The Annals of Applied Probability
2019, Vol. 29, No. 3, 1808–1836
https://doi.org/10.1214/18-AAP1440

THE NESTED KINGMAN COALESCENT: SPEED OF COMING DOWN FROM INFINITY

BY AIRAM BLANCAS[*,1], TIM ROGERS[†,2], JASON SCHWEINSBERG[‡,3] AND ARNO SIRI-JÉGOUSSE[§,4]

Goethe Universität Frankfurt, University of Bath†, University of California, San Diego‡ and Universidad Nacional Autónoma de México§*

The nested Kingman coalescent describes the ancestral tree of a population undergoing neutral evolution at the level of individuals and at the level of species, simultaneously. We study the speed at which the number of lineages descends from infinity in this hierarchical coalescent process and prove the existence of an early-time phase during which the number of lineages at time t decays as $2\gamma/ct^2$, where c is the ratio of the coalescence rates at the individual and species levels, and the constant $\gamma \approx 3.45$ is derived from a recursive distributional equation for the number of lineages contained within a species at a typical time.

1. Introduction. Kingman's coalescent [15] lies at the centre of modern mathematical population genetics. It is a simple probabilistic model describing the ancestral tree of a population undergoing neutral evolution, which has been shown to apply to a wide variety of population dynamical models [18], and gives rise to the hugely important Ewens sampling formula [11] for the expected genetic variation within a population. Work on Kingman's coalescent and its variants has fueled a wealth of developments in the probability literature, summarised succinctly in [6].

A key result of this theory is that Kingman's coalescent *comes down from infinity*, meaning coalescence occurs so quickly that even when the process is started with an infinite number of lineages, only finitely many survive after any positive time. It is in fact possible to be more precise and state the speed of this descent from infinity. Let $K_n(t)$ denote the number of lineages surviving to time t in the Kingman coalescent initialized on a population of size n. Theorem 1 of [5] (see also [1]) states that taking $n \to \infty$ and then $t \to 0$ we have the almost sure convergence $tK_n(t) \to 2$. Thus, for small times the number of surviving lineages in

Received March 2018.
[1]Supported by CONACyT-MEXICO.
[2]Supported by the Royal Society.
[3]Supported in part by NSF Grants DMS-1206195 and DMS-1707953.
[4]Supported by CONACyT Grant CB-2014/243068.
MSC2010 subject classifications. Primary 60J25; secondary 60J80, 92D15, 92D25.
Key words and phrases. Kingman's coalescent, nested coalescent, gene tree, species tree, coming down from infinity, recursive distributional equation.

the Kingman coalescent decays as $1/t$. This result is important to the population genetics community as it characterizes the expected shape of the lineages through time (LTT) plot [13, 19], a popular technique for analyzing phylogenetic trees reconstructed from genetic data. The speed of descent from infinity has also been studied for coalescents with multiple mergers in [5] and for more general birth and death processes in [3].

From the perspective of applications to genetics, a limitation of Kingman's coalescent is that it describes only the historical coalescence of lineages within a species, and cannot at the same time account for macroevolutionary events occurring between species. The problem of how the gene tree is embedded inside the species tree has been one of the central research questions of population genetics for some time now (see, e.g., [17, 24]), and the issue of how to draw the distinction between intra- and inter-specific genetic variation is an important and contested one [21–23].

In this article, we address this deficit in the theory by computing the speed of descent from infinity in a nested (hierarchical) coalescent process which models both the species tree and the embedded gene tree as a Kingman coalescent, with the latter constrained to be embedded in the former; see Figure 1 for an illustration. We prove that this model exhibits an early-time period in which the number of lineages decays as $1/t^2$; much faster than Kingman's coalescent. This result is potentially important for the environmental metagenomics community, where differentiating between inter- and intra-specific genetic variation is a key step in quantifying biodiversity (see, e.g., [8]). Empirical verification of a $1/t^2$ scaling in

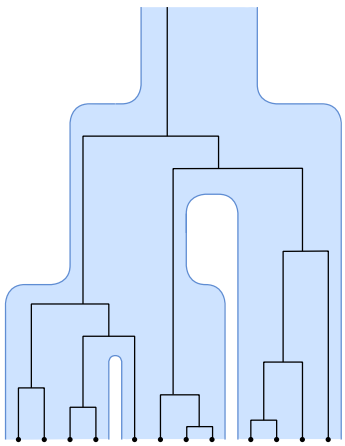

FIG. 1. *Illustration of the nested Kingman coalescent starting with $s = 3$ species and $n = 4$ lineages per species. Black lines show a possible ancestral tree for the sampled individuals, with lineage mergers constrained to lie within the species tree (shown behind in pale blue). The species mergers are described by a Kingman coalescent with rate c, while the within species lineage mergers form a Kingman coalescent with rate* 1.

the LTT plot of an experimentally reconstructed phylogeny would suggest, according to our results, that the gene tree and species tree are evolving on the same time scale, greatly complicating this task.

The article is organised as follows. In the remainder of this section, we give the formal definition of our process (and its population dynamical dual), and state our main theorem. Section 2 develops several results for the standard Kingman coalescent to do with the rate of decrease of the number of lineages, and the asymptotic independence of branches in the ancestral tree. These results are needed for our investigation since, in the nested model, both the species tree and the within-species gene trees (before and between species merger events) are described by Kingman's coalescent. Section 3 brings together the results of Section 2 to deduce a recursion relation between species merger events in the nested coalescent and thence prove our main theorem.

1.1. *Definition of the model.* We consider the following nested coalescent model. We begin with a sample of n individuals from each of s species (including the possibility that one or both of n and s is infinite). Each pair of individuals within a species merges at rate one; also, each pair of species merges at rate $c > 0$. More formally, this process is a continuous-time Markov chain taking its values in the set of labeled partitions of $\{(m, k) \in \mathbb{Z} \times \mathbb{Z} : 1 \leq m \leq n, 1 \leq k \leq s\}$, in which each block of the partition is labeled with one of the integers $1, \ldots, s$. At time zero, the partition consists of ns singleton blocks, and the block (m, k) is labeled by the integer k. Two types of transition are possible:

Lineage mergers Any pair of blocks with the same label may merge into a single block with that label, with rate 1.

Species mergers For any pair of currently surviving labels $i < j$, all blocks with label j have their label changed to i, with rate c.

We refer to this model as the nested Kingman coalescent because, both at the individual and species level, the merging follows the rule of the classical Kingman coalescent [15]. This model has appeared before in the literature in [9]. This model can be alternatively seen as a coalescent process with values in the set of bivariate nested partitions. It is actually an example of simple nested coalescents as defined in [7]. In this reference, a criterion is provided to determine whether nested coalescents come down from infinity or not. However, to our knowledge the speed of descent from infinity has not been computed previously.

The nested Kingman coalescent describes the genealogy in the following population model. Consider a population divided into s species, each composed of N individuals. Within each species, the population evolves according to the classical Moran model [20]. That is, each individual lives for an exponentially distributed time with mean 1; when an individual dies, a new individual is born, and one of the N individuals of the species is chosen at random to be the parent of the new individual. To model the formation of new species, we also suppose that each species

becomes extinct after an exponentially distributed time with rate $c(s-1)/N$, at which time all members of the species simultaneously die. At that time, N new individuals are born, forming a new species. One of the s species is chosen at random, and each member of that species gives birth to one member of the new species. After scaling time by $N/2$, the genealogy of a sample consisting of n individuals from each species converges to the nested Kingman coalescent in the limit as $N \to \infty$ because the large population size ensures that with probability tending to one as $N \to \infty$, the sampled ancestral lines will not merge at the times when new species form. Similar to the standard Kingman coalescent, we expect that the nested Kingman coalescent will also appear as the asymptotic form of various other similar population models under suitable limits. However, this is not the topic of our present study.

1.2. *Main results.* At time t, we write $S(t)$ for the number of species, and $N(t)$ for the total number of blocks (i.e., extant ancestral lines) across all species. Informally, our main result is that, if the initial number of species is large, then there is a period of time during the early evolution of the process in which $N(t)$ decays as $1/t^2$. Since the number of blocks in the standard Kingman coalescent decays as $1/t$, one can understand the $1/t^2$ decay observed in the nested process as a consequence of mergers occurring on both scales (individuals within a species, and whole species mergers) simultaneously.

To state this claim precisely, it is necessary to consider a sequence of processes. For $j \in \mathbb{N}$, consider an instance of the nested Kingman coalescent in which the initial number of species is s_j and the number of individuals sampled from each species is n_j (which, for simplicity, is assumed to be the same for each species). We allow the cases in which $s_j = \infty$ or $n_j = \infty$. Using the notation $a_j \ll b_j$ to mean $\lim_{j\to\infty} a_j/b_j = 0$, $\to_{\mathrm{p}}$ to denote convergence in probability, and $=_{\mathrm{d}}$ to denote equality of distributions, our main result is expressed in the following theorem.

THEOREM 1. *Suppose* $1/s_j \ll t_j \ll 1$, *and* $1/\sqrt{n_j s_j} \ll t_j$. *Then*

$$t_j^2 N(t_j) \to_{\mathrm{p}} \frac{2\gamma}{c} \qquad \text{as } j \to \infty.$$

Here, γ *is the mean of the uniquely determined random variable* W *that takes values in* $[2, \infty)$ *and obeys the recursive distributional equation*

$$W =_{\mathrm{d}} \frac{2}{1 - U(1 - \frac{2}{W_1+W_2})}, \tag{1.1}$$

where U *has a uniform distribution on* $[0, 1]$, W_1 *and* W_2 *have the same distribution as* W, *and the random variables* U, W_1 *and* W_2 *are independent.*

When $s_j \equiv \infty$, Theorem 1 implies that

$$t^2 N(t) \to_{\mathrm{p}} \frac{2\gamma}{c} \qquad \text{as } t \to 0.$$

Therefore, in this case Theorem 1 gives the speed at which $(N(t), t \geq 0)$ descends from infinity. Note that the hypotheses of Theorem 1 require $s_j \to \infty$, but not necessarily that $n_j \to \infty$. For example, the case $n_j \equiv 1$, which corresponds to sampling one individual of each species, is included. When n_j equals some fixed constant n for all j, Theorem 1 implies that, for any fixed $t > 0$ and $\beta \in (0, 1/2)$,

$$\frac{N(ts^{-\beta})}{s^{2\beta}} \to_{\mathrm{p}} \frac{2\gamma}{ct^2} \qquad \text{as } s \to \infty.$$

This scaling can be compared with non-nested models such as Beta-coalescents (see Theorem 4.4 of [10]).

In the case that the initial number of lineages per species vastly exceeds the number of species ($n_j \gg s_j$), the period of $1/t^2$ scaling implied by Theorem 1 is preceded by an earlier phase dominated entirely by within-species coalescence. There the usual $1/t$ scaling is recovered, as we make explicit in the following proposition.

