \psi(T_2^R)$.

Hence, ψ maps each ranked tree shape T^R to a unique matrix, i.e., given an \mathbf{F} -matrix, if it encodes a ranked tree shape, it encodes exactly one ranked tree shape. \square

2. Proof of Proposition 2

Proof. To prove that d_k^w , $k = 1, 2$, is a metric, we need to prove the following properties hold for any ranked genealogies $\mathbf{g}_1^R, \mathbf{g}_2^R, \mathbf{g}_3^R$ with n leaves:

$$\begin{aligned} d_k^w(\mathbf{g}_1^R, \mathbf{g}_2^R) &\geq 0 && \text{non-negativity} \\ d_k^w(\mathbf{g}_1^R, \mathbf{g}_2^R) &= d_k^w(\mathbf{g}_2^R, \mathbf{g}_1^R) && \text{symmetry} \\ d_k^w(\mathbf{g}_1^R, \mathbf{g}_2^R) &= 0 \iff \mathbf{g}_1^R = \mathbf{g}_2^R && \text{identity} \\ d_k^w(\mathbf{g}_1^R, \mathbf{g}_2^R) &\leq d_k^w(\mathbf{g}_1^R, \mathbf{g}_3^R) + d_k^w(\mathbf{g}_3^R, \mathbf{g}_2^R) && \text{triangle inequality.} \end{aligned}$$

The non-negativity and symmetry are trivial. The triangle inequality follows from the Minkowski inequality of L_1 and L_2 norms. It remains to prove the identity property: $d_k^w(\mathbf{g}_1^R, \mathbf{g}_2^R) = 0$ if and only if $\mathbf{g}_1^R = \mathbf{g}_2^R$ for $k = 1, 2$. It is clear that $d_k^w(\mathbf{g}_1^R, \mathbf{g}_2^R) = 0$ if $\mathbf{g}_1^R = \mathbf{g}_2^R$ so we focus on $\mathbf{g}_1^R = \mathbf{g}_2^R$ if $d_k^w(\mathbf{g}_1^R, \mathbf{g}_2^R) = 0$. The following proof is for d_1^w . The proof for d_2^w follows the same arguments.

We assume that the two genealogies have the same number of sampling events m and same number of leaves n , so that the \mathbf{F} -matrices of \mathbf{g}_1^R and \mathbf{g}_2^R have the same dimension $(n+m-2) \times (n+m-2)$ dimension. We define $r = n+m-2$ for notational simplicity.

Because we allow only a single event at each change time point u_i , either coalescent or sampling, the first column of any \mathbf{F} -matrix is $(2, 1, \dots, 1)$ or $(2, 1, \dots, 1, 0, \dots, 0)$. For the latter, we denote the row index of the last occurrence of 1 in the first column by k_1 : $F_{k_1,1}^{(\ell)} = 1$ and $F_{k_1+1,1}^{(\ell)} = 0$ for some index $2 \leq k_1 \leq r$ and $\ell = 1, 2$.

If $\mathbf{F}^{(1)}$ and $\mathbf{F}^{(2)}$ have different first columns, then for some index $k_1 \geq 2$, $(F_{k_1,1}^{(1)}, F_{k_1,1}^{(2)}) = (0, 1)$ or $(F_{k_1,1}^{(1)}, F_{k_1,1}^{(2)}) = (1, 0)$. Because $\left| F_{i,j}^{(1)} W_{i,j}^{(1)} - F_{i,j}^{(2)} W_{i,j}^{(2)} \right| \geq 0$, $d_1^w(\mathbf{g}_1^R, \mathbf{g}_2^R) = 0$ implies $F_{i,j}^{(1)} W_{i,j}^{(1)} = F_{i,j}^{(2)} W_{i,j}^{(2)}$ for all i, j . Therefore, $W_{k_1,1}^{(2)} = 0$ in the first case and $W_{k_1,1}^{(1)} = 0$ in the second case. However, this contradicts our assumption of positive time interval between two change points, and thus $\mathbf{F}^{(1)}$ and $\mathbf{F}^{(2)}$ must have the same first column.

If both \mathbf{F} -matrices share the same first column, then $d_1^w(\mathbf{g}_1^R, \mathbf{g}_2^R) = 0$ implies $W_{i,1}^{(1)} = W_{i,1}^{(2)}$ for all $i = 1, \dots, r$. Recalling $W_{i,j} = u_j - u_{i+1}$, we have $u_1^{(1)} - u_{i+1}^{(1)} = u_1^{(2)} - u_{i+1}^{(2)}$. Because we assume $u_{r+1}^{(1)} = u_{r+1}^{(2)} = 0$, we can traverse through i in decreasing order starting from $i = r$ to get $u_i^{(1)} = u_i^{(2)}$ for all $i = 1, \dots, r+1$, which gives $\mathbf{W}^{(1)} = \mathbf{W}^{(2)}$.

Along with $\mathbf{W}^{(1)} = \mathbf{W}^{(2)}$, $F_{i,j}^{(1)} W_{i,j}^{(1)} = F_{i,j}^{(2)} W_{i,j}^{(2)}$ implies $F_{i,j}^{(1)} = F_{i,j}^{(2)}$ for all i, j , i.e., $\mathbf{F}^{(1)} = \mathbf{F}^{(2)}$, and thus $\mathbf{g}_1^R = \mathbf{g}_2^R$. \square

3. Metric spaces on heterochronous trees with different numbers of sampling events

We extend our distances to include cases in which the numbers of sampling events differ but the total number of samples remains the same.

Consider two heterochronous ranked tree shapes of n leaves, T_1^R and T_2^R , with different numbers of sampling events m_1 and m_2 , respectively. In order to compute the distance between T_1^R and T_2^R with our metrics, we require the two ranked tree shapes to be represented as \mathbf{F} -matrices of the same dimension. We accomplish this by inserting artificial sampling events. The detailed steps are as follows. Note that the following formulation is done going backwards in time with time increasing from the present to the past.

For $i = 1, 2$, let $\mathbf{E}^{(i)} = (e_{m_i+n-1}^{(i)}, \dots, e_1^{(i)})$ be the vector of ordered sampling and coalescence events where $e_{m_i+n-1}^{(i)}$ denotes the most recent sample event ($e_{m_i+n-1}^{(i)} = s$) assumed to occur at time $u_{m_i+n-1}^{(i)} = 0$. $e_1^{(i)} = c$ denotes the coalescent event at time $u_1^{(i)}$ corresponding to the most recent common ancestor of the samples in T_i^R . Each $e_j^{(i)}$ is either a sampling event ($e_j^{(i)} = s$) or a coalescent event ($e_j^{(i)} = c$). The event of type c occurs $n-1$ times and the event of type s occurs m_i times in $\mathbf{E}^{(i)}$. In the example illustrated in Figure S2, the event vectors for T_1^R (Figure S2(A)) and T_2^R (Figure S2(B)) are $\mathbf{E}^{(1)} = (s, c, s, c, s, c)$ and $\mathbf{E}^{(2)} = (s, s, s, c, s, c, c)$, respectively.

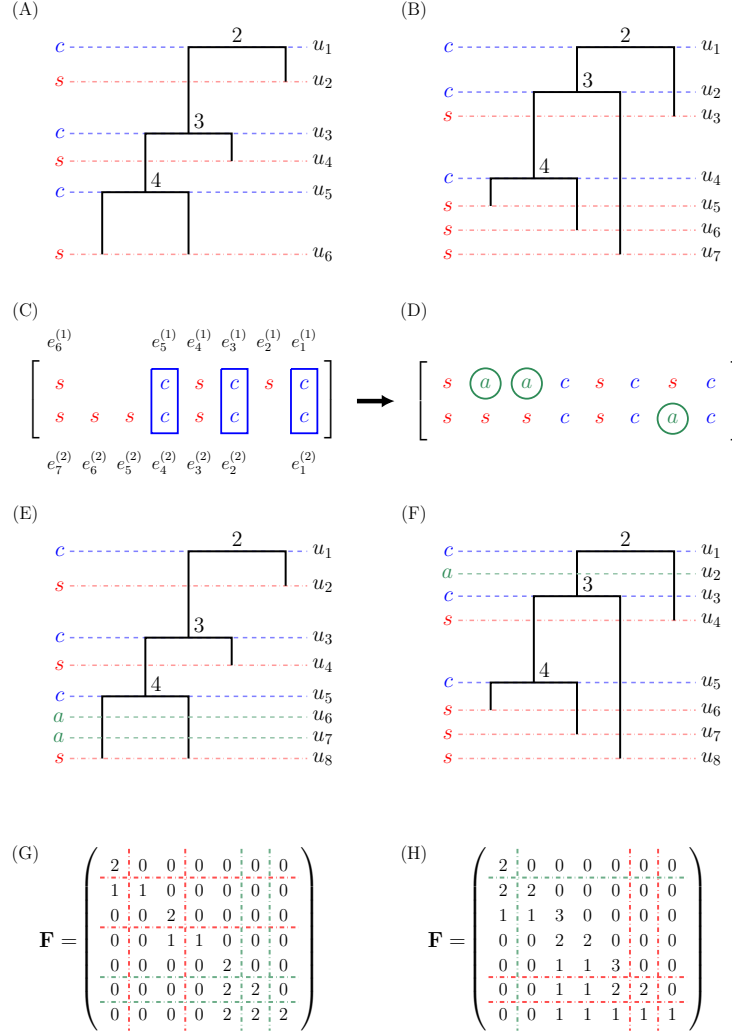


Fig. S2. Example of augmented \mathbf{F} -matrix representation of ranked tree shapes. In order to compute the d_1 or d_2 distances between ranked tree shapes with equally many samples but different numbers of sampling events, such as ranked tree shapes (A) and (B), we insert artificial sampling events with 0 samples in order to match their dimension. (A)-(B) Two ranked tree shapes of 4 samples with different sampling events. (C) Alignment of event vectors of two trees. The $n - 1$ coalescent events are aligned first by matching i th coalescent event of a tree to the i th coalescent event of the other tree ($i = 1, \dots, n - 1$). The sampling events are then matched by increasing index order in the event vector. (D) Augmentation of artificial sampling events a between coalescent events or between the first sampling and the first coalescent event. (E)-(F) Augmented ranked tree shapes. (G)-(H) Augmented \mathbf{F} -matrix representations.

