

AN ADJACENT-SWAP MARKOV CHAIN ON COALESCENT TREES

MACKENZIE SIMPER ^{*}, *and*
JULIA A. PALACIOS,^{**} *Stanford University*

Abstract

The standard coalescent is widely used in evolutionary biology and population genetics to model the ancestral history of a sample of molecular sequences as a rooted and ranked binary tree. In this paper we present a representation of the space of ranked trees as a space of constrained ordered matched pairs. We use this representation to define ergodic Markov chains on labeled and unlabeled ranked tree shapes analogously to transposition chains on the space of permutations. We show that an adjacent-swap chain on labeled and unlabeled ranked tree shapes has a mixing time at least of order n^3 , and at most of order n^4 . Bayesian inference methods rely on Markov chain Monte Carlo methods on the space of trees. Thus it is important to define good Markov chains which are easy to simulate and for which rates of convergence can be studied.

Keywords: Random transpositions; lumped Markov chain; Tajima distribution

2020 Mathematics Subject Classification: Primary 60J10
Secondary 60J90

1. Introduction

The standard Kingman coalescent [16] is often used in evolutionary biology and population genetics to model the ancestry of n molecular sequences, such as DNA from a single segment (locus). This ancestry is represented by a genealogy, a timed and rooted binary tree [33]. In the neutral coalescent, the tree topology and the coalescent times, i.e. the times when two branches merge into a single node, are independent, and the density of the coalescent times is usually parameterized in terms of parameters of interest, such as the effective population size. The central idea of coalescent modeling in population genetics is that the observed molecular sequences at the tips are the result of a process of mutations superimposed on the genealogy. Therefore the more similar sequences are, the more recent their common ancestors (short coalescent times) are in the genealogy.

Bayesian inference methods under this model usually approximate the posterior distribution of model parameters via Markov chain Monte Carlo (MCMC) methods [10]. Therefore it is important to understand the mixing of Markov chains on the space of genealogies. In particular, we are interested in understanding the mixing of Markov chains on discrete tree topologies

Received 31 May 2021; revision received 26 January 2022.

^{*} Postal address: Department of Mathematics, Building 380, 450 Jane Stanford Way, Stanford, CA 94305, USA. Email address: msimper@stanford.edu

^{**} Postal address: Department of Statistics and Department of Biomedical Data Science, Sequoia Hall, 390 Jane Stanford Way, Stanford, CA 94305, USA. Email address: juliapr@stanford.edu

© The Author(s), 2022. Published by Cambridge University Press on behalf of Applied Probability Trust.

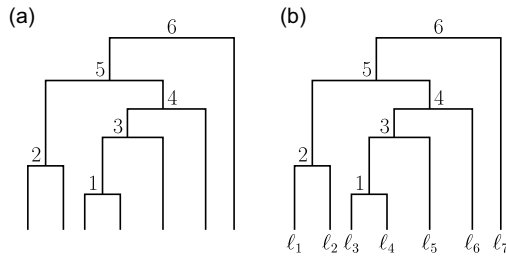


FIGURE 1. Examples of coalescent tree topologies: (a) ranked and unlabeled; (b) ranked and labeled. A ranked and unlabeled tree shape T has internal nodes labeled by their ranking from leaves to root and unlabeled leaves. Tree (a) has ordered matched pairs: $(0, 0)^1, (0, 0)^2, (0, 1)^3, (0, 3)^4, (2, 4)^5, (0, 5)^6$. A ranked and labeled tree T^L has unique leaf and internal node labels. Tree (b) has ordered matched pairs: $(\ell_3, \ell_4)^1, (\ell_1, \ell_2)^2, (\ell_5, 1)^3, (\ell_6, 3)^4, (2, 4)^5, (\ell_7, 5)^6$.

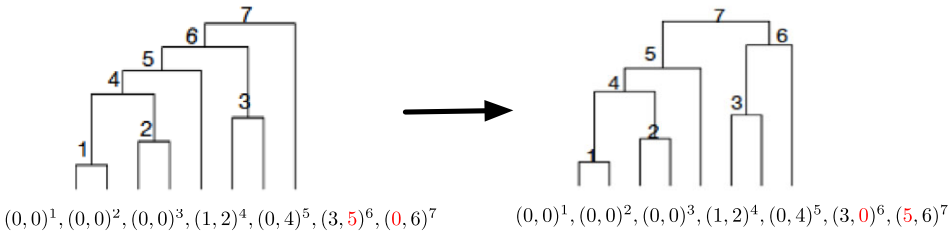


FIGURE 2. Example of an adjacent-swap move on ranked tree shapes. The pair $k = 6$ is selected uniformly at random among pairs $1, \dots, 6$. We then select uniformly at random one element from pair 6 and one element from pair 7. The only allowable swaps are 5 and 0 or 3 and 0. The new state (right tree) is obtained by swapping 5 and 0 from pairs 6 and 7 respectively in the left tree.

only: labeled and unlabeled ranked tree shapes. We will therefore assume a unit interval length between consecutive coalescing (branching) events.

A labeled *ranked tree shape* $T^L \in \mathcal{T}_n^L$ with n leaves is a rooted labeled and ranked binary tree. The leaves are labeled by the set $\{\ell_1, \ell_2, \dots, \ell_n\}$ and internal nodes are labeled $1, \dots, n - 1$ in increasing order, with label $n - 1$ at the root (Figure 1(b)). The cardinality of \mathcal{T}_n^L is given by

$$|\mathcal{T}_n^L| = \frac{n!(n - 1)!}{2^{n-1}}.$$

This equation can be derived by imagining building the tree from leaves to root: Initially there are n leaves and n lineages, and at successive steps two lineages are chosen to merge. It takes $n - 1$ merges to get to a single lineage and at merge $k + 1$ there are $\binom{n-k}{2}$ choices. The total number of trees is

$$\prod_{k=0}^{n-2} \binom{n-k}{2} = \prod_{j=2}^n j(j-1)/2.$$

In many applications, the quantity of interest is the overall shape of the tree [14, 17, 22]; the specific labels of the samples are unimportant. In fact, a key feature of simple coalescent

models, such as the Kingman coalescent, is the assumption that the leaves are *exchangeable*. Recently, Palacios *et al.* [25] proposed a new Bayesian approach from binary molecular data in which the Kingman coalescent model on genealogies is replaced by the Tajima coalescent. The Tajima coalescent is a lumping of the Kingman coalescent that ignores leaf labels. The premise in [25] is that the new target Tajima posterior distribution is more evenly spread among several tree topologies than the Kingman posterior distribution, and hence Markov chain exploration of the whole state space is more efficient [4]. This motivates the study of *unlabeled ranked tree shapes* with n leaves \mathcal{T}_n (Figure 1(a)). The cardinality of \mathcal{T}_n is given by the $(n - 1)$ th Euler zig-zag number E_{n-1} [18] and $E_{n-1} \ll |T_n^L|$. While one appealing feature of modeling unlabeled ranked tree shapes (Tajima coalescent) is the reduction in the cardinality of the state space, when compared to the cardinality of the state space under Kingman coalescent [3], the mixing of the chains can still be of the same order. Although the initial motivation is to compare the mixing of Markov chains on labeled and unlabeled ranked tree shape posterior distributions, we center our work on the Markov chains whose stationary distributions are uniform (or coalescent) distributions on the corresponding spaces (ignoring the data). This study should provide a first step towards a better understanding of coalescent chains in more general settings.