PROPOSITION 2. *Suppose $s_j \to \infty$ and $1/n_j \ll t_j \ll 1/s_j$. Then*

$$\frac{t_j N(t_j)}{s_j} \to_{\mathrm{p}} 2 \qquad \text{as } j \to \infty.$$

The proof of Proposition 2 is given in Section 2.2, and a numerical example showing both $1/t$ and $1/t^2$ phases is shown in Figure 2.

Lambert and Schertzer [16] independently obtained an alternative proof of Theorem 1; see Theorem 5.6 of [16]. Rather than characterizing the random variable W as the unique solution to a recursive distributional equation, they solved the

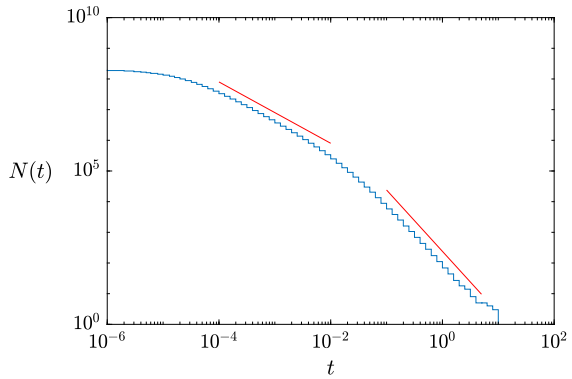

FIG. 2. *Simulation of the nested Kingman coalescent in the case $n \gg s \gg 1$. Here the $1/t^2$ phase is preceded by a period of $1/t$ decay, corresponding to the coalescence occurring within species, but before the species coalescence events kick in. The blue line shows the result of a single simulation with $s = 2000$, $n = 100{,}000$, $c = 0.1$, the red lines indicate slopes of -1 and -2 to illustrate the different scaling regimes.*

problem by showing that the empirical distribution of the number of gene lineages per species evolves like the solution to a certain coagulation-transport equation. They also provided in Theorem 2.8 of [16] a clever alternative characterization of W using the excursion measure of a continuous-state branching process.

1.3. *Heuristics and simulations.* Before presenting our proofs, it is instructive to consider a simple mean-field heuristic for the time-evolution of the process. For the purposes of this discussion, we will focus on the case when $n \gg s \gg 1$. First note that the process $(S(t), t \geq 0)$ has the same law as the number of blocks in Kingman's coalescent (with time scaled by a factor of c). Therefore (following [1]), for small times we can approximate $S(t)$ by the solution to the differential equation

$$\frac{d}{dt}S(t) \approx -\frac{cS(t)^2}{2}, \qquad S(0) = s.$$

It follows that when $t \ll 1$, we have

$$S(t) \approx \frac{2}{ct + \frac{2}{s}}. \tag{1.2}$$

We have $N(t) = N_1(t) + \cdots + N_{S(t)}(t)$, where $N_i(t)$ denotes the number of lineages belonging to the ith of the $S(t)$ species at time t. When $t \ll 1/s$, we see from (1.2) that $S(t) \approx s$, which means very few species mergers have occurred. Within each species, the lineages are merging according to Kingman's coalescent. Therefore, during this period, $N_i(t)$ can be approximated by the solution to the differential equation

$$\frac{d}{dt}N_i(t) \approx -\frac{N_i(t)^2}{2}, \qquad N_i(0) = n.$$

It follows that

$$N_i(t) \approx \frac{2}{t + 2/n} \tag{1.3}$$

for $t \ll 1/s$ and, in particular, $N_i(t) \approx 2/t$ when $1/n \ll t \ll 1/s$. Consequently, we should have $N(t) \approx 2s/t$ when $1/n \ll t \ll 1/s$, which is consistent with Proposition 2.

Note, however, that the number of lineages belonging to a given species will jump upwards when two species merge into one. Consequently, once species mergers start to occur around times of order $1/s$, we can no longer approximate the quantities $N_i(t)$ by solutions to a differential equation. Indeed, these random variables will no longer be well approximated by their expectation, due to the randomness resulting from the timing of the species mergers. Instead, we will argue that

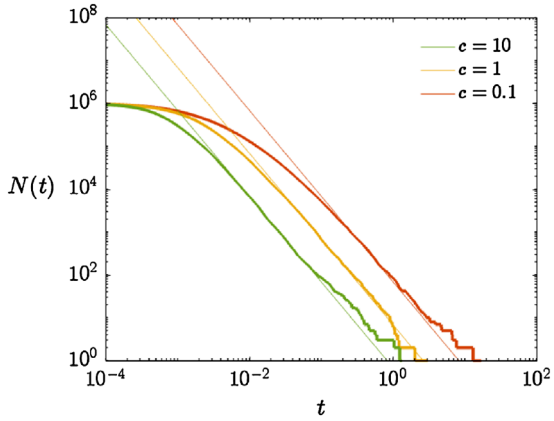

FIG. 3. *Simulations of the nested Kingman coalescent for various values of* c, *compared to the asymptotic form* $N(t) \sim 2\gamma/ct^2$ *given in Theorem* 1. *In each case, the simulation was started with* $s = 1000$ *species each with* $n = 1000$ *members.*

when $1/s \ll t \ll 1$, the distribution of $N_i(t)$ is well approximated by the distribution of W/t, where W satisfies the recursive distributional equation (1.1). The law of large numbers then suggests the approximation

$$N(t) \approx S(t) \cdot \frac{E[W]}{t} \approx \frac{2}{ct} \cdot \frac{\gamma}{t} = \frac{2\gamma}{ct^2}, \tag{1.4}$$

which matches the result of Theorem 1. Therefore, we see the possibility of both $1/t$ and $1/t^2$ behaviour, depending on the parameters. Figure 2 shows an example simulation of the nested Kingman coalescent in which both phases of decay are visible. Figure 3 shows several example simulations for different values of c, compared to the asymptotic result (1.4).

To understand the recursive distributional equation (1.1), we consider choosing at random one of the $S(t)$ species at time t. We then look for the last species merger in the species subtree rooted at this individual at time t. It is well known that this species merger happens at time Ut, where the distribution of U is approximately uniform on $[0, 1]$, as we will explain in more detail in Section 2.3 below. Then, at time Ut, we merge two species with W_1/Ut and W_2/Ut individual lineages, respectively, where W_1 and W_2 are independent and have the same distribution as W. Because the resulting $(W_1 + W_2)/Ut$ lineages then merge as in Kingman's coalescent for the remaining $(1 - U)t$ time, the number of lineages left at time t is given by the right-hand side of (1.3) with $(W_1 + W_2)/Ut$ in place of n and $(1 - U)t$ in place of t. That is, we get the approximation

$$N_i(t) \approx \frac{2}{(1-U)t + \frac{2Ut}{W_1+W_2}} = \frac{1}{t} \cdot \frac{2}{1 - U(1 - \frac{2}{W_1+W_2})}.$$

Writing $N_i(t) \approx W/t$ leads to (1.1).

Straightforward bounds on the constant γ can be obtained based on the conditional expectation

$$E[W|W_1, W_2] = \int_0^1 \frac{2}{1-u(1-2/(W_1+W_2))}\,du = \frac{2}{\frac{2}{W_1+W_2}-1}\log\Big(\frac{2}{W_1+W_2}\Big). \tag{1.5}$$

On the one hand, we know that $W_1, W_2 > 2$, and hence we obtain

$$\gamma > E[W|W_1 = W_2 = 2] = 4\log(2) \approx 2.7726. \tag{1.6}$$

On the other hand, the right-hand-side of (1.5) is a concave function of the sum $W_1 + W_2$, which has expectation 2γ, hence by Jensen's inequality we must have

$$\gamma < \frac{2}{1/\gamma - 1}\log\Big(\frac{1}{\gamma}\Big). \tag{1.7}$$

Solving at equality, we obtain the upper bound

$$\gamma < -2W_{-1}(-1/2\sqrt{e}) \approx 3.5129, \tag{1.8}$$

where W_{-1} denotes the lower branch of the Lambert W function.

We have also simulated from the distribution of W by constructing binary trees of height 12 and using the "recursive tree process" discussed in more detail in Section 2.3 of [2]. Two random variables W_L and W_U were obtained from each run of the procedure. The random variable W_L was obtained by starting with values of 2 at the leaf notes, while W_U was obtained by starting with ∞ at the leaf nodes. The same uniform random variables were used to obtain W_L and W_U, which ensured that $W_L \le W_U$. Furthermore, W stochastically dominates W_L and is stochastically dominated by W_U. This procedure was repeated 10,000,000 times. The values for W_L had a mean of 3.4466, and the values for W_U had a mean of 3.4467. The standard error of these estimates was 0.0009, which means we can be 95 percent confident that $3.4457 < \gamma < 3.4476$.

2. Results on Kingman's coalescent.

2.1. *Estimates on the number of blocks.* Let $(\Pi_\infty(t), t \ge 0)$ be Kingman's coalescent [15], which is a stochastic process taking its values in the set of partitions of $\mathbb{N}$, and let $(\Pi_n(t), t \ge 0)$ be the restriction of $(\Pi_\infty(t), t \ge 0)$ to $\{1, \dots, n\}$. Recall Kingman's coalescent is defined by the property that, for each n, the process $(\Pi_n(t), t \ge 0)$ is a continuous-time Markov chain such that each transition that involves two blocks of the partition merging together happens at rate one, and no other transitions are possible. Let $K_n(t)$ denote the number of blocks of the partition $\Pi_n(t)$, and let $K_\infty(t)$ denote the number of blocks of $\Pi_\infty(t)$. Theorem 1 of [5] (see also [1]) states that

$$\lim_{t\to 0} t K_\infty(t) = 2 \qquad \text{a.s.} \tag{2.1}$$

Theorem 2 of [5] implies that, for all $d \geq 1$,

$$\lim_{t \to 0} E\left[\left|\frac{tK_\infty(t)}{2} - 1\right|^d\right] = 0. \tag{2.2}$$

Our next result provides a first moment estimate for the coalescent started with n blocks.