We first align all $n - 1$ coalescent events between the two trees by adding empty spaces when needed as depicted in Figure S2(C). Once all the type- c events are aligned, we next align the sampling events between two successive coalescent events or between $t = 0$ and the first c event. When aligning the events of type s between the two trees, we pair the events of type s , starting from the most recent events. The event vector alignment is demonstrated in Figure S2(C). If one tree has more type- s events than the other in a given intercoalescent interval, we insert the excess artificial sampling events, denoted by a 's, in the tree with fewer type- s events in that interval. We assign 0 new samples to type- a events. For example, in the tree of Figure S2(A), there is only one sampling event before the first coalescent event, whereas there are three sampling events in the tree of Figure S2(B). In Figure S2(D), we illustrate how the two artificial sampling events are added to the first tree in the first interval. The resulting augmented trees are shown in Figures S2(E) and (F) along with their corresponding \mathbf{F} -matrix representations in Figures S2(G) and (H). We note that by construction, $\mathbf{F}^{(1)} \neq \mathbf{F}^{(2)}$ in these cases. Hence, $d_i(T_1^R, T_2^R) \neq 0$ for $i = 1, 2$.

Consider now two heterochronous ranked genealogies of n leaves, \mathbf{g}_1^R and \mathbf{g}_2^R , with different number of sampling events m_1 and m_2 respectively. In order to compute the distance between \mathbf{g}_1^R and \mathbf{g}_2^R we first augment their \mathbf{F} -matrix representations as with ranked tree shapes. In addition, we augment their weight matrices $W^{(1)}$ and $W^{(2)}$ by assigning a time to each augmented artificial sampling event. If n_a artificial events are inserted between events $e_{j+1}^{(i)}$ and $e_j^{(i)}$, we subdivide the corresponding time interval $[u_{j+1}^{(i)}, u_j^{(i)}]$ into $n_a + 1$ intervals with equal length: the times assigned to the n_a augmented artificial events are $\{u_{j+1}^{(i)} + \Delta, u_{j+1}^{(i)} + 2\Delta, \dots, u_{j+1}^{(i)} + n_a\Delta\}$, where $\Delta = \frac{u_j^{(i)} - u_{j+1}^{(i)}}{n_a + 1}$.

4. Statistical comparison of ranked tree shape and ranked genealogy sampling distributions

We generalize the mean confusion statistics for cases when the L_2 -medoid is not unique. This more general version is more appropriate for small sample spaces such as ranked tree shapes with small number of leaves.

For a given distance function d defined on the space of ranked tree shapes (or ranked genealogies), let \bar{X} be the L_2 -medoid set of the X_1, \dots, X_N samples and let \bar{Y} be the L_2 -medoid set of the Y_1, \dots, Y_N samples, as defined in Equation 5. The *generalized mean confusion* is defined as:

$$C_G^{x,y} = \frac{1}{2} \sum_{j=1}^N \left[\mathbb{1}_{\min_{\bar{Y}} \{d(X_j, \bar{Y})\} \leq \min_{\bar{X}} \{d(X_j, \bar{X})\}} \frac{1}{N - |\bar{X}|} \right. \\ \left. + \mathbb{1}_{\min_{\bar{X}} \{d(Y_j, \bar{X})\} \leq \min_{\bar{Y}} \{d(Y_j, \bar{Y})\}} \frac{1}{N - |\bar{Y}|} \right. \\ \left. - Z_j^x \mathbb{1}_{\min_{\bar{X}} \{d(X_j, \bar{X})\} = \min_{\bar{Y}} \{d(X_j, \bar{Y})\}} \frac{1}{N - |\bar{X}|} \right. \\ \left. - Z_j^y \mathbb{1}_{\min_{\bar{X}} \{d(Y_j, \bar{X})\} = \min_{\bar{Y}} \{d(Y_j, \bar{Y})\}} \frac{1}{N - |\bar{Y}|} \right], \quad [\text{S1}]$$

where Z_j^x and Z_j^y are i.i.d. Bernoulli random variables with probability 0.5 to resolve ties.

5. Adapting other tree metrics to ranked tree shapes and ranked genealogies

A. Other distances on ranked tree shapes. We start with two adaptations of metrics that are originally defined on the space of labeled genealogies: the BHV distance and the KC distance.

The Billera-Holmes-Vogtmann metric (BHV) metric. The BHV space (1) is obtained by representing each labeled genealogy \mathbf{g}_n^L with n leaves and edge set \mathcal{E} by a vector in the Euclidean orthant \mathbb{R}_+^{2n-1} , whose coordinates correspond to edge lengths. The BHV space is the union of $(2n-3)!!$ orthants. The BHV distance (d_{BHV}) between two labeled genealogies is defined as a geodesic, the shortest path connecting two points that lies inside the BHV space. Note that unranked and labeled genealogies with positive intervals between coalescent and sample times are effectively ranked and labeled genealogies. To adapt the BHV distance to the space of ranked tree shapes, we define a modified BHV metric, $d_{\text{BHV-RTS}}$ as follows:

$$d_{\text{BHV-RTS}}(T_1^R, T_2^R) = d_{\text{BHV}}(\psi(T_1^R), \psi(T_2^R)),$$

where ψ maps a ranked tree shape to its corresponding ranked labeled genealogy by assigning a uniquely defined label to each leaf and assigning a unit length to each change point time interval (u_i, u_{i-1}) .

The unique assignment of the leaf labels ℓ_1, \dots, ℓ_n on a ranked tree shape consists in assigning labels in increasing index order starting with leaves subtending from nodes closer to the bottom and ending with leaves subtending closer to the root (Text 6).

The Kendall-Colijn (KC) metric. The KC metric is another metric on labeled genealogies (2). A labeled genealogy \mathbf{g}_n^L with n leaves is represented by an $\frac{n(n+1)}{2}$ -dimensional vector $v_\lambda(\mathbf{g}_n^L)$ that is a convex combination of two vectors: $(1-\lambda)m(\mathbf{g}_n^L) + \lambda M(\mathbf{g}_n^L)$, $\lambda \in [0, 1]$. $m(\mathbf{g}_n^L)$ is a vector that concatenates n repetitions of one and a vector whose entry corresponds to the number of edges between the most recent common ancestor of a pair of leaves and the root; $M(\mathbf{g}_n^L)$ is a vector that concatenates the vector of leaf branch lengths and the branch length between the most recent common ancestor of a pair of tips and the root.

The KC distance $d_{\text{KC},\lambda}$ with $\lambda > 0$ between two labeled genealogies is the Euclidean norm of the difference between the two KC vector representations of the labeled genealogies. When $\lambda = 0$, the KC distance becomes a distance on the space of labeled unranked topologies: $d_{\text{KC},0}$.

To adapt the KC distance to the space of ranked tree shapes, we propose two distances. We first define a KC-based distance on ranked tree shapes, $d_{\text{KC-RTS}}$, as follows

$$d_{\text{KC-RTS}}(T_1^R, T_2^R) = d_{\text{KC},0}(\eta(T_1^R), \eta(T_2^R)),$$

where η maps a ranked tree shape to a labeled unranked tree shape by removing internal node rankings and uniquely labeling leaves following the procedure described for $d_{\text{BHV-RTS}}$.

The Colijn-Plazzotta (CP) metric. The CP metric is defined on tree shapes (3). The CP metric d_{CP} is defined as the Euclidean norm (L_2 -norm) of the difference between two vectors that uniquely describe the two tree shapes. Each node of a tree is labeled by an integer recursively from tips to the root. The i th entry of the CP vector representing a tree shape records the frequency of the tree nodes labeled with integer i . We define a modified CP distance $d_{\text{CP-RTS}}$ on ranked tree shapes as

$$d_{\text{CP-RTS}}(T_1^R, T_2^R) = d_{\text{CP}}(\phi(T_1^R), \phi(T_2^R)),$$

where ϕ returns the corresponding tree shape of a ranked tree shape by removing the labels of its internal nodes. We note that $d_{\text{CP-RTS}}$ is not a metric but a pseudometric: all pairs of different ranked tree shapes with the same shape will have $d_{\text{CP-RTS}} = 0$. In addition, $d_{\text{CP-RTS}}$ does not account for heterochronous sampling, so we exclude $d_{\text{CP-RTS}}$ from our analyses on heterochronous ranked tree shapes.

B. Other distances on ranked genealogies. We now present the modified BHV and KC distances so that they can be compared to our proposed distances on ranked genealogies.

$$d_{\text{BHV-RG}}(\mathbf{g}_1^R, \mathbf{g}_2^R) = d_{\text{BHV}}(\eta_2(\mathbf{g}_1^R), \eta_2(\mathbf{g}_2^R)),$$

where η_2 maps a ranked genealogy to a labeled ranked genealogy by uniquely labeling leaves as described for $d_{\text{BHV-RTS}}$. Similarly,

$$d_{\text{KC-RG}}(\mathbf{g}_1^R, \mathbf{g}_2^R) = d_{\text{KC},0.5}(\eta_2(\mathbf{g}_1^R), \eta_2(\mathbf{g}_2^R)).$$

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.

6. Unique labeling scheme of ranked tree shapes

In order to adapt other distances defined on labeled trees to ranked tree shapes, we use the following labeling scheme. We start by labeling the leaves that descend directly from the internal node with the largest rank n . If the node has two direct descendent leaves with different edge lengths, we label the longer leaf ℓ_1 and the shorter leaf ℓ_2 . If the two leaves have the same edge lengths, we label them ℓ_1 and ℓ_2 from left to right. If the node has only one direct descendent leaf, we label it ℓ_1 . If there is no descending leaf, no labeling is done. We then move to the node with rank $n - 1$ and continue labeling the leaves by traversing through the internal nodes in descending order of rank until all leaves are labeled. If the current node with rank k has two direct descendent leaves and if the last assigned leaf label is ℓ_j , we label the leaf with the longer edge ℓ_{j+1} and the leaf with shorter edge ℓ_{j+2} ; if the leaves have the same edge lengths, we label the pair of leaves ℓ_{j+1} (left) and ℓ_{j+2} (right). If the node k has only one direct descendent leaf, we label it ℓ_{j+1} . If the node k has no direct descendent leaf, no label is assigned. Examples demonstrating our unique labeling scheme are in Figure S3.

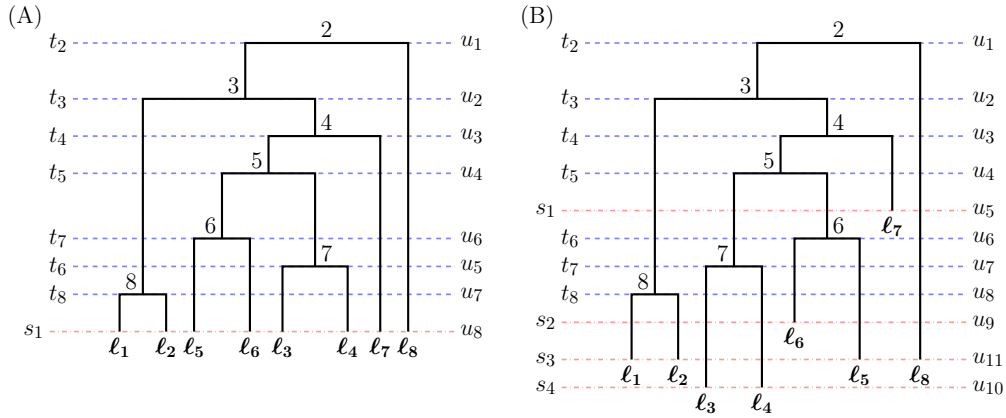


Fig. S3. Unique labeling of ranked tree shapes and ranked genealogies. (A) Example of the unique labeling of a ranked genealogy with isochronous sampling. (B) Example of the unique labeling of a ranked genealogy with heterochronous sampling.