Related work in the context of evolutionary models includes the analysis of Markov chains on cladograms (or phylogenies). Cladograms are rooted or unrooted binary trees with labeled leaves and no ranking of internal nodes. In [5], the space of cladograms with l leaves is shown to be in bijection with the space of perfect matchings of $2n$ labels, with $n = l - 1$. The bijection with perfect matchings is used to define a Markov chain on the space of cladograms. Using the perfect matching perspective, the chain is analogous to a *random transpositions* chain on the space S_{2n} of permutations of $2n$ elements. This representation allows the use of known results for the random transposition chain to be applied to the chain on trees to determine sharp bounds on rates of convergence [6]. Motivated by this result, we propose a new representation of labeled and unlabeled ranked tree shapes as a type of matching. We use this representation to define an adjacent-swap chain on the spaces of labeled and unlabeled ranked tree shapes, and to study their mixing time. It is important to study the mixing time for chains on ranked trees because these results have implications for the design of MCMC algorithms used to infer evolutionary parameters from molecular data in a Bayesian setting [8, 10]. Another use for Markov chains on the space of ranked tree shapes is combinatorial optimization, either for finding the maximum likelihood tree topology [13] or for summarizing samples from the output of a Bayesian algorithm via finding the Fréchet mean ranked tree shape [26]. These models are routinely used for analyzing molecular sequences of pathogens and tracking epidemics [32], and for understanding past population dynamics from present day and ancient DNA samples [21].

In the rest of this section we define a bijective representation of ranked trees as constrained ordered matchings and define the adjacent-swap chain. We then state our main results concerning mixing times and discuss related work. In Section 2, we state known results on Markov chains that will be used in the following sections. In Section 3 we state important properties of the proposed adjacent-swap chain. The lower bound for the mixing time of the adjacent-swap chains is proved in Section 4 and it is obtained by finding a specific function that gives a lower bound for the relaxation time of the chains. In Section 5 we prove the upper bounds on the mixing time of the adjacent-swap chains via a coupling argument. In the discussion in Section 6 we mention future directions and other related chains on labeled and unlabeled coalescent trees which could be defined using the matching representations, and discuss implications for MCMC methods applied in practice.

1.1. Matching representations

We first establish some definitions. Let m be an integer and S a multiset of size $2m$. A *matching* of S objects is a partition of the elements into m disjoint subsets of size two. An *ordered matching* is a matching with a linear ordering (ranking) of the m pairs.

Definition 1. The space $COM_m(S)$ (constrained ordered matchings) is the set of all ordered matchings of S with pairs labeled $\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_m$, which satisfy, for $k = 1, \dots, m$ and $a \in S$,

$$a \in \mathbf{p}_k \implies \begin{cases} a \in \mathbb{Z} \text{ and } a < k, & \text{or} \\ a \in \{\ell_1, \dots, \ell_n\}. \end{cases} \tag{1}$$

To represent a labeled ranked tree shape as a constrained ordered matching, we take S to be the set of leaf and internal node labels, and each pair in the matching represents a coalescence in the tree, with \mathbf{p}_k representing the k th coalescence event. The condition (2) then simply ensures that no interior node can be merged before it has been introduced. A leaf can merge at any point and thus there is no constraint on a leaf’s position in the matching.

Lemma 1. With $S = \{\ell_1, \ell_2, \dots, \ell_n, 1, 2, \dots, n - 2\}$, the space $COM_{n-1}(S)$ is in bijection with \mathcal{T}_n^L . With multiset $S = \{0, 0, \dots, 0, 1, 2, \dots, n - 2\}$, with 0 repeated n times, the space $COM_{n-1}(S)$ is in bijection with \mathcal{T}_n .

Proof. First assume that $S = \{\ell_1, \ell_2, \dots, \ell_n, 1, 2, \dots, n - 2\}$. The first matched pair can be formed by any two leaves in $\binom{n}{2}$ possible ways. The second pair can be formed by any of the $n - 2$ remaining leaves or 1, the label of the first matched pair, in $\binom{n-1}{2}$ ways; in general, the k th matched pair can be formed in $\binom{n-k+1}{2}$ ways, because there are $n - k + 1$ unique labels that could be matched at that time. Thus

$$|COM_{n-1}(S)| = \prod_{i=2}^n \binom{i}{2} = |\mathcal{T}_n^L|$$

and for every element $T^L \in \mathcal{T}_n^L$ there is a one-to-one mapping between every coalescence event in T^L and every matched pair in one element of $COM_{n-1}(S)$. These two facts imply there is a one-to-one mapping between \mathcal{T}_n^L and $COM_{n-1}(S)$.

With the leaf labels ℓ_1, \dots, ℓ_n replaced by the repeated element 0, the space of matchings is in bijection with \mathcal{T}_n due to the fact that the space of \mathcal{T}_n is equivalent to the space of \mathcal{T}_n^L with the leaf labels removed. That is, there is a surjection $\mathcal{T}_n^L \rightarrow \mathcal{T}_n$ equivalent to the surjection $COM_{n-1}(S_1) \rightarrow COM_{n-1}(S_2)$, with $S_1 = \{\ell_1, \ell_2, \dots, \ell_n, 1, 2, \dots, n - 2\}$ and $S_2 = \{0, 0, \dots, 0, 1, 2, \dots, n - 2\}$. The bijective matching representations of the ranked trees \mathcal{T}_n and \mathcal{T}_n^L of Figure 1 are shown in the legend. □

1.2. Markov chain and main results

Definition 2. The adjacent-swap Markov chain on the space $COM_{n-1}(S)$ is defined by the following update move.

1. Pick an index $k \in \{1, \dots, n - 2\}$ uniformly at random.
2. Pick labels a_1 and a_2 uniformly from the pairs \mathbf{p}_k and \mathbf{p}_{k+1} respectively.
 - (a) If swapping the positions of a_1, a_2 does not violate constraint (1), make the swap.
 - (b) Otherwise, remain in the current state.

For the sets S which give the bijections with \mathcal{T}_n and \mathcal{T}_n^L , this Markov chain is connected and reversible. In the case of \mathcal{T}_n^L it is also symmetric and hence it has a uniform stationary distribution. For \mathcal{T}_n , the chain is not symmetric; the stationary distribution is the *Tajima distribution* on the space of ranked unlabeled trees. These facts, as well as the relationship between the two chains, will be proved in a later section.

We are interested in the convergence rates of this Markov chain and how this rate compares depending on the space \mathcal{T}_n or \mathcal{T}_n^L . The measure that we study for convergence rate is the *mixing time*. That is, for a Markov chain on a space Ω with transition probability P and stationary distribution π , the mixing time is defined as

$$t_{\text{mix}} = \sup_{x_0 \in \Omega} \inf\{t > 0: \|P^t(x_0, \cdot) - \pi(\cdot)\|_{TV} < 1/4\}.$$

To state the result, let t_n and t_n^L be the mixing times for the chains on \mathcal{T}_n and \mathcal{T}_n^L respectively. Our main result is as follows.

Theorem 1. *There exist constants C_1, C_2, C_3 such that*

$$C_1 \cdot n^3 \leq t_n \leq C_2 \cdot n^4$$

and

$$C_1 \cdot n^3 \leq t_n^L \leq C_3 \cdot n^4.$$

The lower bound of n^3 comes from the *relaxation time*, and involves a standard trick of finding a specific function for which the variance under the stationary distribution can be bounded. The chain on \mathcal{T}_n is a certain type of ‘lumping’ of the chain on \mathcal{T}_n^L , and so the same lower bound applies to both spaces. The upper bounds are obtained using a coupling argument. We conjecture that the mixing time t_n is indeed smaller than the time t_n^L , though it is not evident whether this is by a significant order, or just a constant factor.