LEMMA 3. *Let $\delta > 0$. There exists a positive number t_0 and a positive integer M, both depending on δ, such that for all $t \leq t_0$ and $n \geq M$, we have*

$$E\left[\left|K_n(t) - \frac{2}{t+2/n}\right|\right] \leq \frac{\delta}{t+2/n}.$$

PROOF. Let $0 < \varepsilon < 1$. By (2.2) with $d = 1$, there exists $t_1 > 0$, depending on ε, such that if $t \leq t_1$ then

$$E\left[\left|K_\infty(t) - \frac{2}{t}\right|\right] \leq \frac{\varepsilon}{t}. \tag{2.3}$$

Also, (2.1) implies that, for sufficiently large n,

$$P\left(K_\infty\left(\frac{2(1+\varepsilon)}{n}\right) \leq n \leq K_\infty\left(\frac{2(1-\varepsilon)}{n}\right)\right) > 1 - \varepsilon. \tag{2.4}$$

The random variable $K_n(t)$ is stochastically bounded from below by a random variable Y_1, which equals $K_\infty(t+2(1+\varepsilon)/n)$ on the event that $K_\infty(2(1+\varepsilon)/n) \leq n$ and zero otherwise. Then, denoting the positive and negative parts of a random variable X by X^+ and X^- and using (2.3) and (2.4), we get that if $t + 2(1+\varepsilon)/n \leq t_1$ and n is sufficiently large, then

$$\begin{aligned}
E\left[\left(K_n(t) - \frac{2}{t+2/n}\right)^-\right] &\leq E\left[\left(Y_1 - \frac{2}{t+2/n}\right)^-\right] \\
&\leq E\left[\left|K_\infty\left(t + \frac{2(1+\varepsilon)}{n}\right) - \frac{2}{t+2/n}\right|\right] \\
&\quad + \frac{2}{t+2/n} P\left(K_\infty\left(\frac{2(1+\varepsilon)}{n}\right) > n\right) \\
&\leq \frac{\varepsilon}{t+2(1+\varepsilon)/n} + \left(\frac{2}{t+2/n} - \frac{2}{t+2(1+\varepsilon)/n}\right) \\
&\quad + \frac{2\varepsilon}{t+2/n} \\
&\leq \frac{3\varepsilon}{t+2/n} + \frac{2}{(t+2/n)^2} \cdot \frac{2\varepsilon}{n} \\
&\leq \frac{5\varepsilon}{t+2/n}.
\end{aligned}$$

Let $(K'_\infty(t), t \geq 0)$ be an independent copy of the process $(K_\infty(t), t \geq 0)$. The random variable $K_n(t)$ is stochastically bounded from above by a random variable Y_2 that equals $K_\infty(t + 2(1-\varepsilon)/n)$ on the event that $K_\infty(2(1-\varepsilon)/n) \geq n$. On the event that $K_\infty(2(1-\varepsilon)/n) < n$, we set $Y_2 = n$ if $n \leq (2+\varepsilon)/t$ and $Y_2 = K'_\infty(t)$ otherwise. If $t + 2(1-\varepsilon)/n \leq t_1$ so that (2.3) can be applied and n is large enough that (2.4) holds, then using that $\min\{\frac{a}{b}, \frac{c}{d}\} \leq \frac{a+c}{b+d}$ for fractions of positive numbers to get the third inequality, we have

$$\begin{aligned}
&E\left[\left(K_n(t) - \frac{2}{t+2/n}\right)^+\right] \\
&\quad \leq E\left[\left|K_\infty\left(t + \frac{2(1-\varepsilon)}{n}\right) - \frac{2}{t+2/n}\right|\right] \\
&\qquad + \left(n\mathbb{1}_{\{n \leq (2+\varepsilon)/t\}} + E\left[K'_\infty(t)\right]\mathbb{1}_{\{n > (2+\varepsilon)/t\}}\right) P\left(K_\infty\left(\frac{2(1-\varepsilon)}{n}\right) < n\right) \\
&\quad \leq \frac{\varepsilon}{t + 2(1-\varepsilon)/n} + \left(\frac{2}{t+2(1-\varepsilon)/n} - \frac{2}{t+2/n}\right) + \min\left\{n, \frac{2+\varepsilon}{t}\right\}\varepsilon \\
&\quad \leq \frac{\varepsilon}{(1-\varepsilon)(t+2/n)} + \frac{2\varepsilon}{(1-\varepsilon)(t+2/n)} + \frac{2(2+\varepsilon)\varepsilon}{t+2/n} \\
&\quad \leq \left(\frac{3\varepsilon}{1-\varepsilon} + 2(2+\varepsilon)\varepsilon\right)\frac{1}{t+2/n}.
\end{aligned}$$

Combining these results gives that if $t + 2(1+\varepsilon)/n \leq t_1$ and n is sufficiently large, we have

$$E\left[\left|K_n(t) - \frac{2}{t+2/n}\right|\right] \leq \left(5\varepsilon + \frac{3\varepsilon}{1-\varepsilon} + 2(2+\varepsilon)\varepsilon\right)\frac{1}{t+2/n}.$$

The result follows. □

COROLLARY 4. *Let $\delta > 0$, and choose t_0 and M as in Lemma 3. Then for all $\varepsilon \in (0,1)$, $t \leq t_0$, and $n \geq M$ such that $\varepsilon > 2/(nt)$, we have*

$$P\left(K_n(t) < \frac{2(1-\varepsilon)}{t}\right) \leq \frac{\delta}{\varepsilon - 2/(nt)}.$$

PROOF. By Lemma 3 and Markov's inequality,

$$\begin{aligned}
P\left(K_n(t) < \frac{2(1-\varepsilon)}{t}\right) &= P\left(K_n(t) < \frac{2}{t+2/n} \cdot (1-\varepsilon)\left(1 + \frac{2}{nt}\right)\right) \\
&\leq P\left(\left|K_n(t) - \frac{2}{t+2/n}\right| > \frac{2(1-(1-\varepsilon)(1+2/(nt)))}{t+2/n}\right)
\end{aligned}$$

$$\leq \frac{\delta}{2(1-(1-\varepsilon)(1+2/(nt)))}$$
$$\leq \frac{\delta}{\varepsilon - 2/(nt)},$$

as claimed. □

2.2. *Proof of Proposition 2.*

PROOF OF PROPOSITION 2. We obtain upper and lower bounds on $N(t_j)$ by comparing our process to simpler coalescent processes. For the upper bound, let $N^+(t)$ denote the number of individuals remaining at time t in a model that is the same as our model, except that all species mergers are suppressed. Suppressing species mergers can only reduce the number of mergers of individual lineages, so $N^+(t)$ stochastically dominates $N(t)$ for all t. Let $N_i^+(t)$ be the number of individual lineages at time t belonging to species i, in this new model. Let $\varepsilon > 0$. Then, using Markov's inequality,

$$P\left(N^+(t_j) > \frac{(2+\varepsilon)s_j}{t_j}\right) \leq \frac{t_j}{\varepsilon s_j} E\left[\left|N^+(t_j) - \frac{2s_j}{t_j}\right|\right]$$
$$\leq \frac{t_j}{\varepsilon s_j} \sum_{i=1}^{s_j} E\left[\left|N_i^+(t_j) - \frac{2}{t_j}\right|\right]. \tag{2.5}$$

Using Lemma 3 and the assumption that $1/n_j \ll t_j$, we get that for all $i \in \{1, \ldots, s_j\}$ and all $\delta > 0$,

$$\limsup_{j\to\infty} t_j E\left[\left|N_i^+(t_j) - \frac{2}{t_j}\right|\right] \leq \limsup_{j\to\infty} t_j \left(\left|\frac{2}{t_j} - \frac{2}{t_j + 2/n_j}\right| + \frac{\delta}{t_j + 2/n_j}\right)$$
$$= \delta. \tag{2.6}$$

Combining (2.5) and (2.6) yields

$$\lim_{j\to\infty} P\left(N^+(t_j) > \frac{(2+\varepsilon)s_j}{t_j}\right) = 0. \tag{2.7}$$

For the lower bound, recall that at time zero, blocks of the partition are labeled by the integers $1, \ldots, s_j$, corresponding to the s_j species. When the two species corresponding to the labels i and j merge, where $i < j$, individuals of both species take the label i. Let $N^-(t_j)$ denote the number of individual lineages at time t_j whose species label has not changed between times 0 and t_j. That is, we keep only the individuals from one of the original species corresponding to each of the $S(t_j)$ species at time t_j. Clearly, $N^-(t_j) \leq N(t_j)$. Conditional on $S(t_j) = s$, the distribution of $N^-(t_j)$ is the same as the distribution of what we get by running s independent copies of Kingman's coalescent, each started with n_j lineages, and

counting the total number of lineages remaining at time t_j. Therefore, the same reasoning that leads to (2.7) gives

$$\lim_{j\to\infty} P\left(N^-(t_j) < \frac{(2-\varepsilon)s}{t_j}\Big| S(t_j) = s\right) = 0, \tag{2.8}$$

and the convergence is uniform in s. However, because $t_j \ll 1/s_j$, another application of Lemma 3 yields

$$\lim_{j\to\infty} P\big(S(t_j) \le (1-\varepsilon)s_j\big) = 0.$$

Combining this result with (2.8) yields

$$\lim_{j\to\infty} P\left(N^-(t_j) < \frac{(2-\varepsilon)(1-\varepsilon)s_j}{t_j}\right) = 0. \tag{2.9}$$

The proposition follows from (2.7) and (2.9). □

2.3. *Kingman's coalescent and time-changed Yule trees.* We now define the coalescent process that describes the species mergers. Let $(\Psi_\infty(t), t \ge 0)$ be a coalescent process having the same law as $(\Pi_\infty(ct), t \ge 0)$. That is, $(\Psi_\infty(t), t \ge 0)$ has the same law as Kingman's coalescent, except that pairs of blocks merge at rate c rather than at rate 1. For $s \in \mathbb{N}$, let $(\Psi_s(t), t \ge 0)$ denote the restriction of $(\Psi_\infty(t), t \ge 0)$ to $\{1, \ldots, s\}$. Let $S_\infty(t)$ be the number of blocks in the partition $\Psi_\infty(t)$, and let $S_s(t)$ denote the number of blocks in the partition $\Psi_s(t)$. We interpret $S_\infty(t)$ as the number of species remaining at time t when we start with infinitely many species at time zero, and $S_s(t)$ as the number of species remaining at time t when we start with s species at time zero. Note that the coalescent process $(\Psi_\infty(t), t \ge 0)$ can also be depicted as a tree $\mathcal{T}$ with infinitely many leaves at height zero and $S_\infty(t)$ branches at height t. The leaves can be labeled by the positive integers.

For positive integers m, let $\tau_m = \inf\{t : S_\infty(t) = m\}$. If we consider the portion of the tree $\mathcal{T}$ below height τ_{m-1}, we have m subtrees, which we place in random order and denote by $\mathcal{T}^{1,m}, \ldots, \mathcal{T}^{m,m}$. One of these trees is pictured in Figure 4.