7. Beta-splitting model on labeled tree shapes

We first consider the single-parameter *beta-splitting model* on labeled tree shapes (4). For a parent clade of size n , the model chooses its child clade size to be i on the left branch and $n - i$ on the right branch with probability

$$q_n(i) = \frac{1}{a_n(\beta)} \frac{\Gamma(\beta + i + 1)\Gamma(\beta + n - i + 1)}{\Gamma(i + 1)\Gamma(n - i + 1)},$$

where $a_n(\beta)$ is a normalizing constant and $i \in \{1, 2, \dots, n-1\}$. The splitting is repeated recursively in each branch independently until the tree is fully resolved. The parameter $\beta \in [-2, \infty)$ controls the degree of balance of the generated trees. With $\beta = -2$, the model generates the perfect unbalanced tree (caterpillar tree) with probability one, and with $\beta = \infty$, the model generates the perfect balanced tree with probability one. We note that Ford's alpha-model (5) is another single parameter family of models on the same class of trees as the beta-splitting model and it is not considered in this manuscript.

8. Alpha-Beta splitting model by Maliet et al. (6)

The *alpha-beta model* of Maliet et al. (6) generates labeled ranked tree shapes according to a size-biased distribution with a stick-breaking construction. The algorithm first generates n independent draws u_1, \dots, u_n from a $\text{Unif}(0, 1)$ distribution. The u_i 's correspond to the n leaves of the tree. The first partition of the n leaves (at the root) is determined by drawing a random number $R_1 \sim \text{Beta}(\beta + 1, \beta + 1)$. All the $n_1 = \sum_{i=1}^n 1(u_i < R_1)$ are placed on the left side of the tree and the rest on the other side. Then, if $n_1 > 1$, the left branch is chosen to bifurcate with probability proportional to n_1^α . The algorithm continues

generating beta-distributed values to bi-partition the leaves and chooses the order (ranking) proportional to their number of descendants to the α power until the interval $(0, 1)$ is partitioned into n intervals, each corresponding to an u_i number. The pseudocode is shown below. Leaf labels are then removed to generate a ranked tree shape. The $\beta \in [-2, \infty)$ parameter determines the balance of the tree as in the beta-splitting model, and the $\alpha \in (-\infty, \infty)$ parameter regulates the relationship between subtree family size (number of descendants) and closeness to the root. More specifically, when $\alpha < 0$, subtrees with small family sizes are closer to the root and when $\alpha > 0$, subtrees with small family sizes are closer to the tips. When $\alpha = 1$, the alpha-beta model becomes the *beta-splitting model on ranked tree shapes*.

Algorithm 1 Simulation of a labeled ranked tree shape according to alpha-beta splitting model

- 1: Draw $u_1, \dots, u_n \sim U(0, 1)$
 - 2: Set $i = 1, r_0 = 0$
 - 3: **while** $i < n$ **do**
 - 4: Draw $R_i \sim \text{Beta}(1 + \beta, 1 + \beta)$
 - 5: Let r_1, \dots, r_i be the ordered permutation of R_1, \dots, R_i such that $r_1 < r_2 < \dots < r_i$
 - 6: Let $y_j = \begin{cases} 1 & \text{if } \sum_{k=1}^n \mathbb{1}(u_k \in (r_{j-1}, r_j)) > 1 \\ 0 & \text{o.w} \end{cases}$, for $j = 1, \dots, i$
 - 7: **if** $\sum_{j=1}^i y_j > 1$ **then**
 - 8: Pick y_k w.p. $\frac{(r_k - r_{k-1})^{\alpha y_k}}{\sum_{j=1}^i (r_j - r_{j-1})^{\alpha y_j}}$. The partition defined by the u_j 's in (r_{k-1}, r_k) is chosen to bifurcate with ranking i .
 - 9: $i = i + 1$
-

9. Tajima coalescent on ranked genealogies

The Tajima coalescent is a model on ranked genealogies (Figure 1(A)). It is a Markov lumping of Kingman's n -coalescent on labeled and ranked genealogies (7, 8). The Tajima coalescent is a pure death process that starts with n unlabeled leaves at time $t_{n+1} = 0$ and proceeds backward in time, merging pairs of branches to create interior nodes labeled by their order of appearance. The merging of two branches is a coalescent event. In the Tajima coalescent, the distribution of coalescence times is the same as in the Kingman coalescent and the probability of a topology, ranked tree shape, is given by

$$P(T^R) = \frac{2^{n-c-1}}{(n-1)!}, \quad [S2]$$

where n is the number of leaves, and c is the number of cherries—the number of pairs of leaves that subtend from a shared interior node. The Tajima coalescent on ranked tree shapes without times corresponds to the beta-splitting model on ranked tree shapes with $\beta = 0$, also called the Yule model. A full description of the Tajima coalescent process can be found in Cappello and Palacios (9).

10. Distributions on branching or coalescent times of heterochronous genealogies

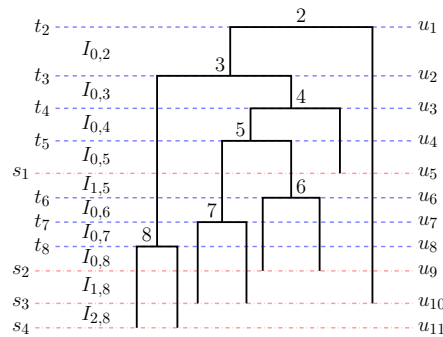


Fig. S4. Heterochronous genealogy. Example of a ranked genealogy with heterochronous sampling. s_4, \dots, s_1 and t_8, \dots, t_2 indicate sampling times (red dotted lines) and coalescent times (blue dotted lines), respectively. u_{11}, \dots, u_1 are the ordered times of change points where the number of lineages changes due to either a sampling event or a coalescent event. The time increases backwards in time starting with $u_{11} = 0$ as the present time. $I_{0,k}$ is the interval that ends with a coalescent event at t_k . $I_{i,k}$ ($i > 0$) represents the interval that ends with a sampling time within the interval (t_{k+1}, t_k) . For $k = n$, we adopt the convention $t_{n+1} = 0$.

The ranked tree shape and ranked genealogy of time-stamped samples are termed *heterochronous ranked tree shapes* and *heterochronous genealogies*, respectively (Figure S4). Here, we assume that samples are collected at times s_m, s_{m-1}, \dots, s_1 , with $s_m = t_{n+1} = 0$ (the present), and $s_j < s_{j-1}$ for $j = m, \dots, 2$. At time s_j , n_j lineages are sampled, and $\sum_{j=1}^m n_j = n$.

The $\lambda(t)$ -heterochronous-coalescent (10) describes the distribution of coalescent times conditioned on collecting n_m, n_{m-1}, \dots, n_1 samples at times s_m, s_{m-1}, \dots, s_1 . As before, t_{n+1}, t_n, \dots, t_2 denote the coalescent times, except that the subindex no longer indicates the number of lineages. Instead, the subindex indicates the rank order of the coalescent events going forward in time (tipward) from the root at t_2 . Define $(u_{n+m-1}, u_{n+m-2}, \dots, u_1)$ as the vector of change points (coalescent or sampling times), with $0 = u_{n+m-1} < u_{n+m-2} < \dots < u_1 = t_2$. To state the density of coalescent times according to the $\lambda(t)$ -heterochronous-coalescent, going backwards in time (rootward), we denote the intervals that end with a coalescent event at t_k by $I_{0,k}$, and the intervals that end with a sampling time within the interval (t_{k+1}, t_k) as $I_{i,k}$, where i is an index tracking the sampling events in (t_{k+1}, t_k) . More specifically,

$$\begin{aligned} I_{0,k} &= (\max\{t_{k+1}, s_j\}, t_k] \text{ for } s_j < t_k \\ I_{i,k} &= (\max\{t_{k+1}, s_{j+i}\}, s_{j+i-1}] \text{ for } s_{j+i-1} > t_{k+1} \text{ and } s_j < t_k, \end{aligned}$$

with $k = 2, \dots, n$ and i ranges from 1 to the number of sampling events in (t_{k+1}, t_k) . An example of the annotated time intervals using $I_{0,k}$ and $I_{i,k}$ is shown in Figure S4.

The conditional density of t_{k-1} is the product of the conditional density of $t_{k-1} \in I_{0,k}$ and the probability of not having a coalescent event during the period of time spanned by intervals $I_{1,k}, \dots, I_{m,k}$. That is, for $k = 2, \dots, n$,

$$\begin{aligned} P[t_{k-1}|t_k, \mathbf{s}, \mathbf{n}, N_e(t)] &= \frac{C_{0,k-1}}{N_e(t_{k-1})} \\ &\times \exp \left[- \left\{ \int_{I_{0,k-1}} \frac{C_{0,k-1} dt}{N_e(t)} + \sum_{i=1}^m \int_{I_{i,k-1}} \frac{C_{i,k-1} dt}{N_e(t)} \right\} \right], \end{aligned} \quad [S3]$$

where $C_{i,k} = \binom{n_{i,k}}{2}$.

11. Embeddings in MDS

We chose MDS in two dimensions to visualize matrices of pairwise distances. In general, our metrics are well explained in the MDS visualization; however, the other distances are usually poorly represented in this space. In this section, we propose a measure of distortion and correlation to assess how well the embedding preserves the pairwise distances for each metric. In our examples, the distortion measure shown in Table S15 suggests that our d_2 metric has the best MDS embedding in general of all distance functions considered with our d_1 metric a close second. Similarly, the correlation measure shown in Table S16 confirms that our d_1 and d_2 metrics, with near perfect correlations, have far better embedding in the 2-dimensional MDS space than the other distances considered.

A. Distortion. To assess our distances and their MDS embeddings in two dimensions, we compute the following distortion statistic (11) defined as follows:

$$\text{distortion} = \text{expansion} \times \text{contraction},$$

where expansion and contraction are defined as follows. For a given sample of ranked tree shapes $\mathcal{T}_S = \{T_1^R, T_2^R, \dots, T_s^R\}$ with n leaves in \mathcal{T}_n^R ,

$$\text{expansion} = \max_{\substack{T_i^R, T_j^R \in \mathcal{T}_S; \\ i \neq j}} \frac{d_{\text{MDS}}(T_i^R, T_j^R)}{d(T_i^R, T_j^R)},$$

where d_{MDS} is the L_2 -Euclidean distance in the reduced MDS space and d is any distance function on ranked tree shapes, and

$$\text{contraction} = \max_{\substack{T_i^R, T_j^R \in \mathcal{T}_S; \\ i \neq j}} \frac{d(T_i^R, T_j^R)}{d_{\text{MDS}}(T_i^R, T_j^R)}.$$

The distortion on the ranked genealogies is defined similarly. The comparison of distortions for simulated ranked tree shapes and ranked genealogies can be found in Table S15.