1.3. Related work

Mixing of chains on other tree spaces have been previously studied, in particular, the mixing of Markov chains on the space of cladograms. An n -leaf *cladogram* is a rooted or unrooted tree with n labeled leaves and unlabeled internal nodes of degree 3. The main difference between cladograms and the tree topologies studied in this manuscript is that cladograms do not rank internal branching events. The cardinality of the space of cladograms of n leaves is $(2(n-1)!/2^{n-1}(n-1)!)$, smaller than the cardinality of the space of ranked and labeled trees but larger than the cardinality of the space of ranked unlabeled trees. Cladograms are fundamental objects in phylogenetics to model ancestral relationships at the species level, while ranked tree shapes are fundamental objects in population genetics and phylodynamics of infectious diseases. Ranked tree shapes are used to model ancestral relationships of a sample of individuals from a single population of the same species [12].

Markov chains on cladograms. Aldous [2] studied a chain on unrooted cladograms that removes a leaf chosen uniformly at random from the current cladogram and reattaches it to a random edge. Using coupling methods, Aldous [2] showed that the relaxation time for this chain is $\tau_{\text{rel}} \leq c_2 n^3$. The same chain was later analyzed by Schweinsberg [28], who showed that $\tau_{\text{rel}} = O(n^2)$ using a modified method of distinguished paths. In [5], Diaconis and Holmes show that the space of rooted cladograms with n leaves is in bijection with the space of perfect matchings in the complete graph on $2(n-1)$ vertices. This bijection was later used by

Diaconis and Holmes [6] to define a Markov chain by randomly choosing two pairs and transposing two randomly selected entries of each pair. They showed that this random transposition chain mixed after $\frac{1}{2}n \log n$ steps. In [29], Spade, Herbei, and Kubatko study the nearest neighbor interchange (NNI) chain and a subtree prune and regraft (SPR) chain on the space of rooted cladograms. They showed that the upper bounds on the relaxation time of the SPR and the NNI chains are $O(n^{5/2})$ and $O(n^4)$ respectively.

Other related work includes the study of mixing times of chains whose stationary distribution is the posterior distribution over cladograms [24, 31]. In [24], Mossel and Vigoda show exponential mixing times under a misspecified model in which data is generated from a mixture of cladograms.

Markov chains on related spaces. The adjacent-swap Markov chain on the space of ranked trees studied in this manuscript is closely related to the adjacent transposition chain on the space S_n of permutations. As is often done, one can think of an element $\sigma \in S_n$ as an ordering of a deck of cards with unique labels $1, \dots, n$. In [1, Example 4.10], the chain is defined by picking at random two adjacent cards from the deck; with probability $1/2$, the cards are swapped and with probability $1/2$ nothing happens. The chain is symmetric with uniform stationary distribution. Aldous [1] showed that the mixing time of this chain has a lower bound of $c_1 n^3$ and an upper bound of $c_2 n^3 \log(n)$. In [34], the now-famous ‘Wilson’s method’ was introduced and applied to the adjacent transposition chain as an example to improve the lower bound to order $n^3 \log(n)$ with the explicit constant $(1/\pi^2)n^3 \log(n)$, as well as an upper bound of $(2/\pi^2)n^3 \log(n)$. Finally, in [19], the upper bound was improved to $(1/\pi^2)n^3 \log(n)$, so that the upper and lower bounds matched; in addition, the chain was proved to follow the *cut-off* phenomenon. Durrett generalized the adjacent transposition chain to the L -reversal chain, introduced as a model for the evolution of DNA in a chromosome [11].

As noted in [23], to the best of our knowledge, bounds on the mixing time of chains on ranked tree shapes have not been studied before.

2. Preliminaries

In this section we review some results of Markov chain theory that will be used to bound the mixing time of the adjacent-swap chains on \mathcal{T}_n^L and \mathcal{T}_n . We use the coupling method to find the upper bounds. A *coupling* of Markov chains with transition matrix P is a process $(X_t, Y_t)_{t \geq 0}$ such that both (X_t) and (Y_t) are marginally Markov chains with transition matrix P . The goal is to define a coupling of two chains from different starting distributions such that the two chains will quickly reach the same state, and once this occurs the two chains stay matched forever. This *coupling time* gives an upper bound on the mixing time; see Chapter 5 from [20] for more details.

Theorem 2. (Theorem 5.4 [20].) *Suppose that for each pair of states $x, y \in \mathcal{X}$ there is a coupling (X_t, Y_t) with $X_0 = x$ and $Y_0 = y$. For each such coupling, let τ_{couple} be the coalescence time of the chains, that is,*

$$\tau_{\text{couple}} := \min\{t: X_s = Y_s \text{ for all } s \geq t\}.$$

Then

$$t_{\text{mix}} \leq 4 \max_{x,y} E_{x,y}(\tau_{\text{couple}}).$$

To find a lower bound on the mixing time, we bound the relaxation time and use the following result.

Theorem 3. (Theorem 12.5 [20].) *Suppose P is the transition matrix of an irreducible, aperiodic, and reversible Markov chain. Then*

$$t_{\text{mix}} \geq (\tau_n - 1) \log(2). \tag{2}$$

The relaxation time is defined as the inverse of the absolute spectral gap, $\tau_n = 1/\gamma^*$, where $\gamma^* = 1 - |\lambda_{n,2}|$, one minus the absolute value of the eigenvalue of the transition matrix of the chain that has second-largest magnitude. The spectral gap γ is one minus the second-largest eigenvalue. If the Markov chain is lazy, then $\gamma = \gamma^*$, since all eigenvalues are positive. As we defined it, the adjacent-swap chain is not lazy, so we have $\gamma^* \leq \gamma$. We will get a lower bound on τ_n by finding an upper bound on γ using the following variational characterization.

Theorem 4. (Lemma 13.7 [20].) *Let P be the transition matrix for a reversible Markov chain. The spectral gap γ satisfies*

$$\gamma = \min_{f: \mathcal{X} \rightarrow \mathbb{R}, \text{Var}_\pi(f) \neq 0} \frac{\mathcal{E}(f)}{\text{Var}_\pi(f)},$$

where

$$\mathcal{E}(f) := \frac{1}{2} \sum_{x,y \in \mathcal{X}} [f(x) - f(y)]^2 \pi(x)P(x, y).$$

3. The adjacent-swap chain on ranked tree shapes

Lemma 2. *The adjacent-swap Markov chain on \mathcal{T}_n^L is irreducible, aperiodic, and reversible with respect to the uniform stationary distribution $\pi^L(x) := 1/|\mathcal{T}_n^L|$.*

Proof. This can be seen by noting the transition matrix is symmetric. Let P^L be the transition matrix for the adjacent-swap chain on \mathcal{T}_n^L . That is, for $x, y \in \mathcal{T}_n^L$,

$$P^L(x, y) = \frac{1}{n-2} \cdot \frac{1}{4},$$

where y is a tree that can be obtained by swapping two elements from two adjacent pairs in x , e.g. consider the following states at pairs k and $k+1$:

$$\begin{aligned} x: & (a, b)^k, (c, d)^{k+1}, \\ y: & (a, c)^k, (b, d)^{k+1}. \end{aligned}$$

In a labeled ranked tree shape \mathcal{T}_n^L , all pairs are formed by distinct elements, so a, b, c, d are all unique and the only way to transition from x to y is to swap the labels b and c , which happens with probability $1/4(n-2)$ if $c < k$ or $c \in \{\ell_1, \dots, \ell_n\}$.