For $k \in \{1, \ldots, m\}$, $\ell \in \{0, 1, 2, \ldots\}$, and $i_1, \ldots, i_\ell \in \{1, 2\}$, we will define random variables $U^{k,m}_{i_1\ldots i_\ell}$ and $V^{k,m}_{i_1\ldots i_\ell}$ as follows. We begin at time τ_{m-1} and follow the tree $\mathcal{T}^{k,m}$ in reversed time from time τ_{m-1} down to time 0, so that branches split instead of coalescing. Define $V^{k,m}$ to be the time when the initial branch splits into two. Then define $V^{k,m}_1$ and $V^{k,m}_2$ to be the times when the two branches created at time $V^{k,m}$, ordered at random, split again. Given $V^{k,m}_{i_1\ldots i_l}$, let $V^{k,m}_{i_1\ldots i_l 1}$ and $V^{k,m}_{i_1\ldots i_l 2}$ denote the times when the two branches created at time $V^{k,m}_{i_1\ldots i_l}$ split into two. Let $U^{k,m} = V^{k,m}/\tau_{m-1}$, and for $\ell \ge 1$, define $U^{k,m}_{i_1\ldots i_\ell} = V^{k,m}_{i_1\ldots i_\ell}/V^{k,m}_{i_1\ldots i_{\ell-1}}$. Then

$$V^{k,m}_{i_1\ldots i_\ell} = U^{k,m}U^{k,m}_{i_1}U^{k,m}_{i_1 i_2}\ldots U^{k,m}_{i_1\ldots i_\ell}\tau_{m-1}. \tag{2.10}$$

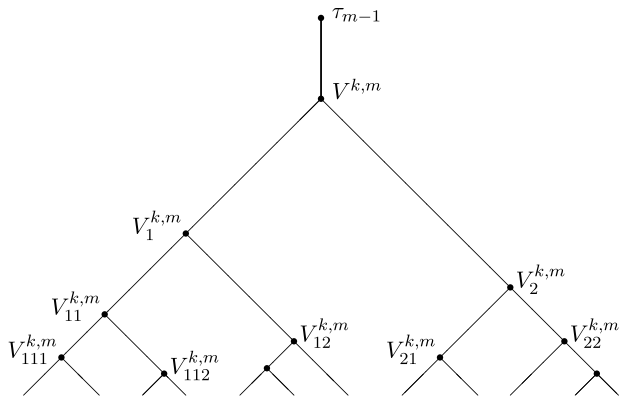

FIG. 4. *The tree* $\mathcal{T}^{k,m}$.

A key ingredient in our proof is that the random variables $U_{i_1 \ldots i_\ell}^{k,m}$ are approximately independent, and have approximately a uniform distribution on [0, 1]. Making this statement rigorous involves coupling the coalescent with a time-changed Yule process. This connection between Kingman's coalescent and a Yule process was discussed in [4], in which both Kingman's coalescent and a Yule process are shown to be embedded in a Brownian excursion.

Consider a Yule process $(Y(t), t \geq 0)$, which is a continuous-time branching process in which there are no deaths and each individual independently gives birth at rate 1. Consider the time-change which maps t to $u = 1 - e^{-t}$, so that $t = -\log(1-u)$. It is well known that for all $u \in [0, 1)$, the next time that an individual at time u gives birth is uniformly distributed on $[u, 1]$. To see this, note that the probability that an individual at time u gives birth before time $u + x(1-u)$ is the same as the probability that an individual in the original Yule process at time $-\log(1-u)$ gives birth before time $-\log(1-u-x(1-u))$, which is

$$1 - e^{\log(1-u-x(1-u))-\log(1-u)} = 1 - \frac{1-u-x(1-u)}{1-u} = x.$$

We can then do the time-reversal $v = 1 - u = e^{-t}$, so $t = -\log v$. After this additional time change, we start at time 1, and individuals branch as we go backwards in time. An individual at time v will branch next at a time which is uniformly distributed on $[0, v]$, and individuals reproduce independently.

Fix a positive integer m. We now obtain a Yule process started with m individuals by starting with Kingman's coalescent and then performing a random time change.

LEMMA 5. *For* $0 < t \leq \tau_{m-1}$, *let*

$$f_m(t) = \log\left(\frac{mc}{2}\right) + \int_t^{\tau_{m-1}} \frac{c(S_\infty(r) - 1)}{2}\, dr.$$

Then $(S_\infty(f_m^{-1}(u)), u > \log(mc/2))$ *is a Yule process started with* m *individuals at time* $\log(mc/2)$.

PROOF. First, note that $f_m(t)$ is a strictly decreasing function of t, so the inverse function is well defined. Now let $k \geq m$, and note that $\tau_{k-1} - \tau_k$, which is the amount of time for which there are k species, has an exponential distribution with rate $ck(k-1)/2$. Because the time change stretches time by a factor of $c(k-1)/2$ during this interval, the distribution of $f_m^{-1}(\tau_k) - f_m^{-1}(\tau_{k-1})$ is exponential with rate k, matching the distribution of the amount of time for which there are k individuals in a Yule process. □

We can now make the further time change discussed in the paragraph before Lemma 5, and define $S^*(v) = S_\infty(f_m^{-1}(-\log v))$ for $0 < v < 2/cm$. Just as there is a coalescent tree $\mathcal{T}$, with subtrees $\mathcal{T}^{1,m}, \ldots, \mathcal{T}^{m,m}$, associated with the original coalescent process $(S_\infty(t), t \geq 0)$, there are m subtrees $\mathcal{T}^{1,m,*}, \ldots, \mathcal{T}^{m,m,*}$ associated with the process $(S^*(v), 0 < v < 2/cm)$, and we can use these trees to define associated random variables $U^{k,m,*}_{i_1 \ldots i_l}$ and $V^{k,m,*}_{i_1 \ldots i_l}$ as before. Furthermore, because $(S^*(v), 0 < v < 2/cm)$ arises by time-changing a Yule process, it follows from the discussion above that the new random variables $U^{k,m,*}_{i_1,\ldots,i_l}$ are independent, and each has exactly the uniform distribution on [0, 1].

For $0 < t < \tau_{m-1}$, we have $S_\infty(t) = S^*(e^{-f_m(t)})$. Lemmas 6 and 7 below establish that this time change is only a small perturbation of time.

LEMMA 6. *We have*

$$\sup_{0<t\leq\tau_{m-1}} \left| \frac{t}{e^{-f_m(t)}} - 1 \right| \to_{\mathrm{p}} 0,$$

where $\to_{\mathrm{p}}$ *denotes convergence in probability as* $m \to \infty$.

PROOF. Taking logarithms, it suffices to show that as $m \to \infty$,

$$\sup_{0<t\leq\tau_{m-1}} \left|\log t + f_m(t)\right| \to_{\mathrm{p}} 0. \tag{2.11}$$

From (2.1), we have $tS_\infty(t) \to 2/c$ almost surely as $t \to 0$. It follows that $\tau_k \sim 2/ck$ almost surely, where $\sim$ means that the ratio of the two sides tends to one as $k \to \infty$. Taking logarithms,

$$\lim_{k\to\infty} \left| \log \tau_k - \log\left(\frac{2}{ck}\right) \right| = 0 \qquad \text{a.s.} \tag{2.12}$$

For $k \geq m-1$, let

$$H_k = \int_{\tau_k}^{\tau_{m-1}} \frac{c(S_\infty(r) - 1)}{2}\, dr = \frac{c}{2} \sum_{j=m}^{k} (\tau_{j-1} - \tau_j)(j-1).$$

Because $\tau_{j-1} - \tau_j$ has an exponential distribution with rate parameter $cj(j-1)/2$, and these random variables are independent for different values of j, we have

$$E[H_k] = \sum_{j=m}^{k} \frac{1}{j}$$

and

$$\operatorname{Var}(H_k) = \sum_{j=m}^{k} \frac{1}{j^2}.$$

By Kolmogorov's maximal inequality applied to the independent mean zero random variables $H_k - H_{k-1} - 1/k$, we have for all $\varepsilon > 0$,

$$P\left(\sup_{k \geq m} \left| H_k - \sum_{j=m}^{k} \frac{1}{j} \right| > \varepsilon \right) \leq \frac{1}{\varepsilon^2} \sum_{j=m}^{\infty} \frac{1}{j^2} \leq \frac{1}{(m-1)\varepsilon^2}. \tag{2.13}$$

Now suppose $\tau_k \leq t < \tau_{k-1}$ for some $k > m$. Then

$$\begin{aligned} \log t + f_m(t) &\leq \log \tau_{k-1} + \log\left(\frac{mc}{2}\right) + H_k \\ &\leq \log\left(\frac{2}{c(k-1)}\right) + \left| \log \tau_{k-1} - \log\left(\frac{2}{c(k-1)}\right) \right| \\ &\quad + \log\left(\frac{mc}{2}\right) + \sum_{j=m}^{k} \frac{1}{j} + \left| H_k - \sum_{j=m}^{k} \frac{1}{j} \right| \\ &= \sum_{j=m}^{k} \frac{1}{j} - \log\left(\frac{k-1}{m}\right) + \left| \log \tau_{k-1} - \log\left(\frac{2}{c(k-1)}\right) \right| \\ &\quad + \left| H_k - \sum_{j=m}^{k} \frac{1}{j} \right|, \end{aligned} \tag{2.14}$$

and likewise

$$\begin{aligned} \log t + f_m(t) &\geq \log \tau_k + \log\left(\frac{mc}{2}\right) + H_{k-1} \\ &\geq \sum_{j=m}^{k-1} \frac{1}{j} - \log\left(\frac{k}{m}\right) - \left| \log \tau_k - \log\left(\frac{2}{ck}\right) \right| - \left| H_{k-1} - \sum_{j=m}^{k-1} \frac{1}{j} \right|. \end{aligned} \tag{2.15}$$

From (2.14) and (2.15), combined with the bounds (2.12) and (2.13) and standard estimates for the harmonic series, we obtain (2.11). □

LEMMA 7. *We have*

$$\sup_{0<u<v\le\tau_{m-1}}\left|\frac{e^{-f_m(u)}-e^{-f_m(v)}}{u-v}-1\right|\to_{\mathrm{p}}0,$$

where $\to_{\mathrm{p}}$ *denotes convergence in probability as* $m\to\infty$.

PROOF. Let $g(t)=e^{-f_m(t)}$. We have

$$\sup_{0<u<v\le\tau_{m-1}}\left|\frac{e^{-f_m(u)}-e^{-f_m(v)}}{u-v}-1\right|\le\sup_{0<t<\tau_m}|g'(t)-1|.$$

Also,

$$g'(t)=-e^{-f_m(t)}f_m'(t)=\frac{e^{-f_m(t)}}{t}\cdot\frac{ct(S_\infty(t)-1)}{2}.$$

Equation (2.1) implies that $ct(S_\infty(t)-1)/2\to 1$ almost surely as $t\to 0$, and also that $\tau_m\to 0$ almost surely as $m\to\infty$. Combining these results with Lemma 6, we see that as $m\to\infty$,

$$\sup_{0<t<\tau_m}|g'(t)-1|\to_{\mathrm{p}}0,$$

which implies the result. □

3. Results on the nested coalescent.

3.1. *Convergence to a unique solution of the RDE.* Let $\mathcal{P}$ denote the set of probability distributions on $[2,\infty]$, and let $\mathcal{P}_1$ denote the set of probability distributions on $[2,\infty]$ with finite mean. Let $T:\mathcal{P}\to\mathcal{P}$ be the mapping defined such that $T(\mu)$ is the distribution of

$$\frac{2}{1-U(1-\frac{2}{W_1+W_2})},\tag{3.1}$$

where U has a uniform distribution on $[0,1]$, the random variables W_1 and W_2 have distribution μ, and the random variables U, W_1 and W_2 are independent. Let $T^n:\mathcal{P}\to\mathcal{P}$ be the map obtained by iterating n times the map T. Our goal in this subsection is to prove the following result.

PROPOSITION 8. *The equation* $T(\mu)=\mu$ *has a unique solution* μ^*, *and* $\mu^*\in\mathcal{P}_1$. *For all* $\mu\in\mathcal{P}$, *the sequence* $T^n(\mu)$ *converges to* μ^* *in the sense of weak convergence of probability measures on* $[2,\infty]$. *Also, the mean of* $T^n(\mu)$ *converges as* $n\to\infty$ *to the mean of* μ^*.