B. Correlation. As a second measure for assessing our distances and their MDS embeddings in two dimensions, we compute the Pearson correlation coefficient between the two vectors of pairwise distances between sampled ranked tree shapes, one from using any distance functions d on ranked tree shapes and the other from the L_2 -Euclidean distance d_{MDS} in the reduced MDS space:

$$\text{correlation} = \frac{\sum_{i=2}^s \sum_{j=1}^i (d(T_i^R, T_j^R) - \mu_d) (d_{\text{MDS}}(T_i^R, T_j^R) - \mu_{d_{\text{MDS}}})}{\sqrt{\sum_{i=2}^s \sum_{j=1}^i (d(T_i^R, T_j^R) - \mu_d)^2 \sum_{i=2}^s \sum_{j=1}^i (d_{\text{MDS}}(T_i^R, T_j^R) - \mu_{d_{\text{MDS}}})^2}},$$

where s is the number of sampled ranked tree shapes. $\mu_{d_{\text{MDS}}}$ and μ_d are the mean of the pairwise distances using L_2 -Euclidean distance in the MDS space and using any distance functions d on the sampled ranked tree shapes, respectively. The comparisons of correlations for simulated ranked tree shapes and ranked genealogies appear in Table S16.

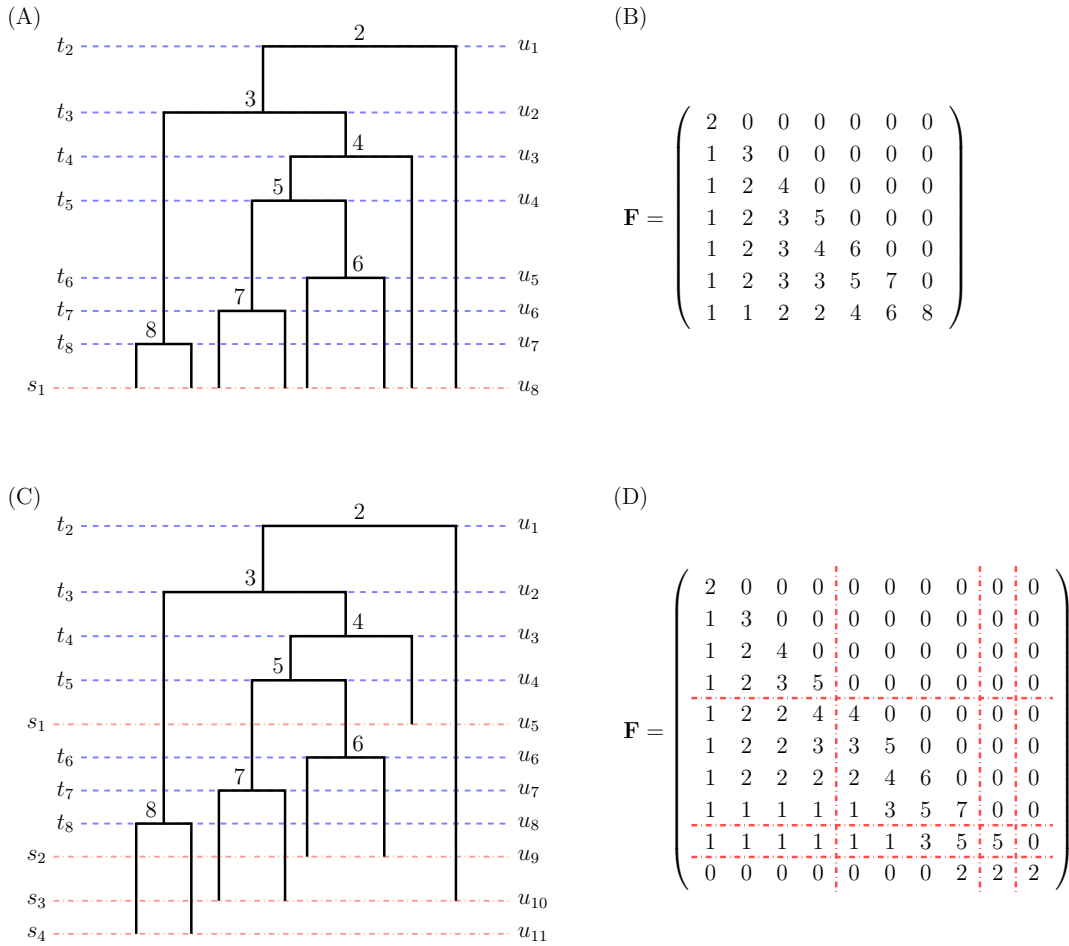


Fig. S5. Bijection of ranked tree shapes and \mathbf{F} -matrices for isochronous and heterochronous trees. (A) Example of a ranked genealogy with isochronous sampling. (B) The corresponding \mathbf{F} -matrix that encodes the ranked tree shape information of the tree in (A). (C) Example of a ranked genealogy with heterochronous sampling. (D) The corresponding \mathbf{F} -matrix of the heterochronous ranked tree shape in (C). Blue dotted lines indicate coalescent events and red dotted lines represent sampling events. In (C), coalescent times are denoted by $\{t_k\}_{k=2}^8$, sampling times by $\{s_k\}_{k=1}^4$, and the number of lineages changes at change points $\{u_k\}_{k=1}^{11}$.

$$\begin{pmatrix}
u_1 - u_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
u_1 - u_3 & u_2 - u_3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
u_1 - u_4 & u_2 - u_4 & u_3 - u_4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
u_1 - u_5 & u_2 - u_5 & u_3 - u_5 & u_4 - u_5 & 0 & 0 & 0 & 0 & 0 & 0 \\
u_1 - u_6 & u_2 - u_6 & u_3 - u_6 & u_4 - u_6 & u_5 - u_6 & 0 & 0 & 0 & 0 & 0 \\
u_1 - u_7 & u_2 - u_7 & u_3 - u_7 & u_4 - u_7 & u_5 - u_7 & u_6 - u_7 & 0 & 0 & 0 & 0 \\
u_1 - u_8 & u_2 - u_8 & u_3 - u_8 & u_4 - u_8 & u_5 - u_8 & u_6 - u_8 & u_7 - u_8 & 0 & 0 & 0 \\
u_1 - u_9 & u_2 - u_9 & u_3 - u_9 & u_4 - u_9 & u_5 - u_9 & u_6 - u_9 & u_7 - u_9 & u_8 - u_9 & 0 & 0 \\
u_1 - u_{10} & u_2 - u_{10} & u_3 - u_{10} & u_4 - u_{10} & u_5 - u_{10} & u_6 - u_{10} & u_7 - u_{10} & u_8 - u_{10} & u_9 - u_{10} & 0 \\
u_1 & u_2 & u_3 & u_4 & u_5 & u_6 & u_7 & u_8 & u_9 & u_{10}
\end{pmatrix}$$

Fig. S6. Example of the weight matrix \mathbf{W} . The weight matrix associated with the example heterochronous ranked genealogy and its \mathbf{F} -matrix in Figures S5(C) and (D). In the last row, u_{11} is suppressed because we set the initial sampling time to be $u_{11} = 0$.

(A) \mathbf{d}_1 (B) \mathbf{d}_2

	T_1^R	T_2^R	T_3^R	T_4^R	T_5^R
T_1^R	0	2	3	2	1
T_2^R		0	1	2	3
T_3^R			0	1	2
T_4^R				0	1
T_5^R					0

	T_1^R	T_2^R	T_3^R	T_4^R	T_5^R
T_1^R	0	1.41	1.73	1.41	1
T_2^R		0	1	1.41	1.73
T_3^R			0	1	1.41
T_4^R				0	1
T_5^R					0

(C) $\mathbf{d}_{\text{BHV-RTS}}$ (D) $\mathbf{d}_{\text{KC-RTS}}$ (E) $\mathbf{d}_{\text{CP-RTS}}$

	T_1^R	T_2^R	T_3^R	T_4^R	T_5^R
T_1^R	0	2.45	3.74	3.46	2.45
T_2^R		0	3.46	3.74	3.74
T_3^R			0	2.45	2.45
T_4^R				0	2.45
T_5^R					0

	T_1^R	T_2^R	T_3^R	T_4^R	T_5^R
T_1^R	0	2.65	3.61	4.12	2
T_2^R		0	2.45	2.83	2.65
T_3^R			0	2.45	2.65
T_4^R				0	2.65
T_5^R					0

	T_1^R	T_2^R	T_3^R	T_4^R	T_5^R
T_1^R	0	2	2	2	2.45
T_2^R		0	0	0	2
T_3^R			0	0	2
T_4^R				0	2
T_5^R					0

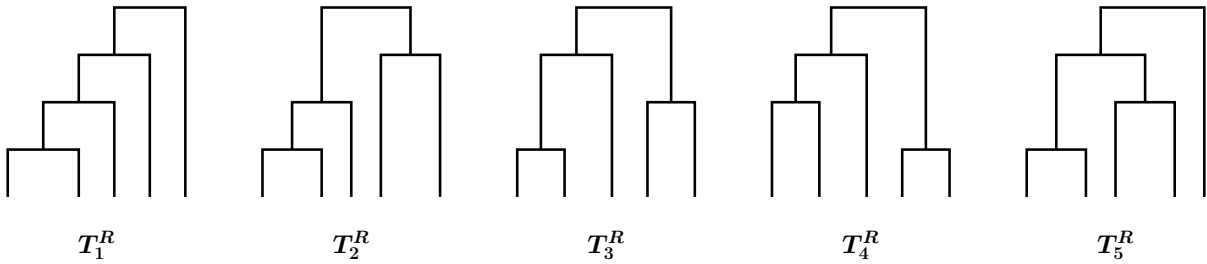


Fig. S7. Comparisons of metrics applied to isochronous ranked tree shapes with $n = 5$.