To see that the chain is irreducible, we note that any label can be moved to any pair to the right, one pair (or step) at a time with positive probability, and that any label can be moved to a pair with index k to the left (one pair at a time) as long as the label corresponds to a leaf or an internal node smaller than k , satisfying constraint (1). To see that the chain is aperiodic, we note that $P^L(x, x) > 0$, for all $x \in \mathcal{T}_n^L$. Independent of the current state, we can pick pair $k = n-2$ and attempt to swap any element of the $n-2$ pair with label $n-2$ from the $n-1$ pair with probability $1/2(n-2)$. Since this move violates constraint (1), the move is rejected and the chain remains in the current state. □

We note that the transition matrix P of the adjacent-swap chain on unlabeled ranked tree shapes is not symmetric. For example, consider the transition x to y of the type

$$(0, a)^k, (0, b)^{k+1} \rightarrow (0, 0)^k, (a, b)^{k+1} \tag{3}$$

for labels $a, b \neq 0$ (recall the 0s represent leaf labels). The probability of this transition is $P(x, y) = 1/(n - 2) \cdot 1/4$, since once pairs k and $k + 1$ are chosen, there is a $1/4$ chance of choosing label a from pair k and label 0 from pair $k + 1$. However, the reverse transition,

$$(0, 0)^k, (a, b)^{k+1} \rightarrow (0, a)^k, (0, b)^{k+1},$$

has probability $1/(n - 2) \cdot 1/2$. It turns out that the stationary distribution π for the chain is the *Tajima coalescent* distribution [27] (also known as the Yule distribution). For $T \in \mathcal{T}_n$,

$$\pi(T) = \frac{2^{n-c(T)-1}}{(n - 1)!},$$

where $c(T)$ is the number of cherries of T , i.e. the number of pairs of the type (0,0) in the $\text{COM}_{n-1}(S)$ representation. Indeed, for transitions x to y of the type of (3), y has one more cherry than x and $\pi(x)P(x, y) = \pi(y)P(y, x)$; for other types of transitions $P(x, y) > 0$ that do not affect the number of cherries, $P(x, y) = P(y, x)$, $\pi(x) = \pi(y)$ and the detailed balance equation is satisfied.

Another way of proving that the Tajima coalescent distribution is the stationary distribution on unlabeled ranked tree shapes is to view the chain as a *lumping* of the chain on \mathcal{T}_n^L . This perspective also gives us an initial comparison of the relaxation times of the chains. The space \mathcal{T}_n of unlabeled ranked tree shapes can be considered as a set of equivalency classes of the trees \mathcal{T}_n^L . That is, for trees $x, y \in \mathcal{T}_n^L$, define the equivalence relation $x \sim y$ if the trees have the same ranked tree shape. From the $\text{COM}_{n-1}(S)$ perspective with $S = \{\ell_1, \dots, \ell_n, 1, 2, \dots, n - 2\}$, two matchings are equivalent if all internal node labels $1, \dots, n - 2$ occur in the same pairs.

This equivalence relation induces a partition of \mathcal{T}_n^L using equivalence classes. That is, we can write \mathcal{T}_n^L as the disjoint union of sets $\Omega_1, \dots, \Omega_M$, where $M = |\mathcal{T}_n|$, and all trees in Ω_i have the same ranked tree shape.

Lemma 3. *For any $x, x' \in \mathcal{T}_n^L$, equivalence class Ω_i , and $x \sim x'$, the following relation holds:*

$$P^L(x, \Omega_i) := \sum_{y \in \Omega_i} P^L(x, y) = P^L(x', \Omega_i).$$

Proof. If $x \in \Omega_i$, then the transition from x to $y \in \Omega_i$ is of the type

$$(a, b)^k, (c, d)^{k+1} \rightarrow (c, b)^k, (a, d)^{k+1},$$

where $a, c \in \{\ell_1, \ell_2, \dots, \ell_n\}$ correspond to leaf labels. Since x and x' differ only at leaf labels, the transition probability of swapping two leaf labels is the same and $P^L(x, \Omega_i) = P^L(x', \Omega_i)$. If $x \notin \Omega_i$, and therefore $x' \notin \Omega_i$, then the transition from x to y is of the type

$$(a, b)^k, (c, d)^{k+1} \rightarrow (c, b)^k, (a, d)^{k+1},$$

where a and c are not both in $\{\ell_1, \ell_2, \dots, \ell_n\}$. Since x and x' differ only at leaf labels, again $P^L(x, \Omega_i) = P^L(x', \Omega_i)$. □

In words, Lemma 3 says that the probability of transitioning to a different ranked tree shape is independent of the leaf configuration of the current tree. A well-known result (e.g. Lemma 2.5 in [20]) is that the induced chain on the space of equivalence classes defined by $\tilde{P}([x], [y]) := P(x, [y])$ is a Markov chain; then observe that the induced chain \tilde{P} is equivalent to adjacent-swap Markov chain P on the space \mathcal{T}_n . This allows a comparison of the spectral gaps of the two chains. Lemma 12.9 in [20] states that the eigenvalues of the transition matrix of the lumped chain are eigenvalues of the transition matrix of the full chain, hence we have the following lemma.

Lemma 4. *Let γ, γ^L be the spectral gaps of the chains P, P^L respectively. Then $\gamma^L \leq \gamma$.*

From Lemma 3 we can immediately see that the stationary distribution for P is indeed the Tajima coalescent distribution. Suppose $T \in \mathcal{T}_n$ corresponds to the equivalence class $\Omega_i \subset \mathcal{T}_n^L$. Then

$$\pi(T) = \sum_{y \in \Omega_i} \pi^L(y) = \frac{|\Omega_i|}{|\mathcal{T}_n^L|}.$$

A uniformly sampled labeled ranked tree shape with the leaf labels erased gives an unlabeled ranked tree shape according to the Tajima distribution [27]. This implies that $|\Omega_i|/|\mathcal{T}_n^L|$ is equal to the probability under the Tajima distribution for the ranked tree shape corresponding to Ω_i .

4. Lower bound

To prove the lower bounds in Theorem 1, we use the variational characterization of the relaxation time from Theorem 4:

$$\tau_n = \sup_{f: \Omega \rightarrow \mathbb{R}} \frac{\text{Var}_\pi(f)}{\mathcal{E}(f)}. \tag{4}$$

The variational characterization can be used to obtain a lower bound on the relaxation time using a specific function f . To achieve a tight lower bound, a common strategy is to find a function f which has a large variance but can only be changed by a constant amount by one step of the Markov chain. That is, for $x, y \in \Omega$ such that $P(x, y) > 0$, we have $(f(x) - f(y))^2 \leq C$.

We will use the *internal tree length* as a function $\varphi(x): \mathcal{T}_n \rightarrow \mathbb{R}$ to find a lower bound for τ_n . Since the internal tree length is independent of the leaf labels, we get the same lower mixing time bound for the adjacent-swap chains on \mathcal{T}_n and \mathcal{T}_n^L . For ease of exposition we will focus the following discussion on \mathcal{T}_n .