For $u \in (0, 1)$ and $x \in (0, \infty]$, define

$$h(u, x) = \frac{2}{1 - u(1 - \frac{2}{x})} = \frac{2}{(1 - u) + \frac{2u}{x}}. \tag{3.2}$$

LEMMA 9. *We have $T(\mathcal{P}_1) \subset \mathcal{P}_1$, and $T^2(\mathcal{P}) \subset \mathcal{P}_1$.*

PROOF. Let $\mu \in \mathcal{P}_1$. Let U, W_1 and W_2 be independent random variables such that U has a uniform distribution on $[0, 1]$, and W_1 and W_2 have distribution μ. Then $T(\mu)$ has the same distribution as $h(U, W_1 + W_2)$, and a stochastic upper bound can be obtained by removing one of the two terms from the denominator on the right-hand side of (3.2). Therefore,

$$\begin{aligned} E\big[h(U, W_1 + W_2)\big] &\le E\left[\frac{2}{1 - U}\mathbb{1}_{\{U \le 1/2\}}\right] + E\left[\frac{W_1 + W_2}{U}\mathbb{1}_{\{U \ge 1/2\}}\right] \\ &\le 4 + 2E[W_1 + W_2]. \end{aligned}$$

It follows that $T(\mu) \in \mathcal{P}_1$, which proves the first statement of the lemma.

Let δ_a denote the unit mass at a. Because the expression in (3.1) is an increasing function of W_1 and W_2, if we can show that $T^2(\delta_\infty) \in \mathcal{P}_1$, then it will follow that $T^2(\mu) \in \mathcal{P}_1$ for all $\mu \in \mathcal{P}$, which will establish the second part of the lemma. Note that $T(\delta_\infty)$ has the same distribution as $2/(1 - U)$, which has the same distribution as $2/U$. Therefore, $T^2(\delta_\infty)$ has the same distribution as

$$Y = \frac{2}{1 - U(1 - \frac{2}{2/U_1 + 2/U_2})} = \frac{2}{(1 - U) + \frac{U}{1/U_1 + 1/U_2}},$$

where U, U_1 and U_2 are independent random variables, each having the uniform distribution on $[0, 1]$. Thus, it suffices to show that $E[Y] < \infty$. We have

$$Y \le \min\left\{\frac{2}{1 - U}, \frac{2}{U}\left(\frac{1}{U_1} + \frac{1}{U_2}\right)\right\}.$$

Let $x \ge 4$. If $Y \ge x$, then we must have $2/(1 - U) \ge x$ and, therefore, $U > 1 - 2/x$. We also must have $(2/U)(1/U_1 + 1/U_2) \ge x$. When $U > 1 - 2/x \ge 1/2$, this can only happen if $1/U_1 + 1/U_2 \ge x/4$, which requires either $U_1 \le 8/x$ or $U_2 \le 8/x$. Thus,

$$\begin{aligned} P(Y \ge x) &\le P(U_1 \ge 1 - 2/x \text{ and either } U_1 \le 8/x \text{ or } U_2 \le 8/x) \\ &\le \frac{2}{x}\left(\frac{8}{x} + \frac{8}{x}\right) = \frac{32}{x^2}. \end{aligned}$$

It follows that

$$E[Y] = \int_0^\infty P(Y \ge x)\, dx \le 4 + \int_4^\infty \frac{32}{x^2}\, dx < \infty,$$

which completes the proof. □

Let d denote the Kantorovich–Rubinstein metric on $\mathcal{P}_1$, which goes back to [14] and is also the Wasserstein metric for $p = 1$. That is,

$$d(\mu, \nu) = \inf\{E[|X - Y|] : X \text{ has distribution } \mu \text{ and } Y \text{ has distribution } \nu\}.$$

It is well known that d is a complete metric on $\mathcal{P}_1$ (see, e.g., [12]). Because $2(1 - \log 2) < 1$, the following lemma shows that, with respect to this metric, T is a strict contraction.

LEMMA 10. *Suppose $\mu, \nu \in \mathcal{P}_1$ and $\mu \neq \nu$. Then*

$$d(T(\mu), T(\nu)) \leq 2(1 - \log 2)d(\mu, \nu).$$

PROOF. Let $\varepsilon > 0$. Let $\mu, \nu \in \mathcal{P}_1$. By the definition of d, on some probability space one can construct random variables X_1 and Y_1 such that X_1 has distribution μ, Y_1 has distribution ν and $E[|X_1 - Y_1|] \leq d(\mu, \nu) + \varepsilon$. One can construct X_2 and Y_2, independently of (X_1, Y_1), so that they satisfy these same conditions. Let U be a random variable that has a uniform distribution on $(0, 1)$ and is independent of (X_1, X_2, Y_1, Y_2). Let $X = h(U, X_1 + X_2)$ and $Y = h(U, Y_1 + Y_2)$, where h is the function defined in (3.2). Note that X has the same distribution as $T(\mu)$, and Y has the same distribution as $T(\nu)$. For $x \geq 4$,

$$\frac{\partial h}{\partial x}(u, x) = \frac{4u}{(x(1-u) + 2u)^2} \leq \frac{4u}{(4(1-u) + 2u)^2} = \frac{u}{4(1 - u/2)^2}.$$

Therefore,

$$|X - Y| \leq \frac{U}{4(1 - U/2)^2} |(X_1 + X_2) - (Y_1 + Y_2)|.$$

Taking expectations, we get

$$E[|X - Y|] \leq \frac{d(\mu, \nu) + \varepsilon}{2} E\left[\frac{U}{(1 - U/2)^2}\right] = \frac{d(\mu, \nu) + \varepsilon}{2} \cdot 4(1 - \log 2).$$

Letting $\varepsilon \to 0$ gives $E[|X - Y|] \leq 2(1 - \log 2)d(\mu, \nu)$, which implies the result. □

PROOF OF PROPOSITION 8. Because $T^2(\mu) \in \mathcal{P}_1$ for all $\mu \in \mathcal{P}$ (by Lemma 9), any solution to the equation $T(\mu) = \mu$ must be in $\mathcal{P}_1$. By Lemma 10, the map T is a strict contraction with respect to the Kantorovich–Rubinstein metric on $\mathcal{P}_1$. Therefore, as noted in Lemma 5 of [2], it follows from the Banach contraction theorem that the equation $T(\mu) = \mu$ has a unique solution μ^*, and $T^n(\mu)$ converges to μ^* as $n \to \infty$ with respect to the Kantorovich–Rubinstein metric for all $\mu \in \mathcal{P}_1$. Because convergence with respect to the Kantorovich–Rubinstein metric implies both weak convergence and convergence of means (see, e.g., [12]), the result follows. □

3.2. *Mergers of individual ancestral lines.* We now consider the merging of individual ancestral lines within a species. Recall that, at time zero, there are s_j species, and we sample n_j individuals from each of the s_j species. Pairs of ancestral lines belonging to the same species merge at rate one.

Recall the definition of the trees $\mathcal{T}^{1,m}, \ldots, \mathcal{T}^{m,m}$ derived from the species tree in Section 2.3. Let $N^{k,m}$ be the number of individual lineages remaining at time $\tau_{m-1}-$ that belong to the species represented by the tree $\mathcal{T}^{k,m}$. Note that this number could be zero when s_j is finite because $\mathcal{T}^{k,m}$ is derived from a species tree starting from infinitely many species, whereas we only sample n_j lineages from s_j of these species. Let $N^{k,m}_{i_1\ldots i_\ell}$ be the number of individual lineages, belonging to the species created by the merger at time $V^{k,m}_{i_1\ldots i_\ell}$, that remain at time $V^{k,m}_{i_1,\ldots,i_{\ell-1}}-$. If we know the values of $N^{k,m}_{i_1\ldots i_\ell}$ for all $i_1, \ldots, i_\ell \in \{1,2\}$, then we obtain $N^{k,m}_{i_1\ldots i_{\ell-1}}$ by starting with $N^{k,m}_{i_1\ldots i_{\ell-1}1} + N^{k,m}_{i_1\ldots i_{\ell-1}2}$ lineages and running Kingman's coalescent for time $V^{k,m}_{i_1\ldots i_{\ell-2}} - V^{k,m}_{i_1\ldots i_{\ell-1}}$. Also, let $W^{k,m}_{i_1\ldots i_\ell} = V^{k,m}_{i_1\ldots i_{\ell-1}} N^{k,m}_{i_1\ldots i_\ell}$ for $\ell \geq 1$, and let $W^{k,m} = \tau_{m-1} N^{k,m}$.

Fix a positive integer d. Let $W^{k,m,*,+}_{i_1\ldots i_d} = \infty$, and let $W^{k,m,*,-}_{i_1\ldots i_d} = 2$. Let $W^{k,m,*}_{i_1\ldots i_d}$ and $\bar{W}^{k,m}_{i_1\ldots i_d}$ both equal $\max\{W^{k,m}_{i_1\ldots i_d}, 2\}$. Recall the definition of the function h from (3.2). For $0 \leq \ell \leq d-1$ and $i_1, \ldots, i_\ell \in \{1,2\}$, let

$$
\begin{aligned}
W^{k,m,*}_{i_1\ldots i_\ell} &= h\big(U^{k,m,*}_{i_1\ldots i_\ell}, W^{k,m,*}_{i_1\ldots i_\ell 1} + W^{k,m,*}_{i_1\ldots i_\ell 2}\big),\\
W^{k,m,*,+}_{i_1\ldots i_\ell} &= h\big(U^{k,m,*}_{i_1\ldots i_\ell}, W^{k,m,*,+}_{i_1\ldots i_\ell 1} + W^{k,m,*,+}_{i_1\ldots i_\ell 2}\big),\\
W^{k,m,*,-}_{i_1\ldots i_\ell} &= h\big(U^{k,m,*}_{i_1\ldots i_\ell}, W^{k,m,*,-}_{i_1\ldots i_\ell 1} + W^{k,m,*,-}_{i_1\ldots i_\ell 2}\big),\\
\bar{W}^{k,m}_{i_1\ldots i_\ell} &= h\big(U^{k,m}_{i_1\ldots i_\ell}, \bar{W}^{k,m}_{i_1\ldots i_\ell 1} + \bar{W}^{k,m}_{i_1\ldots i_\ell 2}\big).
\end{aligned}
$$

Because the random variables $U^{k,m,*}_{i_1\ldots i_\ell}$ are independent and have a uniform distribution on $[0,1]$, the distribution of $W^{k,m,*,+}_{i_1\ldots i_{d-1}}$ is $T(\delta_\infty)$, while the distribution of $W^{k,m,-}_{i_1\ldots i_{d-1}}$ is $T(\delta_2)$. More generally, for $0 \leq \ell \leq d$, the distributions of $W^{k,m,*,+}_{i_1\ldots i_\ell}$ and $W^{k,m,*,-}_{i_1\ldots i_\ell}$ are $T^{d-\ell}(\delta_\infty)$ and $T^{d-\ell}(\delta_2)$, respectively. In particular, the distributions of $W^{k,m,*,+}$ and $W^{k,m,*,-}$ are $T^d(\delta_\infty)$ and $T^d(\delta_2)$, respectively. Also, because $h(u,x)$ is an increasing function of x, we have

$$
W^{k,m,*,-} \leq W^{k,m,*} \leq W^{k,m,*,+}. \tag{3.3}
$$

To prove Theorem 1, we will consider a sequence $(m_j)_{j=1}^\infty$ tending to infinity. That is, for the process in which there are s_j species and n_j individuals sampled from each of these species, we will consider the trees $\mathcal{T}^{k,m_j}$. Throughout the rest of this section, we will occasionally drop the superscripts k and m_j to lighten notation, when doing so seems unlikely to cause confusion.