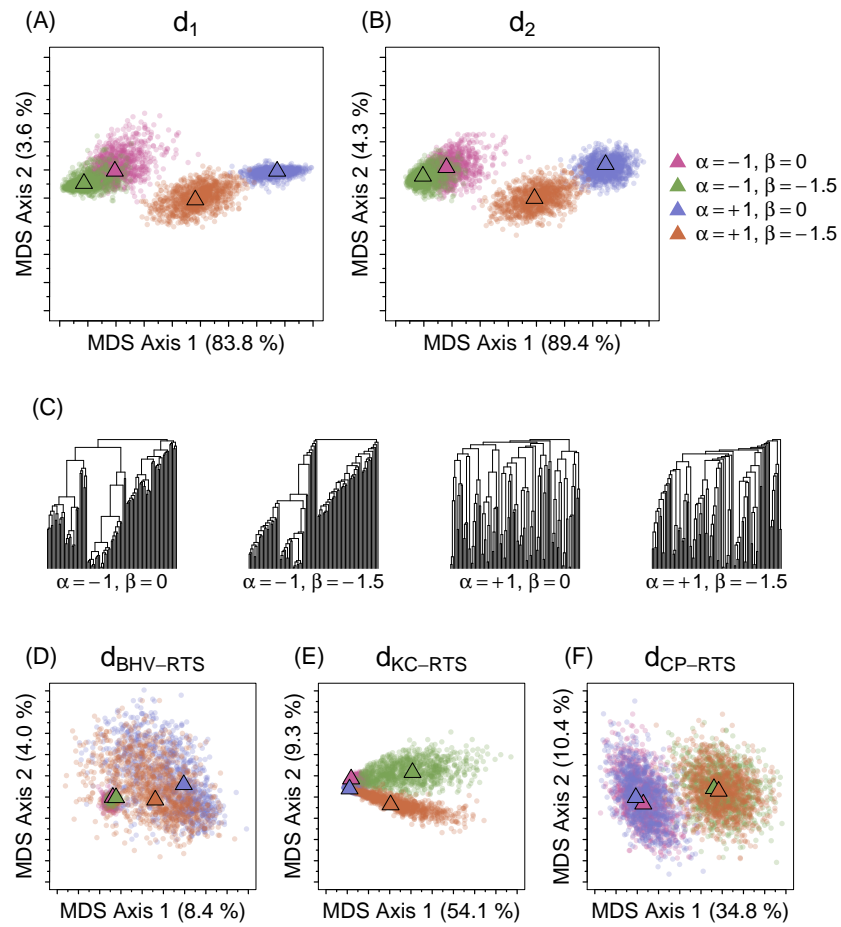


Fig. S8. MDS representation of distances between 4,000 simulated isochronous ranked tree shapes of $n = 100$ leaves, aggregated from four different alpha-beta splitting models. 1,000 isochronous ranked tree shapes were simulated for each of pairs of (α, β) values in $\{(-1, 0), (-1, -1.5), (+1, 0), (+1, -1.5)\}$. Different α generates different distributions of internal node ranking while different β generates different distributions of tree balance. (A) MDS of the d_1 metric. (B) MDS of the d_2 metric. (C) L_2 -medoid trees from each distribution using the d_1 metric. MDS plots for (D) $d_{\text{BHV-RTS}}$, (E) $d_{\text{KC-RTS}}$, and (F) $d_{\text{CP-RTS}}$. In each MDS plot, the triangle represents the L_2 -medoid tree of 1,000 points for a specified model.

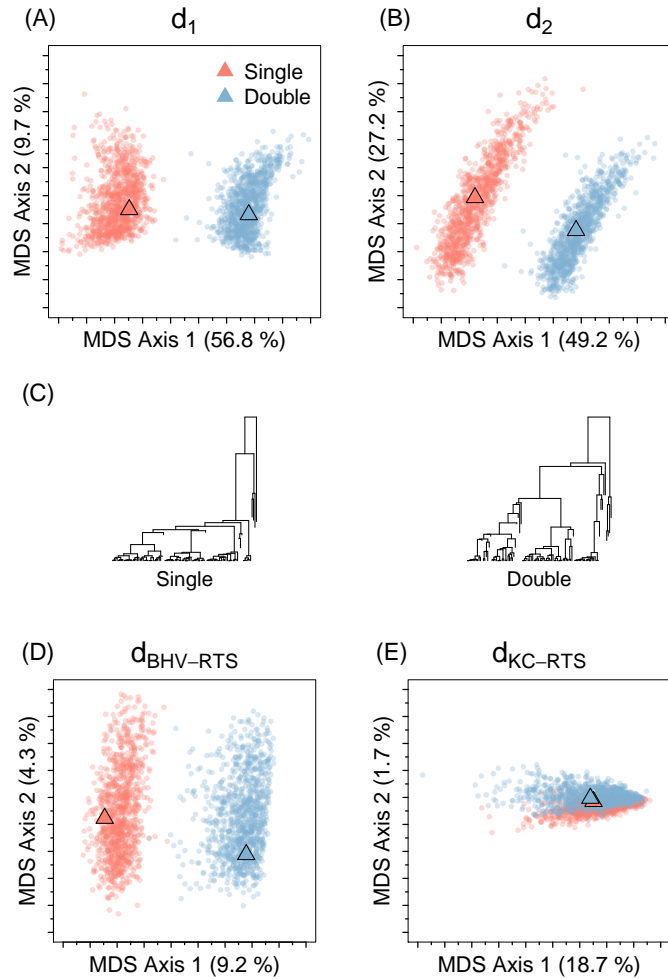


Fig. S9. MDS representation of distances between 2,000 simulated heterochronous ranked tree shapes of $n = 100$ with different sampling events. (A) MDS of the d_1 metric. (B) MDS of the d_2 metric. (C) L_2 -medoid trees from each distribution using the d_1 metric. MDS plots for (D) $d_{\text{BHV-RTS}}$ and (E) $d_{\text{KC-RTS}}$. To demonstrate that our metrics are sensitive to sampling schedules, we simulated trees with $n = 100$ leaves under heterochronous sampling with two sampling scenarios. For the first set of trees (the “Single” distribution), we selected 90 samples at time 0 and the remaining 10 samples at distinct times with sampling times drawn uniformly at random from $(0, 10^4]$, i.e., $\mathbf{n} = (90, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1)$. In the second set of trees (the “Double” distribution), 80 samples were drawn at time 0 and the remaining 20 were sampled in pairs at ten distinct random sampling times drawn uniformly from $(0, 10^4]$, i.e., $\mathbf{n} = (80, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2)$. We generated 1,000 coalescent trees per sampling scheme assuming a constant effective population size trajectory of $N_0 = 10^4$. We then removed leaf labels to produce the $2,000 \times 2,000$ distance matrices with all applicable distances. Our metrics and $d_{\text{BHV-RTS}}$ show a clear separation between the two distributions along the first two MDS axes. The total distance explained in the two-dimensional space is higher using our metrics, 66.5% and 76.4% for d_1 and d_2 , respectively, compared to 13.5% of $d_{\text{BHV-RTS}}$. $d_{\text{KC-RTS}}$ exhibits less discrimination than the other three distances. The confusion matrices in Table S7 and the confusion statistics in Table S8 confirm that our metrics distinguish different sampling schemes better than the other two metrics compared.

Table S1. Summary of dispersions for ranked tree shapes. Comparison of dispersion (Eq. 6) of ranked tree shape distribution using distance functions between ranked tree shapes. (A) Isochronous ranked tree shapes simulated from the beta-splitting model with varying β parameters (Figure 4). (B) Isochronous ranked tree shapes simulated from the alpha-beta splitting model with varying α parameters and fixed $\beta = 0$ (Figure 5). (C) Heterochronous ranked tree shapes simulated from different sampling schemes (Figure S9). Note that d_{CP-RTS} does not account for heterochronous sampling, so we exclude d_{CP-RTS} from our analyses on heterochronous ranked tree shapes (Text 5).

(A) Isochronous ranked tree shapes, beta-splitting model

	d_1	d_2	$d_{BHV-RTS}$	d_{KC-RTS}	d_{CP-RTS}
Balanced	5541.21	117.14	377.84	128.20	8.51
Yule	5889.31	119.40	363.33	182.52	8.97
AB	6579.90	127.37	345.82	271.78	9.58
PDA	6966.50	133.24	304.49	490.98	10.48
Unbalanced	4391.36	90.88	154.44	944.90	12.50

(B) Isochronous ranked tree shapes, alpha-beta splitting model

	d_1	d_2	$d_{BHV-RTS}$	d_{KC-RTS}	d_{CP-RTS}
$\alpha = -2$	4701.27	90.81	85.33	110.76	8.95
$\alpha = -1$	7202.89	135.70	110.64	126.06	9.00
$\alpha = 0$	7719.71	150.57	262.18	172.73	8.96
$\alpha = 1$	6084.73	122.77	363.17	182.49	9.06
$\alpha = 2$	5668.18	115.80	376.98	179.18	9.08

(C) Heterochronous ranked tree shapes, different sampling schemes

	d_1	d_2	$d_{BHV-RTS}$	d_{KC-RTS}	d_{CP-RTS}
Single	10359.43	277.88	350.26	230.17	—
Double	10423.86	238.90	324.47	257.65	—

Table S2. Comparison of metrics: discrimination of isochronous ranked tree shapes under different beta-splitting models. We compare the performance of different distances on ranked tree shapes according to how well they separate trees simulated from the beta-splitting distribution of ranked tree shapes with different balance parameters β . Rows indicate the sampling distribution and columns indicate the L_2 -medoid of each distribution. Each matrix corresponds to a different distance metric. Entry (i, j) corresponds to the percentage of trees simulated from the i th distribution that are closer to the j th L_2 -medoid than to the medoids of any other columns. The color scheme of the L_2 -medoids follows Figure 4. The mean diagonal values are 83.28, 82.02, 20.32, 70.50, and 74.96 for matrices (A)-(E), respectively.

	(A) d_1	(B) d_2	
			L_2 -medoid
Balanced	$\begin{bmatrix} 85.9 & 14.0 & 0.1 & 0 & 0 \\ 22.9 & 64.4 & 12.7 & 0 & 0 \\ 1.1 & 19.7 & 73.6 & 5.6 & 0 \\ 0 & 0 & 7.2 & 92.5 & 0.3 \\ 0 & 0 & 0 & 0 & 100.0 \end{bmatrix}$	$\begin{bmatrix} 82.6 & 17.3 & 0.1 & 0 & 0 \\ 23.1 & 62.8 & 14.1 & 0 & 0 \\ 1.4 & 19.7 & 74.2 & 4.7 & 0 \\ 0 & 0 & 9.2 & 90.6 & 0.2 \\ 0 & 0 & 0 & 0.1 & 99.9 \end{bmatrix}$	Balanced Yule AB PDA Unbalanced
Yule			
AB			
PDA			
Unbalanced			
	(C) $d_{\text{BHV-RTS}}$	(D) $d_{\text{KC-RTS}}$	(E) $d_{\text{CP-RTS}}$
Balanced	$\begin{bmatrix} 1.0 & 0 & 0 & 0.2 & 98.8 \\ 0.5 & 0.2 & 0 & 0.2 & 99.1 \\ 0.5 & 0.1 & 0.2 & 0.2 & 99.0 \\ 0.2 & 0 & 0 & 0.3 & 99.5 \\ 0 & 0 & 0 & 0.1 & 99.9 \end{bmatrix}$	$\begin{bmatrix} 100.0 & 0 & 0 & 0 & 0 \\ 56.3 & 36.0 & 7.7 & 0 & 0 \\ 7.9 & 27.3 & 55.2 & 9.6 & 0 \\ 0 & 2.5 & 20.9 & 76.3 & 0.3 \\ 0 & 0 & 0.2 & 14.8 & 85.0 \end{bmatrix}$	$\begin{bmatrix} 77.1 & 22.5 & 0.4 & 0 & 0 \\ 29.8 & 51.3 & 18.7 & 0.2 & 0 \\ 4.8 & 26.4 & 57.7 & 11.1 & 0 \\ 0 & 0.5 & 10.7 & 88.8 & 0 \\ 0 & 0 & 0 & 0.1 & 99.9 \end{bmatrix}$
Yule			
AB			
PDA			
Unbalanced			

Table S3. Summary of the pairwise mean confusion statistics and associated P values for isochronous ranked tree shapes under different beta-splitting models. Each entry in the table contains a pair of values: the mean confusion statistic of two ranked tree shape sampling distributions (Eq. 8) and its associated P value (Eq. 9). The simulated distributions of ranked tree shapes for the off-diagonal entries are the same as those considered in Figure 4 and Table S2. The diagonal entries represent the tests of true null cases, where we generated additional 1,000 random ranked tree shapes per each β -value and computed the mean confusion statistic between two samples drawn from the same distribution.