We will define the internal tree length function on the space \mathcal{T}_n using the correspondence with $\text{COM}_{n-1}(S)$ and assuming unit length between consecutive coalescence events. For a label $j \in \{1, \dots, n-2\}$, let $I_x(j)$ denote the pair index k such that $j \in \mathbf{p}_k$. In the example of Figure 3, $I_x(2) = 5$. Note that from the constraints (equation (1)), we have

$$I_x(j) \in \{j + 1, \dots, n - 1\}.$$

We now define the internal tree length function on \mathcal{T}_n as follows:

$$\varphi(x) = \sum_{j=1}^{n-2} (I_x(j) - j).$$

As an example consider the ranked unlabeled tree shape with $n = 7$ depicted in Figure 3. Note that the function is minimized for the ‘caterpillar tree’, $(0, 0)^1, (0, 1)^2, (0, 2)^3, \dots, (0, n - 2)^{n-1}$, where $I_x(j) = j + 1$ for all j , giving $\varphi(x) = n - 2$.

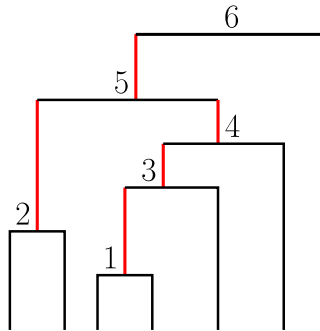


FIGURE 3. Example of internal tree length calculation. A ranked tree shape T with ordered matched pairs: $x = (0, 0)^1, (0, 0)^2, (0, 1)^3, (0, 3)^4, (2, 4)^5, (0, 5)^6$. The corresponding total internal tree length is the sum of the lengths of the internal branches: $\varphi(x) = (3 - 1) + (5 - 2) + (4 - 3) + (5 - 4) + (6 - 5) = 8$.

This function φ is useful in bounding the relaxation time because it is ‘local’ with respect to the Markov chain. That is, suppose $P(x, y) > 0$ and $x \neq y$. The change from x to y could have moved an interior node label to the left, in which case $\varphi(y) = \varphi(x) - 1$. If it moved an interior node label to the right, then $\varphi(y) = \varphi(x) + 1$. If two interior node labels were swapped, then $\varphi(y) = \varphi(x)$. Thus $(\varphi(x) - \varphi(y))^2 \leq 1$. The denominator of equation (4) can then be upper-bounded by $1/2$.

To find the variance of the internal tree length, we note that $\text{Var}_\pi[\varphi(X)] = \text{Var}_{\pi^L}[\varphi(X)]$ since the internal tree length is independent of the leaf labels. We now restate the standard coalescent of labeled ranked tree shapes as an urn process.

Consider an urn containing n balls labeled ℓ_1, \dots, ℓ_n . At step k , draw two balls without replacement from the urn. Let a_1, a_2 be the labels of the balls drawn, then set $\mathbf{p}_k = (a_1, a_2)^k$. Add in a new ball with label k . Repeat this process for $k = 1, \dots, n - 1$ until only a single ball labeled $(n - 1)$ remains in the urn. The resulting sequence of pairs $\mathbf{p} = (\mathbf{p}_1, \dots, \mathbf{p}_{n-1})$ corresponds to a labeled ranked tree shape $T \in \mathcal{T}_n^L$ drawn from the standard coalescent, i.e. $\mathbf{p} \sim \pi^L$.

We can now simplify the urn process as follows. Start with n white balls in the urn. At each step, draw two balls and add back in a single red ball (representing an interior node of the tree). Let $R_0 = 0$ and R_k be the number of red balls in the urn after k coalescence events. Note that after k mergers (coalescence events), there are $n - k$ total balls left and $R_{n-1} = 1$.

The simplified urn process is useful because the internal tree length can be computed by counting the number of red balls at each step and adding them all together (Figure 3), that is,

$$\varphi(\mathbf{p}) = \sum_{k=1}^{n-2} R_k.$$

The quantities $\{R_k\}_{k=1}^{n-2}$ are fairly easy to analyze and have been studied before in various contexts. In [15], the values are used to study the asymptotic behavior of the external tree length of the Kingman coalescent and it is shown that

$$\begin{aligned} \mathbb{E}[R_k] &= \frac{k(n-k)}{n-1}, \\ \text{Cov}(R_k \cdot R_l) &= \frac{k(k-1)(n-l)(n-l-1)}{(n-1)^2(n-2)}, \quad k \leq l. \end{aligned}$$

Using this, we get

$$\mathbb{E}_\pi[\varphi(\mathbf{p})] = \sum_{k=1}^{n-2} \frac{k(n-k)}{n-1} = \frac{1}{6}(n^2 + n - 6).$$

Then we calculate

$$\sum_{k=1}^{n-2} \mathbb{E}[R_k^2] = \frac{1}{30}(n^3 + 3n^2 + 2n - 30), \tag{5}$$

$$2 \cdot \sum_{k=1}^{n-3} \sum_{l=k+1}^{n-2} \mathbb{E}[R_k R_l] = \frac{1}{180}(n-3)(5n^3 + 21n^2 - 14n - 120). \tag{6}$$

The second moment $\mathbb{E}_\pi[\varphi(\mathbf{p})^2]$ is the sum of these two lines (5) and (6), which gives

$$\text{Var}_\pi(\varphi(\mathbf{p})) = \mathbb{E}_\pi[\varphi(\mathbf{p})^2] - \mathbb{E}_\pi[\varphi(\mathbf{p})]^2 = \frac{1}{90}n(n+1)(n-3).$$

Theorem 5. Let τ_n, τ_n^L be the relaxation time of the adjacent-swap chains on \mathcal{T}_n and \mathcal{T}_n^L , respectively. Then

$$\tau_n^L \geq \tau_n \geq \frac{\text{Var}_\pi(\varphi)}{\mathcal{E}(\varphi)} \geq \frac{2}{90}n(n+1)(n-3).$$

The lower bound on the relaxation time is applicable to labeled and unlabeled ranked tree shapes, but Lemma 4 further allows us to state the first inequality on the left-hand side of Theorem 5.

5. Upper bound

In this section we prove the upper bounds in Theorem 1 using a coupling argument. The coupling is similar to the one used by Aldous for analyzing the adjacent transposition chain on S_n [1]. The main difference from our approach is that in order to preserve the constraints of the ranked tree shapes, the coupling matches the labels in a specific order. This results in the pairs matching, starting at the root and becoming matched successively from right to left, ending with the first cherry.

We analyze a lazy version of the chain to make the coupling. That is, at each step, we generate a random coin flip $\theta \sim \text{Bernoulli}(1/2)$. If $\theta = 1$, attempt a move of the chain, and if $\theta = 0$ then make no change. Let $X_t = (\mathbf{p}_1, \dots, \mathbf{p}_{n-1})$ and $Y_t = (\mathbf{q}_1, \dots, \mathbf{q}_{n-1})$ be the two copies of the chain at time t , one started from x and the other from y . We will first describe the coupling for the chain on \mathcal{T}_n .

Coupling of unlabeled ranked tree shapes. We will define a coupling that jointly matches the internal node labels of the two copies in the order $n - 2, n - 3, n - 4, \dots, 1$. Note that label $n - 2$ is already jointly matched in the $n - 1$ pairs because it can only occur in the final pair. Let $N_t \in \{1, 2, \dots, n - 2\}$ be the minimum label that is jointly matched, i.e. $N_t - 1$ is the maximum element that occurs in different pairs in X_t and Y_t . Let $X_t(a)$ denote the index of the air in the matching X_t that contains a , i.e. $X_t(a) = i$ if $a \in \mathbf{p}_i$ at time t . Then, at time t , the elements $\geq N_t$ match in both X_t and Y_t , i.e. for every $a \geq N_t$ there is $X_t(a) = Y_t(a)$.

For any label $a \in \{1, 2, \dots, n - 2\}$, the coupling will have two properties.