LEMMA 11. *We have*

$$\sup_{1\le k\le m_j}\left|\frac{\bar{W}^{k,m_j}}{W^{k,m_j,*}}-1\right|\to_{\mathrm{p}}0,$$

where $\to_{\mathrm{p}}$ *denotes convergence in probability as* $j\to\infty$.

PROOF. Recall the definition of the function h from (3.2). Note that

$$\frac{h(u_1,x_1)}{h(u_2,x_2)}=\frac{(1-u_2)+2u_2/x_2}{(1-u_1)+2u_1/x_1}.$$

Therefore, using that $\min\{\frac{a}{b},\frac{c}{d}\}\le\frac{a+c}{b+d}\le\max\{\frac{a}{b},\frac{c}{d}\}$,

$$\begin{aligned}\frac{h(u_1,x_1)}{h(u_2,x_2)}&\le\max\left\{\frac{1-u_2}{1-u_1},\frac{u_2x_1}{u_1x_2}\right\}\\&\le\max\left\{\frac{1-u_2}{1-u_1},\left(\frac{u_2}{u_1}\right)^2,\left(\frac{x_1}{x_2}\right)^2\right\}\\&\le\max\left\{\frac{1-u_2}{1-u_1},\frac{u_2}{u_1},\frac{x_1}{x_2}\right\}^2\end{aligned}$$

and

$$\frac{h(u_1,x_1)}{h(u_2,x_2)}\ge\min\left\{\frac{1-u_2}{1-u_1},\frac{u_2}{u_1},\frac{x_1}{x_2}\right\}^2.$$

Recall that $W^*_{i_1\ldots i_d}=\bar{W}_{i_1\ldots i_d}$, and for $0\le\ell\le d-1$, we have

$$\frac{\bar{W}_{i_1\ldots i_\ell}}{W^*_{i_1\ldots i_\ell}}=\frac{h(U_{i_1\ldots i_\ell},\bar{W}_{i_1\ldots i_\ell1}+\bar{W}_{i_1\ldots i_\ell2})}{h(U^*_{i_1\ldots i_\ell},W^*_{i_1\ldots i_\ell1}+W^*_{i_1\ldots i_\ell2})}.\tag{3.4}$$

Recall also that, defining the random function f_m as in Lemma 5 and defining $V^*_{i_1\ldots i_\ell}$ as in the discussion following that lemma, we have $V^*_{i_1\ldots i_\ell}=e^{-f_{m_j}(V_{i_1\ldots i_\ell})}$. For $1\le\ell\le d-1$, we have

$$\frac{U^*_{i_1\ldots i_\ell}}{U_{i_1\ldots i_\ell}}=\frac{V^*_{i_1\ldots i_\ell}}{V^*_{i_1\ldots i_{\ell-1}}}\cdot\frac{V_{i_1\ldots i_{\ell-1}}}{V_{i_1\ldots i_\ell}}=\frac{e^{-f_{m_j}(V_{i_1\ldots i_\ell})}}{V_{i_1\ldots i_\ell}}\cdot\frac{V_{i_1\ldots i_{\ell-1}}}{e^{-f_{m_j}(V_{i_1\ldots i_{\ell-1}})}}\tag{3.5}$$

and

$$\begin{aligned}\frac{1-U^*_{i_1\ldots i_\ell}}{1-U_{i_1\ldots i_\ell}}&=\frac{(V^*_{i_1\ldots i_{\ell-1}}-V^*_{i_1\ldots i_\ell})V_{i_1\ldots i_{\ell-1}}}{V^*_{i_1\ldots i_{\ell-1}}(V_{i_1\ldots i_{\ell-1}}-V_{i_1\ldots i_\ell})}\\&=\frac{e^{-f_{m_j}(V_{i_1\ldots i_{\ell-1}})}-e^{-f_{m_j}(V_{i_1\ldots i_\ell})}}{V_{i_1\ldots i_{\ell-1}}-V_{i_1\ldots i_\ell}}\cdot\frac{V_{i_1\ldots i_{\ell-1}}}{e^{-f_{m_j}(V_{i_1\ldots i_{\ell-1}})}}.\end{aligned}$$

Likewise, for the $\ell = 0$ case,

$$\frac{U^*}{U} = \frac{e^{-f_{m_j}(V)}}{V} \cdot \frac{\tau_{m_j-1}}{e^{-f_{m_j}(\tau_{m_j-1})}}$$

and

$$\frac{1-U^*}{1-U} = \frac{e^{-f_{m_j}(\tau_{m_j-1})} - e^{-f_{m_j}(V)}}{\tau_{m_j-1} - V} \cdot \frac{\tau_{m_j-1}}{e^{-f_{m_j}(\tau_{m_j-1})}}.$$

Let

$$R = \max\left\{ \sup_{0<t\le\tau_{m_j-1}} \left| \frac{t}{e^{-f_{m_j}(t)}} - 1 \right|, \sup_{0<u<v\le\tau_{m_j-1}} \left| \frac{e^{-f_{m_j}(u)} - e^{-f_{m_j}(v)}}{u-v} - 1 \right| \right\},$$

which converges in probability to zero as $j \to \infty$ by Lemmas 6 and 7. Using the fact that if $|a-1| \le R$ then a and $1/a$ are both between $1-R$ and $1/(1-R)$, it follows from these results with (3.4), we obtain

$$(1-R)^4 \le \frac{\bar{W}_{i_1\ldots i_{d-1}}}{W^*_{i_1\ldots i_{d-1}}} \le \left(\frac{1}{1-R}\right)^4.$$

Then by induction, we end up with

$$(1-R)^{2^{d+1}} \le \frac{\bar{W}^{k,m_j}}{W^{k,m_j,*}} \le \left(\frac{1}{1-R}\right)^{2^{d+1}}$$

for all $k \in \{1,\ldots,m_j\}$. Because d is a fixed positive integer, the result follows. □

LEMMA 12. *Suppose $s_j \gg m_j$. Let $\varepsilon > 0$. For $i_1,\ldots,i_d \in \{1,2\}$ and $k \in \{1,\ldots,m_j\}$, let $R^{k,m_j}_{i_1\ldots i_d}$ be the number of species, among the s_j present at time zero, that are descended from the species created by the merger at time $V^{k,m_j}_{i_1\ldots i_d}$. Then there exists $\delta > 0$ such that for sufficiently large j, we have*

$$P\left(R^{k,m_j}_{i_1\ldots i_d} > \frac{\delta s_j}{m_j}\right) > 1-\varepsilon.$$

PROOF. For all $t \ge 0$, the partition given by Kingman's coalescent at time t, $\Pi_\infty(t)$, is an exchangeable random partition of $\mathbb{N}$. Therefore, if B is a block of the partition $\Pi_\infty(t)$, then the limit

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{i\in B\}}$$

exists and is called the asymptotic frequency of B. Let $K_\infty(t)$ be the number of blocks of $\Pi_\infty(t)$, and let $\tau_m = \inf\{t : K_\infty(t) = m\}$ be the first time that the coalescent has m blocks. Denote by $\Lambda(t)$ the sequence consisting of the asymptotic

frequencies of the blocks of $\Pi_\infty(t)$, ranked in decreasing order. It is shown in [15] that the distribution of $\Lambda(\tau_m)$ is uniform on the simplex

$$\Delta_m = \{x_1 \geq \cdots \geq x_m \geq 0 : x_1 + \cdots + x_m = 1\}.$$

In particular, if we choose one of the m blocks uniformly at random, the distribution of the asymptotic frequency of this block is Beta$(1, m-1)$. Furthermore, if we follow Kingman's coalescent in reversed time, so that blocks split instead of merging, and B is a block with asymptotic frequency λ, then immediately after this block splits into two, the new blocks will have asymptotic frequencies λU and $\lambda(1-U)$, where U has a uniform distribution on $[0, 1]$.

By the discussion above, the asymptotic frequency of the block of $\Psi_\infty(\tau_{m_j})$ corresponding to the species represented by the tree $\mathcal{T}^{k,m_j}$ has the Beta$(1, m_j - 1)$ distribution. Moreover, let $\Lambda_{i_1\ldots i_d}$ be the asymptotic frequency of the block of $\Psi_\infty(V_{i_1\ldots i_d})$ created by the merger at time $V_{i_1\ldots i_d}$. Then the distribution of $\Lambda_{i_1\ldots i_d}$ is the same the distribution of the product of $d+1$ independent random variables, one of them having the Beta$(1, m_j - 1)$ distribution and d of them having the Uniform$(0, 1)$ distribution. Because d is a fixed positive integer, it follows that there exists $\delta > 0$ such that for all j, we have

$$P\left(\Lambda_{i_1\ldots i_d} > \frac{2\delta}{m_j}\right) > 1 - \frac{\varepsilon}{2}.$$

Conditional on $\Lambda_{i_1\ldots i_d}$, the distribution of $R_{i_1\ldots i_d}$ is Binomial$(s_j, \Lambda_{i_1\ldots i_d})$. Because $s_j \gg m_j$, the result now follows from elementary concentration results for the binomial distribution. □

LEMMA 13. *Let*

$$L^{k,m_j} = \max_{i_1,\ldots,i_d \in \{1,2\}} \left| \frac{W^{k,m_j}_{i_1\ldots i_d}}{\bar{W}^{k,m_j}_{i_1\ldots i_d}} - 1 \right|. \tag{3.6}$$

Suppose $s_j \gg m_j$ and $s_j n_j \gg m_j^2$. Then

$$\lim_{j\to\infty} E\big[L^{k,m_j} W^{k,m_j,*,+}\big] = 0.$$

PROOF. Note that $W_{i_1\ldots i_d} = \bar{W}_{i_1\ldots i_d}$ unless $W_{i_1\ldots i_d} < 2$. Therefore, $|L^{k,m_j}| \leq 1$, and $L^{k,m_j} = 0$ unless $W_{i_1\ldots i_d} < 2$ for some $i_1, \ldots, i_d$. Because the distribution of $W^{k,m_j,*,+}$ is exactly $T^d(\delta_\infty)$ for all k and m_j, the collection of random variables $\{L^{k,m_j} W^{k,m_j,*,+} : j \in \mathbb{N}, 1 \leq k \leq m_j\}$ is uniformly integrable. Therefore, noting also that the distribution of $L^{k,m_j} W^{k,m_j,*,+}$ does not depend on k, it suffices to show that $L^{k,m_j} W^{k,m_j,*,+} \to_{\mathrm{p}} 0$ as $m \to \infty$. Because the random variables $W^{k,m_j,*,+}$ are identically distributed and finite, it suffices to show that $L^{k,m_j} \to_{\mathrm{p}} 0$ as $j \to \infty$.