(A) d_1

	Balanced	Yule	AB	PDA	Unbalanced
Balanced	0.4935, 0.4628				
Yule	0.1850, 0.0001	0.4990, 0.5088			
AB	0.0360, 0.0001	0.1675, 0.0001	0.5025, 0.6396		
PDA	0.0005, 0.0001	0.0110, 0.0001	0.0640, 0.0001	0.5105, 0.8766	
Unbalanced	0.0000, 0.0001	0.0000, 0.0001	0.0000, 0.0001	0.0015, 0.0001	0.5245, 0.9919

(B) d_2

	Balanced	Yule	AB	PDA	Unbalanced
Balanced	0.5040, 0.7583				
Yule	0.2025, 0.0001	0.5015, 0.6240			
AB	0.0405, 0.0001	0.1755, 0.0001	0.4960, 0.3915		
PDA	0.0015, 0.0001	0.0125, 0.0001	0.0695, 0.0001	0.5020, 0.6494	
Unbalanced	0.0000, 0.0001	0.0000, 0.0001	0.0000, 0.0001	0.0015, 0.0001	0.4720, 0.0087

(C) $d_{\text{BHV-RTS}}$

	Balanced	Yule	AB	PDA	Unbalanced
Balanced	0.4900, 0.2052				
Yule	0.4795, 0.0452	0.4760, 0.0194			
AB	0.4570, 0.0001	0.4685, 0.0028	0.4955, 0.4076		
PDA	0.4695, 0.0033	0.4700, 0.0041	0.4825, 0.0639	0.5150, 0.9245	
Unbalanced	0.4950, 0.3609	0.4990, 0.4886	0.4990, 0.4864	0.4990, 0.5025	0.5085, 0.8043








(D) $d_{\text{KC-RTS}}$

	Balanced	Yule	AB	PDA	Unbalanced
Balanced	0.5015, 0.5973				
Yule	0.2815, 0.0001	0.4940, 0.3448			
AB	0.0945, 0.0001	0.2145, 0.0001	0.4880, 0.1071		
PDA	0.0675, 0.0001	0.0900, 0.0001	0.1650, 0.0001	0.4760, 0.0126	
Unbalanced	0.0400, 0.0001	0.0460, 0.0001	0.0535, 0.0001	0.0765, 0.0001	0.5115, 0.8692

(E) $d_{\text{CP-RTS}}$

	Balanced	Yule	AB	PDA	Unbalanced
Balanced	0.4935, 0.3205				
Yule	0.2590, 0.0001	0.5150, 0.9263			
AB	0.0755, 0.0001	0.2555, 0.0001	0.5015, 0.6544		
PDA	0.0035, 0.0001	0.0260, 0.0001	0.1110, 0.0001	0.5035, 0.7410	
Unbalanced	0.0000, 0.0001	0.0000, 0.0001	0.0000, 0.0001	0.0005, 0.0001	0.4750, 0.0274

Table S4. Comparison of metrics: discrimination of isochronous ranked tree shapes under different alpha-beta splitting models with a fixed beta value. We compare the performance of different distances on ranked tree shapes according to how well they separate trees simulated from the alpha-beta splitting distribution of ranked tree shapes with different parameter values α which regulates the internal node ranking of a given tree shape. The format of the matrices follows Table S2. The simulation values and the color scheme of the L_2 -medoids follow Figure 5. The mean diagonal values are 75.40, 76.94, 20.14, 36.48, and 19.56 for matrices (A)-(E), respectively.

	(A) d_1	(B) d_2		L_2 -medoid								
												
$\alpha = -2$	95.6	4.4	0	0	0	92.5	7.5	0	0	0		$\alpha = -2$
$\alpha = -1$	30.4	69.0	0.6	0	0	22.9	76.6	0.5	0	0		$\alpha = -1$
$\alpha = 0$	0	0.6	86.0	12.8	0.6	0	0.4	88.0	11.2	0.4		$\alpha = 0$
$\alpha = 1$	0	0	4.9	61.4	33.7	0	0	4.3	62.3	33.4		$\alpha = 1$
$\alpha = 2$	0	0	0	35.0	65.0	0	0	0	34.7	65.3		$\alpha = 2$




	(C) $d_{\text{BHV-RTS}}$	(D) $d_{\text{KC-RTS}}$	(E) $d_{\text{CP-RTS}}$												
															
$\alpha = -2$	97.5	2.5	0	0	0	71.5	28.5	0	0	0	29.6	13.1	17.0	14.4	25.9
$\alpha = -1$	97.1	2.9	0	0	0	71.4	28.5	0	0	0.1	31.4	13.5	15.3	11.2	28.6
$\alpha = 0$	88.2	11.7	0.1	0	0	34.2	34.1	17.3	1.5	12.9	29.9	12.3	15.3	14.3	28.2
$\alpha = 1$	80.1	18.2	1.6	0.1	0	1.3	2.9	35.7	19.6	40.5	28.8	11.1	17.3	11.1	31.7
$\alpha = 2$	80.2	18.4	1.2	0.1	0.1	1.5	3.8	32.2	17.0	45.5	28.2	13.9	16.9	12.7	28.3

Table S5. Summary of the pairwise mean confusion statistics and associated P values for isochronous ranked tree shapes under different alpha-beta splitting models. The format of the table follows Table S3. The simulated distributions of ranked tree shapes for the off-diagonal entries are the same as those considered in Figure 5 and Table S4. The diagonal entries represent the tests of true null cases, where we generated additional 1,000 random ranked tree shapes per each α -value.

(A) d_1

	$\alpha = -2$	$\alpha = -1$	$\alpha = 0$	$\alpha = 1$	$\alpha = 2$
$\alpha = -2$	0.4845, 0.1206				
$\alpha = -1$	0.1740, 0.0001	0.5020, 0.6866			
$\alpha = 0$	0.0000, 0.0001	0.0060, 0.0001	0.5175, 0.9541		
$\alpha = 1$	0.0000, 0.0001	0.0005, 0.0001	0.0915, 0.0001	0.4970, 0.4946	
$\alpha = 2$	0.0000, 0.0001	0.0000, 0.0001	0.0505, 0.0001	0.3430, 0.0001	0.4900, 0.2482

(B) d_2

	$\alpha = -2$	$\alpha = -1$	$\alpha = 0$	$\alpha = 1$	$\alpha = 2$
$\alpha = -2$	0.4970, 0.5093				
$\alpha = -1$	0.1520, 0.0001	0.4990, 0.5878			
$\alpha = 0$	0.0000, 0.0001	0.0045, 0.0001	0.5120, 0.8818		
$\alpha = 1$	0.0000, 0.0001	0.0000, 0.0001	0.0795, 0.0001	0.5035, 0.7277	
$\alpha = 2$	0.0000, 0.0001	0.0000, 0.0001	0.0415, 0.0001	0.3400, 0.0001	0.4805, 0.0777

(C) $d_{\text{BHV-RTS}}$

	$\alpha = -2$	$\alpha = -1$	$\alpha = 0$	$\alpha = 1$	$\alpha = 2$
$\alpha = -2$	0.4930, 0.2680				
$\alpha = -1$	0.4980, 0.4308	0.5075, 0.7598			
$\alpha = 0$	0.4995, 0.5020	0.4990, 0.4879	0.5160, 0.9350		
$\alpha = 1$	0.4995, 0.4940	0.4995, 0.5025	0.4975, 0.4313	0.4890, 0.1853	
$\alpha = 2$	0.4995, 0.5217	0.4995, 0.5303	0.4950, 0.3698	0.4675, 0.0019	0.4965, 0.4360

(D) $d_{\text{KC-RTS}}$

	$\alpha = -2$	$\alpha = -1$	$\alpha = 0$	$\alpha = 1$	$\alpha = 2$
$\alpha = -2$	0.5060, 0.7526				
$\alpha = -1$	0.5000, 0.5893	0.5070, 0.7606			
$\alpha = 0$	0.3395, 0.0001	0.3350, 0.0001	0.4940, 0.3101		
$\alpha = 1$	0.1420, 0.0001	0.1210, 0.0001	0.4065, 0.0001	0.4895, 0.2360	
$\alpha = 2$	0.0610, 0.0001	0.0780, 0.0001	0.3425, 0.0001	0.4885, 0.0001	0.4860, 0.1200

(E) $d_{\text{CP-RTS}}$

	$\alpha = -2$	$\alpha = -1$	$\alpha = 0$	$\alpha = 1$	$\alpha = 2$
$\alpha = -2$	0.5090, 0.8159				
$\alpha = -1$	0.4920, 0.2788	0.5020, 0.6775			
$\alpha = 0$	0.4890, 0.1934	0.4970, 0.4219	0.5080, 0.7847		
$\alpha = 1$	0.4925, 0.2898	0.4865, 0.1333	0.5035, 0.6906	0.4870, 0.1825	
$\alpha = 2$	0.4895, 0.1984	0.4975, 0.4956	0.4990, 0.5067	0.5120, 0.9041	0.5105, 0.8784

Table S6. Comparison of metrics: discrimination of isochronous ranked tree shapes under different alpha-beta splitting models. We compare the performance of different distances on ranked tree shapes according to how well they separate trees simulated from the alpha-beta splitting distribution of ranked tree shapes with different parameter values α and β . The format of the matrices follows Table S2. The simulation values and the color scheme of the L_2 -medoids follow Figure S8. The respective mean separation indices (mean diagonal) are 89.8, 89.8, 25.4, 78.8, and 49.5 for d_1 , d_2 , $d_{\text{BHV-RTS}}$, $d_{\text{KC-RTS}}$, and $d_{\text{CP-RTS}}$.