Property 1. If $a \geq N_t$, then $X_s(a) = Y_s(a)$ for all $s \geq t$.

Property 2. *If $a = N_t - 1$ and $X_t(a) < Y_t(a)$ then $X_{t+1}(a) \leq Y_{t+1}(a)$, and if $X_t(a) > Y_t(a)$ then $X_{t+1}(a) \geq Y_{t+1}(a)$. This condition will ensure that the label ‘ a ’ will eventually get matched in the two copies.*

Define the following quantities.

- Let M_t be the set of all indices $2 \leq i \leq n - 2$ that contain labels $\geq N_t$ jointly matched. That is, there is a label $a \geq N_t$ such that $X_t(a) = Y_t(a) = i$. Note that it is possible to have other matches for labels $< N_t$, but we do not keep track of those matches and breaking those matches does not violate the two properties of the coupling.
- Let $AM_t = i$ if label $N_t - 1$ is in pairs \mathbf{p}_i and \mathbf{q}_{i+1} , or if it is in pairs \mathbf{p}_{i+1} and \mathbf{q}_i . Otherwise, set $AM_t = 0$.

At each step, pick an index $1 \leq i \leq n - 2$ uniformly at random and consider the following cases.

1. $i, i + 1 \notin M_t$ and $i \neq AM_t$. There are no joint matchings of labels $\geq N_t$ in pairs i and $i + 1$ and swapping two elements in both copies will not match label $N_t - 1$. In this case, propose independent swaps in X_t and Y_t according to their (lazy) marginal dynamics.
2. $i \in M_t$ and/or $i + 1 \in M_t$, and $i \neq AM_t$. There is at least one joint matching of labels $\geq N_t$ in pairs i or $i + 1$ and swapping two elements in both copies will not create a new joint matching of $N_t - 1$. To preserve Property 1 it is necessary to perform the *same* move (or no move at all) on both chains. Toss a single coin $\theta \sim \text{Bernoulli}(1/2)$ to determine whether a move will be proposed on both copies or not.

(a) Suppose $\theta = 1$ and the pairs at i and $i + 1$ are of the type

$$\begin{aligned} X: (a, b)^i, (c, d)^{i+1}, \\ Y: (a, e)^i, (f, g)^{i+1}, \end{aligned}$$

with the possibility of $b = e$, and $\{c, d\} \cap \{f, g\} = \emptyset$. The concern here comes since the probability that a specific label in pair i moves to the right depends on whether or not label i occurs in pair $i + 1$. That is, for X , if $c, d \neq i$ then the probability of moving a , given that index i has been chosen, is $1/2$. If $c = i$ or $d = i$ then the probability of moving a is $1/4$.

The coupling procedure, however, ensures that none of c, d, f, g are equal to i : because label a occurs in pair i , it must either represent a leaf or have value less than i ; and $a \geq N_t$ since it is a label that was jointly matched before. Then c, d, f or g cannot be label i since i is already a match. This implies that the labels c, d, f, g can all occur in either pair i or $i + 1$, thus there are no constraints on possible swap moves, and the marginal probability of swapping a would be the same in each chain.

(b) Suppose $\theta = 1$ and the pairs at i and $i + 1$ are of the type

$$\begin{aligned} X: (a, b)^i, (c, d)^{i+1}, \\ Y: (e, f)^i, (g, d)^{i+1}, \end{aligned}$$

with the possibility of $d = i$, or

$$\begin{aligned} X: (a, b)^i, (i, d)^{i+1}, \\ Y: (e, f)^i, (i, d)^{i+1}, \end{aligned}$$

TABLE 1. Coupling transition probabilities in case 4(a).

Move in X	Move in Y	Probability
No change	No change	3/8
No change	$c \leftrightarrow f$	1/8
$b \leftrightarrow c$	No change	1/8
$b \leftrightarrow d$	$c \leftrightarrow g$	1/8
$a \leftrightarrow c$	$e \leftrightarrow f$	1/8
$a \leftrightarrow d$	$e \leftrightarrow g$	1/8

and allowing the possibility $i \in M_t$ in both cases. Then, if a matched label, e.g. d , is chosen to be swapped in chain X , it must also be swapped in Y .

3. $i, i + 1 \notin M_t$ and $i = AM_t$. There are no joint matchings of labels $\geq N_t$ in pairs i and $i + 1$, and label $N_t - 1$ is either in \mathbf{p}_i and \mathbf{q}_{i+1} or in \mathbf{p}_{i+1} and \mathbf{q}_i . In order to preserve Property 2, simply set $\theta_Y = 1 - \theta_X$. This ensures that label $a = N_t - 1$ will only possibly move in one of the chains.
4. $i \in M_t$ and/or $i + 1 \in M_t$, and $i = AM_t$. There is one joint matching in pair i and/or $i + 1$, and label $N_t - 1$ is either in \mathbf{p}_i and \mathbf{q}_{i+1} or in \mathbf{p}_{i+1} and \mathbf{q}_i . To preserve Properties 1 and 2 while keeping the correct marginal transition probabilities for each chain we define the joint transitions as follows.

- (a) Suppose $c = N_t - 1$ and the pairs at indices $i, i + 1$ are of the type

$$\begin{aligned}
 X: & (a, b)^i, (c, d)^{i+1}, \\
 Y: & (e, c)^i, (f, g)^{i+1},
 \end{aligned}$$

with $a = e$ and/or $d = g$. By the same argument as for case 2, $c, d, g, f \neq i$. Table 1 defines the joint proposal probabilities that preserve Properties 1 and 2 with the correct marginal transition probabilities.

Note that this would preserve the matches if $a = e$ and/or $d = g$.

- (b) Suppose now that $c = N_t - 1$ and that pairs at indices i and $i + 1$ are of the type

$$\begin{aligned}
 X: & (a, b)^i, (c, i)^{i+1}, \\
 Y: & (e, c)^i, (f, i)^{i+1},
 \end{aligned}$$

with possibly $a = e$. Table 2 defines the joint proposal probabilities that preserve Properties 1 and 2 with correct marginal transition probabilities.

Coupling of labeled ranked tree shapes. The coupling for \mathcal{T}_n^L can proceed exactly as the coupling for \mathcal{T}_n until every interior node label is matched, say at time T . After that point, the leaf labels can be matched in any order. We extend the set M_t to contain the indices $1 \leq i \leq n - 2$ of any label $a \in \{\ell_1, \dots, \ell_n\}$ such that $X_t(a) = Y_t(a)$ and AM_t to be the set of indices i such that any label $a \in \{\ell_1, \dots, \ell_n\}$ is in pair \mathbf{p}_i and \mathbf{q}_{i+1} or \mathbf{p}_{i+1} and \mathbf{q}_i .

The transition probabilities defined in cases 1–4 above work with these sets M_t, AM_t and have the following properties, for $t \geq T$ and label a .

TABLE 2. Coupling transition probabilities in case 4(b).

Move in X	Move in Y	Probability
No change	No change	5/8
No change	$c \leftrightarrow f$	1/8
$b \leftrightarrow c$	No change	1/8
$a \leftrightarrow c$	$e \leftrightarrow f$	1/8

1. *Property 1.* If $X_t(a) = Y_t(a)$, then $X_s(a) = Y_s(a)$ for all $s \geq t$.
2. *Property 2.* If $X_t(a) < Y_t(a)$, then $X_s(a) \leq Y_s(a)$ for all $s \geq t$. If $X_t(a) > Y_t(a)$, then $X_s(a) \geq Y_s(a)$ for all $s \geq t$.