Let $\varepsilon > 0$. We have

$$\begin{aligned} P\big(L^{k,m_j} > \varepsilon\big) &\le 2^d P\left(\left|\frac{W_{i_1\ldots i_d}}{\bar{W}_{i_1\ldots i_d}} - 1\right| > \varepsilon\right) \\ &= 2^d P\big(W_{i_1\ldots i_d} < 2(1-\varepsilon)\big) \\ &= 2^d P\big(V_{i_1\ldots i_{d-1}} N_{i_1\ldots i_d} < 2(1-\varepsilon)\big). \end{aligned} \tag{3.7}$$

Recall the definition of $R_{i_1\ldots i_d}$ from Lemma 12. Note that there are $R_{i_1\ldots i_d} n_j$ individual lineages at time zero descended from the species created by the merger at time $V_{i_1\ldots i_d}$. Pairs of these individual lineages are subject to mergers at rate one, once the corresponding species lineages have merged, which means we can obtain a stochastic lower bound on the number of individual lineages by allowing all pairs of these lineages to merge at rate one. Therefore, a stochastic lower bound for $N_{i_1\ldots i_d}$ can be obtained first constructing the species tree and then running the block-counting process associated with Kingman's coalescent, started with $R_{i_1\ldots i_d} n_j$ lineages, for time $V_{i_1\ldots i_{d-1}}$. In particular, denoting by $\mathcal{G}$ the σ-field generated by the process $(\Psi_\infty(t), t \ge 0)$ that governs the species mergers, we have

$$P\big(V_{i_1\ldots i_{d-1}} N_{i_1\ldots i_d} < 2(1-\varepsilon)|\mathcal{G}\big) \le P\big(V_{i_1\ldots i_{d-1}} K_{R_{i_1\ldots i_d} n_j}(V_{i_1\ldots i_{d-1}}) < 2(1-\varepsilon)|\mathcal{G}\big).$$

Now let $\delta = \varepsilon^2$ and apply Corollary 4 with $V_{i_1\ldots i_{d-1}}$ in place of t and $R_{i_1\ldots i_d} n_i$ in place of n to get

$$P\big(V_{i_1\ldots i_{d-1}} K_{R_{i_1\ldots i_d} n_j}(V_{i_1\ldots i_{d-1}}) < 2(1-\varepsilon)|\mathcal{G}\big) \le \frac{\varepsilon^2}{\varepsilon - 2/(V_{i_1\ldots i_{d-1}} R_{i_1\ldots i_d} n_i)} \tag{3.8}$$

on the event that $V_{i_1\ldots i_{d-1}} < t_0$, $R_{i_1\ldots i_d} n_j \ge M$, and $2/(V_{i_1\ldots i_{d-1}} R_{i_1\ldots i_d} n_j) < \varepsilon$. Note that $P(V_{i_1\ldots i_{d-1}} < t_0) \ge P(\tau_{m_j - 1} < t_0) \to 1$ as $j \to \infty$ by (2.1). Therefore, the result that $L^{k,m_j} \to_{\mathrm{p}} 0$ and, therefore, the result of the lemma, will follow from (3.7) and (3.8) provided we can show that

$$V_{i_1\ldots i_{d-1}} R_{i_1\ldots i_d} n_j \to_{\mathrm{p}} \infty \qquad \text{as } j \to \infty.$$

Recall from equation (2.10) that $V_{i_1\ldots i_{d-1}} = U U_{i_1} U_{i_1 i_2} \ldots U_{i_1\ldots i_{d-1}} \tau_{m_j - 1}$. It follows from (2.1) that $m_j \tau_{m_j - 1} \to 2/c$ almost surely as $j \to \infty$. Combining this observation with (3.5) and Lemma 6, we see that there is a constant $\delta_1 > 0$ such that $P(V_{i_1\ldots i_{d-1}} > \delta_1/m_j) > 1 - \varepsilon/2$ for sufficiently large j. By Lemma 12 and the assumption that $s_j \gg m_j$, there is a constant $\delta_2 > 0$ such that $P(R_{i_1\ldots i_d} > \delta_2 s_j/m_j) > 1 - \varepsilon/2$ for sufficiently large j. Combining these results, we get

$$P\left(V_{i_1\ldots i_{d-1}} R_{i_1\ldots i_d} n_j > \frac{\delta_1 \delta_2 s_j n_j}{m_j^2}\right) > 1 - \varepsilon$$

for sufficiently large j. Because $s_j n_j \gg m_j^2$ by assumption, the result follows. □

LEMMA 14. *Let $\varepsilon > 0$. Suppose $s_j \gg m_j$ and $s_j n_j \gg m_j^2$. There is a positive constant t_0, depending on ε, such that if we define the events*

$$A_j = \{\tau_{m_j - 1} \le t_0\}, \qquad B_j = \left\{ \sup_{1 \le k \le m_j} \left| \frac{\bar{W}^{k,m_j}}{W^{k,m_j,*}} - 1 \right| \le \frac{1}{2} \right\},$$

then for sufficiently large j, we have

$$E\big[\big|W^{k,m_j} - \bar{W}^{k,m_j}\big| \mathbb{1}_{A_j \cap B_j}\big] < \varepsilon.$$

PROOF. By Proposition 8, we can choose a positive integer d large enough that the mean of the distribution $T^d(\delta_\infty)$ is less than $4\gamma/3$. Choose $0 < \delta < 1$ small enough that $(\delta/2)\sum_{n=0}^{d-1}(2+\delta)^n < \varepsilon/2$ and $6^d \gamma \delta < \varepsilon/4$. Choose a positive integer M and then choose $t_0 < \delta/M$ such that if $t \le t_0$ and $n \ge M$, then the conclusion of Lemma 3 holds for this choice of δ.

Suppose $0 \le \ell \le d-1$. Then, dropping the superscripts k and m_j to lighten notation,

$$\left| \frac{W_{i_1 \ldots i_\ell}}{\bar{W}_{i_1 \ldots i_\ell}} - 1 \right| = \left| \frac{W_{i_1 \ldots i_\ell}}{h(U_{i_1 \ldots i_\ell}, W_{i_1 \ldots i_\ell 1} + W_{i_1 \ldots i_\ell 2})} \cdot \frac{h(U_{i_1 \ldots i_\ell}, W_{i_1 \ldots i_\ell 1} + W_{i_1 \ldots i_\ell 2})}{h(U_{i_1 \ldots i_\ell}, \bar{W}_{i_1 \ldots i_\ell 1} + \bar{W}_{i_1 \ldots i_\ell 2})} - 1 \right|.$$

Recall the definition of the function h from (3.2). Because

$$|xy - 1| = \big|x - 1 + x(y-1)\big| \le |x-1| + x|y-1|$$

for positive real numbers x and y, we have

$$\begin{aligned} &\left| \frac{W_{i_1 \ldots i_\ell}}{\bar{W}_{i_1 \ldots i_\ell}} - 1 \right| \\ &\quad \le \left| \frac{W_{i_1 \ldots i_\ell}}{h(U_{i_1 \ldots i_\ell}, W_{i_1 \ldots i_\ell 1} + W_{i_1 \ldots i_\ell 2})} - 1 \right| \\ &\qquad + \frac{W_{i_1 \ldots i_\ell}}{h(U_{i_1 \ldots i_\ell}, W_{i_1 \ldots i_\ell 1} + W_{i_1 \ldots i_\ell 2})} \left| \frac{h(U_{i_1 \ldots i_\ell}, W_{i_1 \ldots i_\ell 1} + W_{i_1 \ldots i_\ell 2})}{h(U_{i_1 \ldots i_\ell}, \bar{W}_{i_1 \ldots i_\ell 1} + \bar{W}_{i_1 \ldots i_\ell 2})} - 1 \right|. \end{aligned}$$

If $0 \le u \le 1$ and $x_1, x_2, y_1, y_2 > 0$, then

$$\left| \frac{h(u, x_1 + x_2)}{h(u, y_1 + y_2)} - 1 \right| = \left| \frac{(1-u) + \frac{2u}{y_1 + y_2}}{(1-u) + \frac{2u}{x_1 + x_2}} - 1 \right| \le \left| \frac{x_1 + x_2}{y_1 + y_2} - 1 \right| \le \left| \frac{x_1}{y_1} - 1 \right| + \left| \frac{x_2}{y_2} - 1 \right|.$$

Therefore,

$$\begin{aligned} &\left| \frac{W_{i_1 \ldots i_\ell}}{\bar{W}_{i_1 \ldots i_\ell}} - 1 \right| \\ &\quad \le \left| \frac{W_{i_1 \ldots i_\ell}}{h(U_{i_1 \ldots i_\ell}, W_{i_1 \ldots i_\ell 1} + W_{i_1 \ldots i_\ell 2})} - 1 \right| \\ &\qquad + \frac{W_{i_1 \ldots i_\ell}}{h(U_{i_1 \ldots i_\ell}, W_{i_1 \ldots i_\ell 1} + W_{i_1 \ldots i_\ell 2})} \left(\left| \frac{W_{i_1 \ldots i_\ell 1}}{\bar{W}_{i_1 \ldots i_\ell 1}} - 1 \right| + \left| \frac{W_{i_1 \ldots i_\ell 2}}{\bar{W}_{i_1 \ldots i_\ell 2}} - 1 \right| \right). \end{aligned} \tag{3.9}$$

Interpreting $V_{i_1\dots i_{\ell-1}}$ to be τ_{m_j-1} when $\ell=0$, we have

$$\begin{aligned}&\left|\frac{W_{i_1\dots i_\ell}}{h(U_{i_1\dots i_\ell}, W_{i_1\dots i_\ell 1}+W_{i_1\dots i_\ell 2})}-1\right|\\&\quad=\left|\frac{N_{i_1\dots i_\ell}V_{i_1\dots i_{\ell-1}}}{h(V_{i_1\dots i_\ell}/V_{i_1\dots i_{\ell-1}}, V_{i_1\dots i_\ell}(N_{i_1\dots i_\ell 1}+N_{i_1\dots i_\ell 2}))}-1\right|\\&\quad=\left|N_{i_1\dots i_\ell}\left(\frac{(V_{i_1\dots i_{\ell-1}}-V_{i_1\dots i_\ell})+2/(N_{i_1\dots i_\ell 1}+N_{i_1\dots i_\ell 2})}{2}\right)-1\right|.\end{aligned}$$