(A)	d_1	(B)	d_2	L_2 -medoid	
$\alpha = -1, \beta = 0$	75.6	23.3	0.0	1.1	$\alpha = -1, \beta = 0$
$\alpha = -1, \beta = -1.5$	13.2	86.8	0.0	0.0	$\alpha = -1, \beta = -1.5$
$\alpha = +1, \beta = 0$	0.0	0.0	99.8	0.2	$\alpha = +1, \beta = 0$
$\alpha = +1, \beta = -1.5$	0.4	0.0	2.7	96.9	$\alpha = +1, \beta = -1.5$

(C)	$d_{\text{BHV-RTS}}$	(D)	$d_{\text{KC-RTS}}$	(E)	$d_{\text{CP-RTS}}$
$\alpha = -1, \beta = 0$	0.1	99.9	0.0	0.0	56.7
$\alpha = -1, \beta = -1.5$	0.0	100.0	0.0	0.0	1.9
$\alpha = +1, \beta = 0$	0.1	99.3	0.2	0.4	54.2
$\alpha = +1, \beta = -1.5$	0.1	98.6	0.0	1.3	2.2

Table S7. Comparison of metrics: heterochronous ranked tree shapes with different sampling schemes. We compare the performance of different distances on ranked tree shapes according to how well they separate trees simulated from the heterochronous Tajima coalescent with different sampling sequences s and n . The format of the matrices follows Table S2. The simulation values and the color scheme of the L_2 -medoids follow Figure S9. The mean diagonal values are 99.85, 99.90, 95.60, and 56.60 for matrices (A)-(D), respectively.

<p>(A) d_1</p> <div style="display: flex; justify-content: center; gap: 20px; margin-bottom: 5px;"> ▲ ▲ </div> <table style="border-collapse: collapse;"> <tr> <td style="padding-right: 10px;">Single</td> <td style="border-left: 1px solid black; border-right: 1px solid black; padding: 5px 10px;">100.0</td> <td style="border-right: 1px solid black; padding: 5px 10px;">0</td> </tr> <tr> <td>Double</td> <td style="border-left: 1px solid black; padding: 5px 10px;">0.3</td> <td style="padding: 5px 10px;">99.7</td> </tr> </table>	Single	100.0	0	Double	0.3	99.7	<p>(B) d_2</p> <div style="display: flex; justify-content: center; gap: 20px; margin-bottom: 5px;"> ▲ ▲ </div> <table style="border-collapse: collapse;"> <tr> <td style="padding-right: 10px;">Single</td> <td style="border-left: 1px solid black; border-right: 1px solid black; padding: 5px 10px;">99.9</td> <td style="border-right: 1px solid black; padding: 5px 10px;">0.1</td> </tr> <tr> <td>Double</td> <td style="border-left: 1px solid black; padding: 5px 10px;">0.1</td> <td style="padding: 5px 10px;">99.9</td> </tr> </table>	Single	99.9	0.1	Double	0.1	99.9	<p>L_2-medoid</p> <p style="margin-top: 5px;">▲ Single</p> <p style="margin-top: 5px;">▲ Double</p>
Single	100.0	0												
Double	0.3	99.7												
Single	99.9	0.1												
Double	0.1	99.9												
<p>(C) $d_{\text{BHV-RTS}}$</p> <div style="display: flex; justify-content: center; gap: 20px; margin-bottom: 5px;"> ▲ ▲ </div> <table style="border-collapse: collapse;"> <tr> <td style="padding-right: 10px;">Single</td> <td style="border-left: 1px solid black; border-right: 1px solid black; padding: 5px 10px;">92.1</td> <td style="border-right: 1px solid black; padding: 5px 10px;">7.9</td> </tr> <tr> <td>Double</td> <td style="border-left: 1px solid black; padding: 5px 10px;">0.9</td> <td style="padding: 5px 10px;">99.1</td> </tr> </table>	Single	92.1	7.9	Double	0.9	99.1	<p>(D) $d_{\text{KC-RTS}}$</p> <div style="display: flex; justify-content: center; gap: 20px; margin-bottom: 5px;"> ▲ ▲ </div> <table style="border-collapse: collapse;"> <tr> <td style="padding-right: 10px;">Single</td> <td style="border-left: 1px solid black; border-right: 1px solid black; padding: 5px 10px;">93.9</td> <td style="border-right: 1px solid black; padding: 5px 10px;">6.1</td> </tr> <tr> <td>Double</td> <td style="border-left: 1px solid black; padding: 5px 10px;">80.7</td> <td style="padding: 5px 10px;">19.3</td> </tr> </table>	Single	93.9	6.1	Double	80.7	19.3	
Single	92.1	7.9												
Double	0.9	99.1												
Single	93.9	6.1												
Double	80.7	19.3												

Table S8. Summary of the pairwise mean confusion statistics and associated P values for heterochronous ranked tree shapes under different sampling schemes. The format of the table follows Table S3. The simulated distributions of ranked tree shapes for the off-diagonal entries are the same as those considered in Figure S9 and Table S7. The diagonal entries represent the tests of true null cases, where we generated additional 1,000 random ranked tree shapes per each sampling scenario.

(A) d_1		
	Single	Double
Single	0.5175, 0.9782	
Double	0.0015, 0.0001	0.5065, 0.7484

(B) d_2		
	Single	Double
Single	0.5215, 0.9832	
Double	0.0010, 0.0001	0.5075, 0.8271

(C) $d_{\text{BHV-RTS}}$		
	Single	Double
Single	0.5135, 0.9413	
Double	0.0440, 0.0001	0.4935, 0.2702

(D) $d_{\text{KC-RTS}}$		
	Single	Double
Single	0.4860, 0.1288	
Double	0.4340, 0.0001	0.5000, 0.6279

Table S9. Summary of dispersion comparisons for ranked genealogies. Comparison of dispersion of ranked genealogies using distance functions between ranked genealogies. (A) Isochronous ranked genealogies simulated from different population trajectories under neutral coalescent model (Figure 6). (B) Heterochronous ranked genealogies of real and simulated human influenza A/H3N2 virus data from different geographical regions (Figure 8).

(A) Isochronous ranked genealogies, different demographic models

	d_1^w	d_2^w	$d_{\text{BHV-RG}}$	$d_{\text{KC-RG}}$
Constant	6420771.00	142726.04	33007.42	391150.52
Exponential	1331068.00	24847.09	1397.17	4743.48
Logistic	1744504.00	39291.83	8751.45	109258.59

(B) Heterochronous ranked genealogies, human influenza A/H3N2 virus

	d_1^w	d_2^w	$d_{\text{BHV-RG}}$	$d_{\text{KC-RG}}$
New York – New York	3368.28	53.41	3.90	88.91
New York – Southeast Asia	4581.15	65.93	9.21	263.54
Southeast Asia – Southeast Asia	6370.47	85.97	6.93	152.59
Southeast Asia – New York	5547.11	77.86	8.20	259.08

Table S10. Comparison of metrics: isochronous ranked genealogies under different demographic models. We compare the performance of different distances on ranked genealogies according to how well they separate trees simulated from the $\lambda(t)$ -coalescent with different population histories. The format of the matrices follows Table S2. The simulation values and the color scheme of the L_2 -medoids follow Figure 6. The mean diagonal values are 99.80, 99.87, 33.40, and 73.97 for matrices (A)-(D), respectively.

(A)	d_1^w	(B)	d_2^w	
Uniform	$\begin{bmatrix} 99.6 & 0.4 & 0 \\ 0 & 99.9 & 0.1 \\ 0 & 0.1 & 99.9 \end{bmatrix}$	$\begin{bmatrix} 99.7 & 0.3 & 0 \\ 0 & 99.9 & 0.1 \\ 0 & 0 & 100.0 \end{bmatrix}$		L_2-medoid Uniform Exponential Logistic
Exponential				
Logistic				
(C)	$d_{\text{BHV-RG}}$	(D)	$d_{\text{KC-RG}}$	
Uniform	$\begin{bmatrix} 0.1 & 99.9 & 0 \\ 0 & 100.0 & 0 \\ 0 & 99.9 & 0.1 \end{bmatrix}$	$\begin{bmatrix} 60.0 & 1.0 & 39.0 \\ 0 & 100.0 & 0 \\ 4.4 & 33.7 & 61.9 \end{bmatrix}$		
Exponential				
Logistic				

Table S11. Summary of the pairwise mean confusion statistics and associated P values for isochronous ranked genealogies under different demographic models. The format of the table follows Table S3. The simulated distributions of ranked genealogies for the off-diagonal entries are the same as those considered in Figure 6 and Table S10. The diagonal entries represent the tests of true null cases, where we generated additional 1,000 random ranked genealogies per each demographic scenario.

(A) d_1^w

	Uniform	Exponential	Logistic
Uniform	0.4910, 0.2565		
Exponential	0.0020, 0.0001	0.4870, 0.1493	
Logistic	0.0005, 0.0001	0.0010, 0.0001	0.5015, 0.5805

(B) d_2^w

	Uniform	Exponential	Logistic
Uniform	0.5010, 0.6534		
Exponential	0.0015, 0.0001	0.4840, 0.1144	
Logistic	0.0005, 0.0001	0.0005, 0.0001	0.5125, 0.9081

(C) $d_{\text{BHV-RG}}$

	Uniform	Exponential	Logistic
Uniform	0.4990, 0.4722		
Exponential	0.4995, 0.5093	0.5135, 0.8979	
Logistic	0.4995, 0.4875	0.4995, 0.5139	0.5030, 0.8152

(D) $d_{\text{KC-RG}}$

	Uniform	Exponential	Logistic
Uniform	0.4890, 0.1819		
Exponential	0.1335, 0.0001	0.5070, 0.8513	
Logistic	0.2220, 0.0001	0.1685, 0.0001	0.5060, 0.7667

Table S12. Comparison of metrics: human influenza A/H3N2 virus from different regions. We compare the performance of different distances on ranked genealogies according to how well they separate trees from human influenza A virus from different regions. The format of the matrices follows Table S2. The simulation values and the color scheme of the L_2 -medoids follow Figure 8. The mean diagonal values are 99.9, 99.9, 28.9, and 65.6 for matrices (A)-(D), respectively.