Note that now in case 4(a) we have the possibility $b = f$, and the transition probabilities defined in Table 1 will work to preserve Property 2.

5.1. Coupling time

For the chain on unlabeled ranked tree shapes \mathcal{T}_n , the time to couple is $T = T_{n-3} + T_{n-4} + \dots + T_1$, where T_a is the time it takes to match interior node label a , after the labels $a + 1, \dots, n - 2$ are already matched. Property 2 is crucial to study T_a . Suppose that at time $t = T_{n-3} + \dots + T_{a+1}$, when label $a + 1$ is matched we have $X_t(a) < Y_t(a)$. Let S_a be the time it takes for Y_t to hit the left boundary of its range, i.e. $Y_t(a) = a + 1$. Then necessarily at this time, by Property 2, $X_t(a) = Y_t(a)$.

Note that $X_t(a) \in \{a + 1, \dots, n - 2\}$. If a is not at the left boundary, the probability of moving a to the left is

$$\mathbb{P}(X_{t+1}(a) = X_t(a) - 1 \mid X_t \neq a + 1) = \frac{1}{n - 2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4(n - 2)}.$$

The probability that a moves to the right will depend on the exact configuration and whether or not there is a constraint, e.g. in the situation

$$(a, b)^i, (c, i)^{i+1},$$

in which the probability a moves to the right is $1/8(n - 2)$. Thus we can bound

$$\frac{1}{8(n - 2)} \leq \mathbb{P}(X_{t+1}(a) = X_t(a) + 1 \mid X_t \neq n - 2) \leq \frac{1}{4(n - 2)}.$$

The following is an easy result about the hitting time of a symmetric random walk on a line.

Lemma 5. Let $(Z_t)_{t \geq 0}$ be a random walk on the line $\{1, 2, 3, \dots, m\}$ with the transitions

$$\begin{aligned} P(i, i + 1) &= p, & i \neq m, \\ P(i, i - 1) &= p, & i \neq 1, \\ P(i, i) &= 1 - 2p, & i \neq 1, m, \\ P(m, m) &= 1 - p, \\ P(1, 1) &= 1, \end{aligned}$$

for some $0 < p \leq 1/2$. Suppose $Z_0 = m$ and let $T_m = \inf\{t > 0: Z_t = 1\}$. Then $\mathbb{E}[T_m] = (1/2p) \cdot m(m - 1)$.

Proof. We can prove the result for $p = 1/2$, then scale. Let a_x be the expected first time of hitting 1 if $Z_0 = x$. Note that the following identities are satisfied:

$$\begin{aligned} a_1 &= 0, \\ a_x &= 1 + \frac{1}{2}(a_{x+1} + a_{x-1}), \quad x \in \{2, 3, \dots, m - 1\}, \\ a_m &= 1 + \frac{1}{2}a_{m-1} + \frac{1}{2}a_m. \end{aligned}$$

This is a second-order recursion, with solution $a_x = x(2m - x + 1) - 2m$. So this means $a_m = m(m - 1)$. When $p < 1/2$ the probability of moving is $2p < 1$, so the expected time to hit 1 started from m is $(1/2p)m(m - 1)$. \square

Let Z_t be a random walk on the line $a + 1, \dots, n - 2$ with $p = 1/4(n - 2)$, defined as in Lemma 5. We can couple Z_t with $Y_t(a)$ so that $Y_t(a) \leq Z_t$ almost surely for all $t \geq 0$, because Z_t is moving to the right with probability larger than the probability that $Y(t)$ moves to the right. In conclusion,

$$\mathbb{E}[T_a] \leq 2(n - 2)(n - 2 - a)^2$$

and thus the total coupling time is

$$\mathbb{E}[\tau_{\text{couple}}] \leq \sum_{a=1}^{n-3} \mathbb{E}[T_a] \leq \sum_{a=1}^{n-3} 2(n - 2)(n - 2 - a)^2 = \frac{1}{3}(n - 2)^2(n - 3)(2n - 5) \leq \frac{2}{3}n^4.$$

This together with Theorem 2 gives the upper bound in Theorem 1 for \mathcal{T}_n :

$$t_n \leq 4 \cdot \mathbb{E}[\tau_{\text{couple}}] \leq \frac{8}{3}n^4.$$

Leaf-labeled trees. The coupling time for leaf-labeled trees can be written $\tau_{\text{couple}}^L = T + T^{\text{leaves}}$, where T is the time from the previous section for the interior labels to match. For the time T^{leaves} for the leaves to match, note that time is saved because the leaves are allowed to match in any order. Moreover, leaf labels can be moved without constraints. In analogy with the result from Example 4.10 in [1] for adjacent transpositions on S_n , this would take time of order $n^3 \log(n)$. Therefore

$$\mathbb{E}[\tau_{\text{couple}}^L] \leq \frac{2}{3}n^4 + C_4n^3 \log(n).$$

6. Discussion

The representation of ranked tree shapes as ordered matchings can be used to define many other Markov chains in analogy to well-studied chains on S_n . For example, in addition to the adjacent-swap chain, another natural chain would be a random-swap Markov chain: pick two pairs uniformly at random, within each pair pick a label, and swap the two elements if it is allowed. The internal tree length function defined in Section 4 gives a lower bound of n for this chain, because a single move could change the function by at most n , and thus the Dirichlet form can be bounded by n^2 . A naive coupling argument would give an upper bound of order

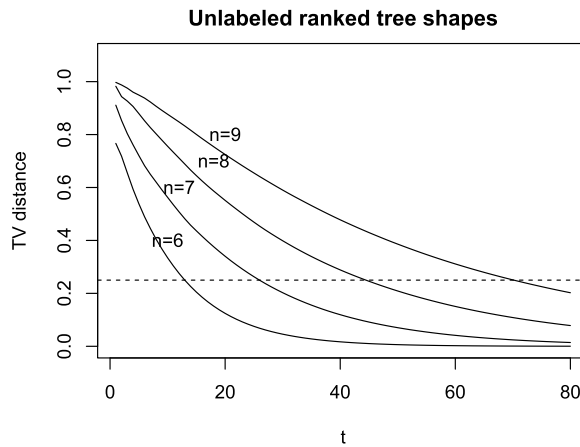


FIGURE 4. Total variation distance for adjacent-swap chain on unlabeled ranked tree shapes. Each curve depicts the total variation distance between the probability of the chain at time t and the Tajima stationary distribution for trees with $n = 6, 7, 8,$ and 9 leaves.

n^3 . The focus of this paper was on the adjacent-swap chain because it is a local-move chain which could be more useful for Metropolis algorithms in applications.

The upper bound on the mixing time relies on a coupling that matches one label at a time in a specific order giving a sub-optimal upper bound. We conjecture that the mixing time of the adjacent-swap chain on unlabeled ranked tree shapes is of the order of $n^3 \log(n)$, as it is in the case of adjacent transpositions on S_n . Figure 4 shows the total variation distance between the adjacent-swap chain on unlabeled ranked tree shapes and the Tajima stationary distribution for trees with $n < 10$ leaves. Due to the large size of the state space, this calculation could not be extended for larger trees in order to inform about the presence of a cut-off phenomenon.

We also note another connection between trees and permutations: Unlabeled ranked tree shapes are in bijection with alternating permutations [9, 30] and hence some results concerning Markov chains on alternating permutations could be used to analyze the mixing of Markov chains on unlabeled ranked trees and vice versa. However, we are not aware of mixing time results of Markov chains on the space of alternating permutations [7].

Finally, in this work we analyzed Markov chains on trees whose stationary distributions are coalescent distributions. These coalescent distributions are used as *prior* distributions over the set of ancestral relationships of samples of DNA [25]. In practice, these Markov chains are used to approximate the posterior distribution, and hence the state space is more restricted in the sense that some trees will not have positive likelihood under some mutation models. Future research is needed to investigate the adjacent-swap Markov chain and its convergence times on these restricted state spaces.

Acknowledgement

We wish to thank the two anonymous referees for helpful comments.

Funding information

MS is supported by a National Defense Science & Engineering Graduate fellowship. JAP is supported by National Institutes of Health Grant R01-GM-131404 and the Alfred P. Sloan Foundation.

Competing interests

There were no competing interests to declare which arose during the preparation or publication process of this article.

References

- [1] ALDOUS, D. (1983). Random walks on finite groups and rapidly mixing Markov chains. In *Séminaire de Probabilités XVII 1981/82*, pp. 243–297. Springer.
- [2] ALDOUS, D. J. (2000). Mixing time for a Markov chain on cladograms. *Combinatorics Prob. Comput.* **9**, 191–204.
- [3] CAPPELLO, L. AND PALACIOS, J. A. (2020). Sequential importance sampling for multiresolution Kingman–Tajima coalescent counting. *Ann. Appl. Statist.* **14**, 727–751.
- [4] CAPPELLO, L., VEBER, A. AND PALACIOS, J. A. (2020). The Tajima heterochronous n -coalescent: Inference from heterochronously sampled molecular data. Available at [arXiv:2004.06826](https://arxiv.org/abs/2004.06826).
- [5] DIACONIS, P. W. AND HOLMES, S. P. (1998). Matchings and phylogenetic trees. *Proc. Nat. Acad. Sci. USA* **95**, 14600–14602.
- [6] DIACONIS, P. AND HOLMES, S. (2002). Random walks on trees and matchings. *Electron. J. Prob.* **7**, 1–17.
- [7] DIACONIS, P. AND WOOD, P. M. (2013). Random doubly stochastic tridiagonal matrices. *Random Structures Algorithms* **42**, 403–437.
- [8] DINH, V., DARLING, A. E. AND MATSEN IV, F. A. (2018). Online Bayesian phylogenetic inference: Theoretical foundations via sequential Monte Carlo. *Syst. Biol.* **67**, 503–517.
- [9] DONAGHEY, R. (1975). Alternating permutations and binary increasing trees. *J. Combinatorial Theory A* **18**, 141–148.
- [10] DRUMMOND, A., SUCHARD, M., XIE, D. AND RAMBAUT, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* **29**, 1969–1973.
- [11] DURRETT, R. (2003). Shuffling chromosomes. *J. Theoret. Prob.* **16**, 725–750.
- [12] FELSENSTEIN, J. (2004). *Inferring Phylogenies*, Vol. 2. Sinauer Associates, Sunderland, MA.
- [13] FRIEDMAN, N., NINIO, M., PE’ER, I. AND PUPKO, T. (2001). A structural EM algorithm for phylogenetic inference. In *Proceedings of the Fifth Annual International Conference on Computational Biology*, pp. 132–140.
- [14] FROST, S. D. W. AND VOLZ, E. M. (2013). Modelling tree shape and structure in viral phylogenetics. *Phil. Trans. R. Soc. B* **368**, 20120208.
- [15] JANSON, S. AND KERSTING, G. (2011). On the total external length of the Kingman coalescent. *Electron. J. Prob.* **16**, 2203–2218.
- [16] KINGMAN, J. (1982). The coalescent. *Stoch. Process. Appl.* **13**, 235–248.
- [17] KIRKPATRICK, M. AND SLATKIN, M. (1993). Searching for evolutionary patterns in the shape of a phylogenetic tree. *Evolution* **47**, 1171–1181.
- [18] KUZNETSOV, A. G., PAK, I. M. AND POSTNIKOV, A. E. (1994). Increasing trees and alternating permutations. *Russian Math. Surveys* **49**, 79.
- [19] LACON, H. (2016). Mixing time and cutoff for the adjacent transposition shuffle and the simple exclusion. *Ann. Prob.* **44**, 1426–1487.
- [20] LEVIN, D. A. AND PERES, Y. (2017). *Markov Chains and Mixing Times*, American Mathematical Society.
- [21] LIU, S., WESTBURY, M. V., DUSSEX, N., MITCHELL, K. J., SINDING, M.-H. S., HEINTZMAN, P. D., DUCHÊNE, D. A., KAPP, J. D., VON SETH, J., HEINIGER, H. et al. (2021). Ancient and modern genomes unravel the evolutionary history of the rhinoceros family. *Cell* **184**, 4874–4885.
- [22] MALIET, O., GASCUEL, F. AND LAMBERT, A. (2018). Ranked tree shapes, nonrandom extinctions, and the loss of phylogenetic diversity. *Syst. Biol.* **67**, 1025–1040.
- [23] MISRA, N., BLELLOCH, G., RAVI, R. AND SCHWARTZ, R. (2011). An optimization-based sampling scheme for phylogenetic trees. In *International Conference on Research in Computational Molecular Biology* (Lecture Notes in Computer Science **6577**), pp. 252–266. Springer, Berlin and Heidelberg.
- [24] MOSSEL, E. AND VIGODA, E. (2006). Limitations of Markov chain Monte Carlo algorithms for Bayesian inference of phylogeny. *Ann. Appl. Prob.* **16**, 2215–2234.
- [25] PALACIOS, J. A., VÉBER, A., CAPPELLO, L., WANG, Z., WAKELEY, J. AND RAMACHANDRAN, S. (2019). Bayesian estimation of population size changes by sampling Tajima’s trees. *Genetics* **213**, 967–986.
- [26] RAJANALA, S. AND PALACIOS, J. A. (2021). Statistical summaries of unlabelled evolutionary trees and ranked hierarchical clustering trees. Available at [arXiv:2106.02724](https://arxiv.org/abs/2106.02724).
- [27] SAINUDIIN, R., STADLER, T. AND VÉBER, A. (2015). Finding the best resolution for the Kingman–Tajima coalescent: Theory and applications. *J. Math. Biology* **70**, 1207–1247.
- [28] SCHWEINSBERG, J. (2002). An $O(n^2)$ bound for the relaxation time of a Markov chain on cladograms. *Random Structures Algorithms* **20**, 59–70.

- [29] SPADE, D., HERBEI, R. AND KUBATKO, L. (2014). A note on the relaxation time of two Markov chains on rooted phylogenetic tree spaces. *Statist. Prob. Lett.* **84**, 247–252.
- [30] STANLEY, R. P. (1999). *Enumerative Combinatorics*, Vol. I. Wadsworth & Brooks/Cole.
- [31] ŠTEFANKOVIČ, D. AND VIGODA, E. (2011). Fast convergence of Markov chain Monte Carlo algorithms for phylogenetic reconstruction with homogeneous data on closely related species. *SIAM J. Discrete Math.* **25**, 1194–1211.
- [32] VOLZ, E. M., KOELLE, K. AND BEDFORD, T. (2013). Viral phylodynamics. *PLoS Comput. Biol.* **9**, e1002947.
- [33] WAKELEY, J. (2008). *Coalescent Theory: An Introduction*. Roberts.
- [34] WILSON, D. B. (2004). Mixing times of lozenge tiling and card shuffling Markov chains. *Ann. Appl. Prob.* **14**, 274–325.