(A)	d_1^w	(B)	d_2^w	L_2 -medoid
NY – NY	$\begin{bmatrix} 99.8 & 0.2 & 0.0 & 0.0 \\ 0.1 & 99.9 & 0.0 & 0.0 \\ 0.0 & 0.0 & 100.0 & 0.0 \\ 0.0 & 0.0 & 0.1 & 99.9 \end{bmatrix}$	NY – NY	$\begin{bmatrix} 99.8 & 0.2 & 0.0 & 0.0 \\ 0.1 & 99.9 & 0.0 & 0.0 \\ 0.0 & 0.0 & 100.0 & 0.0 \\ 0.0 & 0.0 & 0.1 & 99.9 \end{bmatrix}$	New York – New York
NY – SEA		NY – SEA		New York – Southeast Asia
SEA – SEA		SEA – SEA		Southeast Asia – Southeast Asia
SEA – NY		SEA – NY		Southeast Asia – New York
(C)	$d_{\text{BHV-RG}}$	(D)	$d_{\text{KC-RG}}$	
NY – NY	$\begin{bmatrix} 100.0 & 0.0 & 0.0 & 0.0 \\ 98.7 & 0.1 & 1.2 & 0.0 \\ 84.6 & 0.2 & 15.2 & 0.0 \\ 98.7 & 0.0 & 1.1 & 0.2 \end{bmatrix}$	NY – NY	$\begin{bmatrix} 99.7 & 0.3 & 0.0 & 0.0 \\ 46.8 & 38.2 & 11.6 & 3.4 \\ 0.0 & 0.8 & 99.1 & 0.1 \\ 55.6 & 13.0 & 6.1 & 25.3 \end{bmatrix}$	
NY – SEA		NY – SEA		
SEA – SEA		SEA – SEA		
SEA – NY		SEA – NY		

Table S13. Summary of the pairwise mean confusion statistics and associated P values for heterochronous ranked genealogies sampled from two posterior distributions of human influenza A/H3N2 virus (NY-NY and SEA-SEA) and two sets of simulated trees (NY-SEA and SEA-NY). The format of the table follows Table S3. The simulated distributions of ranked genealogies for the off-diagonal entries are the same as those considered in Figure 8 and Table S12. The diagonal entries represent the tests of true null cases, where we generated additional 1,000 random ranked genealogies per each scenario.

(A) d_1^w				
	NY – NY	NY – SEA	SEA – SEA	SEA – NY
NY – NY	0.4985, 0.7797			
NY – SEA	0.0015, 0.0001	0.4965, 0.4107		
SEA – SEA	0.0000, 0.0001	0.0000, 0.0001	0.4780, 0.0365	
SEA – NY	0.0000, 0.0001	0.0000, 0.0001	0.0005, 0.0001	0.5030, 0.6356

(B) d_2^w				
	NY – NY	NY – SEA	SEA – SEA	SEA – NY
NY – NY	0.4925, 0.3424			
NY – SEA	0.0015, 0.0001	0.5075, 0.8267		
SEA – SEA	0.0000, 0.0001	0.0000, 0.0001	0.4915, 0.2798	
SEA – NY	0.0000, 0.0001	0.0000, 0.0001	0.0005, 0.0001	0.4850, 0.1196

(C) $d_{\text{BHV-RG}}$				
	NY – NY	NY – SEA	SEA – SEA	SEA – NY
NY – NY	0.4890, 0.1764			
NY – SEA	0.4995, 0.5242	0.4720, 0.0034		
SEA – SEA	0.4240, 0.0001	0.3025, 0.0001	0.5000, 0.5256	
SEA – NY	0.4990, 0.5154	0.1435, 0.0001	0.4225, 0.0001	0.5005, 0.5278

(D) $d_{\text{KC-RG}}$				
	NY – NY	NY – SEA	SEA – SEA	SEA – NY
NY – NY	0.4820, 0.0632			
NY – SEA	0.2395, 0.0001	0.4955, 0.3624		
SEA – SEA	0.0005, 0.0001	0.0625, 0.0001	0.4860, 0.1215	
SEA – NY	0.2810, 0.0001	0.2540, 0.0001	0.0355, 0.0001	0.5160, 0.9560

Table S14. The NCBI GenBank accession numbers for the HA sequences used in the analysis of human influenza A/H3N2 virus from different geographical regions.

(A) New York

CY000001	CY000017	CY000025	CY000057	CY000177	CY000065	CY000089	CY000097	CY000113
CY000145	CY000209	CY000217	CY000233	CY000257	CY000265	CY000345	CY000353	CY000361
CY000033	CY000401	CY000409	CY000417	CY000433	CY000441	CY000489	CY000545	CY000585
CY000753	CY000761	CY000777	CY000873	CY000769	CY000933	CY000941	CY000965	CY001037
CY001088	CY000049	CY001072	CY001061	CY001045	CY000137	CY001237	CY001144	CY001301
CY001317	CY001333	CY001325	CY001293	CY001429	CY001469	CY001512	CY001552	CY001624
CY001640	CY000073	CY001736	CY002032	CY002072	CY002176	CY002224	CY002200	CY002248
CY002240	CY002264	CY000865	CY002424	CY002408	CY002416	CY002456	CY002472	CY002488
CY002480	CY002504	CY002608	CY002712	CY002720	CY002728	CY002736	CY002784	CY002776
CY003056	CY003096	CY003104	CY003120	CY003144	CY003160	CY003168	CY003192	CY003112
CY003040	CY003344	CY003424	CY003656	CY003777	CY006115	CY006147	CY006291	CY006371
CY006435	CY007643	CY008884	CY008868	CY009260	CY009244	CY013805	CY019165	CY019173
CY019189	CY019253	CY019285	CY019293	CY019811	EF473618	EF473619	EF473625	CY020533
EU501484	EU502299	EU502307	EU502310	EU502316				

(B) Southeast Asia

DQ865945	DQ865949	DQ865951	DQ865955	DQ865957	DQ865958	DQ865959	DQ865961	DQ865962
DQ865970	AB281195	AB281200	AB281205	AB281210	AB281215	AB281232	AB281235	AB281238
AB281241	AB281244	AB281247	AB281250	EF566141	EF566142	EF566155	EF566164	EF566173
EF566176	EF566224	EF566227	EF566229	EF566332	EF566361	EF566362	EF566365	EF566051
EF566053	EF566067	EF566068	EF566074	EU501124	EU501125	EU501126	EU501127	EU501161
EU501168	EU501171	EU501172	EU501174	EU501176	EU501177	EU501221	EU501279	EU501281
EU501283	EU501308	EU501312	EU501317	EU501318	EU501319	EU501320	EU501321	EU501372
EU501373	EU501374	EU501375	EU501385	EU501441	EU501458	EU501459	EU501460	EU501461
EU501462	EU501463	EU501464	EU501465	EU501466	EU501467	EU501474	EU501531	EU501541
EU501542	EU501544	EU501545	EU501546	EU501547	EU501548	EU501558	EU501622	EU501773
EU501774	EU501787	EU501788	EU501789	EU501790	EU501791	EU501792	EU501793	EU501795
EU501800	EU502344	EU514639	EU514653	EU514654	EU514667	FJ229884	FJ561060	FJ865282
FJ865283	FJ865284	CY091149	CY091157	CY091165	CY091229	CY091245	CY091253	CY091261
CY091437	CY091445	CY091453	CY091469	CY091485				

Table S15. Summary of distortions. Comparison of distortion (Text 11A) using distance functions between ranked trees shapes and ranked genealogies.

(A) Comparison of distortions on ranked tree shapes. The simulated data used for computation are the same considered for Table S1

	d_1	d_2	$d_{\text{BHV-RTS}}$	$d_{\text{KC-RTS}}$	$d_{\text{CP-RTS}}$
Isochronous ranked tree shapes (beta splitting)	5582.02	2369.98	8586.00	14331.64	27721.77
Isochronous ranked tree shapes (alpha-beta splitting)	3979.43	3028.30	10941.54	23989.64	5446.20
Heterochronous ranked tree shapes	2343.93	2969.33	4553.04	12034.03	—

(B) Comparison of distortions on ranked genealogies. The simulated data used for computation are the same considered for Table S9.

	d_1^w	d_2^w	$d_{\text{BHV-RG}}$	$d_{\text{KC-RG}}$
Isochronous ranked genealogies	6751.44	1836.89	18569.28	13716.09
Heterochronous ranked genealogies	13938.84	7196.04	16197.56	3368.99

Table S16. Summary of correlations. Comparison of correlation (Text 11B) between original distances and Euclidean distances in 2-dimensional MDS comparisons.

(A) Comparison of correlations of original distances and Euclidean distances in two-dimensional MDS plots on ranked tree shapes. The simulated data used for computation are the same considered for Table S1.

	d_1	d_2	$d_{\text{BHV-RTS}}$	$d_{\text{KC-RTS}}$	$d_{\text{CP-RTS}}$
Isochronous ranked tree shapes (beta splitting)	0.998	0.998	0.401	0.982	0.988
Isochronous ranked tree shapes (alpha-beta splitting)	0.998	0.999	0.648	0.518	0.909
Heterochronous ranked tree shapes	0.984	0.986	0.634	0.734	—

(B) Comparison of correlations of original distances and Euclidean distances in two-dimensional MDS plots on ranked genealogies. The simulated data used for computation are the same considered for Table S9.

	d_1^w	d_2^w	$d_{\text{BHV-RG}}$	$d_{\text{KC-RG}}$
Isochronous ranked genealogies	0.981	0.954	0.527	0.946
Heterochronous ranked genealogies	0.980	0.955	0.096	0.844

References

1. Billera LJ, Holmes SP, Vogtmann K (2001) Geometry of the space of phylogenetic trees. *Advances in Applied Mathematics* 27(4):733–767.
2. Kendall M, Colijn C (2016) Mapping phylogenetic trees to reveal distinct patterns of evolution. *Molecular Biology and Evolution* 33(10):2735–2743.
3. Colijn C, Plazzotta G (2018) A metric on phylogenetic tree shapes. *Systematic Biology* 67(1):113–126.
4. Aldous D (1996) Probability distributions on cladograms in *Random Discrete Structures*, eds. Aldous D, Pemantle R. (Springer New York, New York, NY), pp. 1–18.
5. Ford DJ (2005) Probabilities on cladograms: introduction to the alpha model. *arXiv:math/0511246 [math.PR]*.
6. Maliet O, Gascuel F, Lambert A (2018) Ranked tree shapes, nonrandom extinctions, and the loss of phylogenetic diversity. *Systematic Biology* 67(6):1025–1040.
7. Sainudiin R, Stadler T, Véber A (2015) Finding the best resolution for the Kingman-Tajima coalescent: theory and applications. *Journal of Mathematical Biology* 70(6):1207–1247.
8. Palacios JA, et al. (2019) Bayesian estimation of population size changes by sampling Tajima’s trees. *Genetics* 213(3): 967–986.
9. Cappello L, Palacios JA (2019) Sequential importance sampling for multi-resolution Kingman-Tajima coalescent counting. *arXiv:1902.05527 [stat.AP]*.
10. Felsenstein J, Rodrigo AG (1999) Coalescent approaches to HIV population genetics in *The Evolution of HIV*, ed. Crandall KA. (Johns Hopkins University Press, Baltimore, Maryland), pp. 233–272.
11. Bourgain J (1985) On Lipschitz embedding of finite metric spaces in Hilbert space. *Israel Journal of Mathematics* 52(1-2): 46–52.