Recall that $N_{i_1\dots i_\ell}$ is obtained by running Kingman's coalescent started with $N_{i_1\dots i_\ell 1}+N_{i_1\dots i_\ell 2}$ blocks for time $V_{i_1\dots i_{\ell-1}}-V_{i_1\dots i_\ell}$. Now let $\mathcal{F}_{\ell,j}$ denote the σ-field generated by the process $(\Psi_\infty(t), t\ge 0)$ and the random variables $W_{i_1\dots i_b}$ with $\ell+1\le b\le d$ and $i_1\dots i_b\in\{1,2\}$. By Lemma 3,

$$E\left[\left|\frac{W_{i_1\dots i_\ell}}{h(U_{i_1\dots i_\ell}, W_{i_1\dots i_\ell 1}+W_{i_1\dots i_\ell 2})}-1\right|\mathbb{1}_{A_j}\middle|\mathcal{F}_{\ell,j}\right]\le\frac{\delta}{2}$$

on the event $D_{i_1\dots i_\ell}=\{N_{i_1\dots i_\ell 1}+N_{i_1\dots i_\ell 2}\ge M\}$. Combining this result with (3.9) yields that on $D_{i_1\dots i_\ell}$,

$$\begin{aligned}&E\left[\left|\frac{W_{i_1\dots i_\ell}}{\bar{W}_{i_1\dots i_\ell}}-1\right|\mathbb{1}_{A_j}\middle|\mathcal{F}_{\ell,j}\right]\\&\quad\le\frac{\delta}{2}+\left(1+\frac{\delta}{2}\right)\left(\left|\frac{W_{i_1\dots i_\ell 1}}{\bar{W}_{i_1\dots i_\ell 1}}-1\right|+\left|\frac{W_{i_1\dots i_\ell 2}}{\bar{W}_{i_1\dots i_\ell 2}}-1\right|\right)\mathbb{1}_{A_j}.\end{aligned}\tag{3.10}$$

Now suppose $D^c_{i_1\dots i_\ell}$ occurs. Then $W_{i_1\dots i_\ell}=V_{i_1\dots i_\ell}N_{i_1\dots i_\ell}\le V_{i_1\dots i_\ell}(N_{i_1\dots i_\ell 1}+N_{i_1\dots i_\ell 2})\le\tau_{m_j-1}M$. Because $\bar{W}_{i_1\dots i_\ell}\ge 2$ by construction in view of the definition of the function h, it follows that $W_{i_1\dots i_\ell}/\bar{W}_{i_1\dots i_\ell}\le\tau_{m_j-1}M/2\le Mt_0/2\le\delta/2$ on A_j. By the same reasoning, if $N_{i_1\dots i_\ell 1}<M$, we have $W_{i_1\dots i_\ell 1}/\bar{W}_{i_1\dots i_\ell 1}\le\delta/2$ on A_j, and likewise if $N_{i_1\dots i_\ell 2}<M$. Thus, on the event $D^c_{i_1\dots i_\ell}$, the left-hand side of (3.10) is bounded above by 1, while the right-hand side is bounded below by $\delta/2+(1+\delta/2)(1-\delta/2)\ge 1$ on A_j. Therefore, (3.10) also holds on $D^c_{i_1,\dots,i_\ell}$. Now taking conditional expectations with respect to $\mathcal{F}_{d-1,j}$ on both sides of (3.10), we get

$$E\left[\left|\frac{W_{i_1\dots i_\ell}}{\bar{W}_{i_1\dots i_\ell}}-1\right|\mathbb{1}_{A_j}\middle|\mathcal{F}_{d-1,j}\right]\le\frac{\delta}{2}+(2+\delta)E\left[\left|\frac{W_{i_1\dots i_\ell 1}}{\bar{W}_{i_1\dots i_\ell 1}}-1\right|\mathbb{1}_{A_j}\middle|\mathcal{F}_{d-1,j}\right].\tag{3.11}$$

Recall the definition of L^{k,m_j} from (3.6). Note that L^{k,m_j} is $\mathcal{F}_{d-1,j}$-measurable and $A_j\in\mathcal{F}_{d-1,j}$. Therefore, when $\ell=d-1$, equation (3.11) implies

$$E\left[\left|\frac{W_{i_1\dots i_{d-1}}}{\bar{W}_{i_1\dots i_{d-1}}}-1\right|\mathbb{1}_{A_j}\middle|\mathcal{F}_{d-1,j}\right]\le\left(\frac{\delta}{2}+(2+\delta)L^{k,m_j}\right)\mathbb{1}_{A_j}.$$

Applying (3.11) inductively as ℓ goes from $d-2$ down to 0 gives

$$E\left[\left|\frac{W^{k,m_j}}{\bar{W}^{k,m_j}}-1\right|\mathbb{1}_{A_j}\middle|\mathcal{F}_{d-1,j}\right]\le\left(\frac{\delta}{2}\sum_{n=0}^{d-1}(2+\delta)^n+(2+\delta)^dL^{k,m_j}\right)\mathbb{1}_{A_j}$$
$$\le\left(\frac{\varepsilon}{2}+3^dL^{k,m_j}\right)\mathbb{1}_{A_j}.$$

Because $\bar{W}^{k,m_j}$ is $\mathcal{F}_{d-1,j}$-measurable and $B_j\in\mathcal{F}_{d-1,j}$, we can multiply both sides by $\bar{W}^{k,m_j}\mathbb{1}_{B_j}$ to get

$$E\big[|W^{k,m_j}-\bar{W}^{k,m_j}|\mathbb{1}_{A_j\cap B_j}|\mathcal{F}_{d-1,j}\big]\le\left(\frac{\varepsilon}{2}+3^dL^{k,m_j}\right)\bar{W}^{k,m_j}\mathbb{1}_{A_j\cap B_j}$$
$$\le\frac{3}{2}\left(\frac{\varepsilon}{2}+3^dL^{k,m_j}\right)W^{k,m_j,*,+}.$$

Taking expectations of both sides, we get

$$E\big[|W^{k,m_j}-\bar{W}^{k,m_j}|\mathbb{1}_{A_j\cap B_j}\big]\le\frac{3\varepsilon}{4}+\frac{3^{d+1}}{2}E\big[W^{k,m_j,*,+}L^{k,m_j}\big].$$

The result now follows from Lemma 13. □

LEMMA 15. *Define μ^* as in Proposition 8, and let $\gamma=\int_2^\infty x\mu^*(dx)$ be the mean of μ^*. Suppose $s_j\gg m_j$ and $s_jn_j\gg m_j^2$. Then*

$$\frac{1}{m_j}\sum_{k=1}^{m_j}W^{k,m_j}\to_{\mathrm{p}}\gamma,$$

where $\to_{\mathrm{p}}$ denotes convergence in probability as $j\to\infty$.

PROOF. Let $\eta>0$, and let $\varepsilon=\eta^2$. By Proposition 8, we can choose a positive integer d sufficiently large that the mean of $T^d(\delta_2)$ is greater than $\gamma-\eta$, and the mean of $T^d(\delta_\infty)$ is less than $\gamma+\eta$. Because the random variables $W^{k,m_j,*,+}$ are independent of one another and have the distribution $T^d(\delta_\infty)$, and the random variables $W^{k,m_j,*,-}$ are independent of one another and have distribution $T^d(\delta_2)$, it follows from the law of large numbers and the assumption that $\lim_{j\to\infty}m_j=\infty$ that

$$\lim_{j\to\infty}P\left(\gamma-\eta<\frac{1}{m_j}\sum_{k=1}^{m_j}W^{k,m_j,*,-}\le\frac{1}{m_j}\sum_{k=1}^{m_j}W^{k,m_j,*,+}<\gamma+\eta\right)=1.$$

Therefore, by (3.3),

$$\lim_{j\to\infty}P\left(\gamma-\eta<\frac{1}{m_j}\sum_{k=1}^{m_j}W^{k,m_j,*}<\gamma+\eta\right)=1.$$

It now follows from Lemma 11 that

$$\lim_{j\to\infty} P\left(\gamma-\eta<\frac{1}{m_j}\sum_{k=1}^{m_j}\bar{W}^{k,m_j}<\gamma+\eta\right)=1. \tag{3.12}$$

Define t_0 and the events A_j and B_j as in Lemma 14. We have

$$\begin{aligned} &P\left(\left|\frac{1}{m_j}\sum_{k=1}^{m_j}W^{k,m_j}-\gamma\right|\geq 2\eta\right)\\ &\leq P(A_j^c)+P(B_j^c)+P\left(\left|\frac{1}{m_j}\sum_{k=1}^{m_j}\bar{W}^{k,m_j}-\gamma\right|\geq\eta\right)\\ &\quad+P\left(\left|\frac{1}{m_j}\sum_{k=1}^{m_j}W^{k,m_j}-\frac{1}{m_j}\sum_{k=1}^{m_j}\bar{W}^{k,m_j}\right|\mathbb{1}_{A_j\cap B_j}\geq\eta\right). \end{aligned} \tag{3.13}$$

Note that $\lim_{j\to\infty}P(A_j^c)=0$ by (2.1) and $\lim_{j\to\infty}P(B_j^c)=0$ by Lemma 11. The third term on the right-hand side of (3.13) tends to zero as $j\to\infty$ by (3.12). By Lemma 14 and Markov's inequality, for sufficiently large j, we have

$$\begin{aligned} &P\left(\left|\frac{1}{m_j}\sum_{k=1}^{m_j}W^{k,m_j}-\frac{1}{m_j}\sum_{k=1}^{m_j}\bar{W}^{k,m_j}\right|\mathbb{1}_{A_j\cap B_j}\geq\eta\right)\\ &\leq\frac{1}{\eta m_j}\sum_{k=1}^{m_j}E\left[\left|W^{k,m_j}-\bar{W}^{k,m_j}\right|\mathbb{1}_{A_j\cap B_j}\right]\leq\frac{\varepsilon}{\eta}=\eta. \end{aligned}$$

Because $\eta>0$ is arbitrary, the result follows. □

3.3. *Proof of Theorem 1.*

PROOF OF THEOREM 1. For positive integers j, let $m_j^+=2(1+\varepsilon)/ct_j$ and $m_j^-=2(1-\varepsilon)/ct_j$. It follows from (2.1) that $m\tau_m\to 2/c$ as $m\to\infty$, which implies that almost surely $\tau_{m_j^+}\leq t_j\leq\tau_{m_j^-}$ for sufficiently large j. Therefore, almost surely

$$N(\tau_{m_j^-})\leq N(t_j)\leq N(\tau_{m_j^+}) \tag{3.14}$$

for sufficiently large j. The assumptions of Theorem 1 imply that $s_j\gg m_j^+$ and $n_js_j\gg(m_j^+)^2$ and the same is true for m_j^-. Therefore, by Lemma 15, using $\to_{\mathrm{p}}$ to denote convergence in probability as $j\to\infty$, we have

$$\frac{\tau_{m_j^+-1}N(\tau_{m_j^+-1})}{m_j^+}=\frac{\tau_{m_j^+-1}}{m_j^+}\sum_{k=1}^{m_j^+}N^{k,m_j^+}=\frac{1}{m_j^+}\sum_{k=1}^{m_j^+}W^{k,m_j^+}\to_{\mathrm{p}}\gamma.$$

Now using again that $m\tau_m \to 2/c$ almost surely as $m \to \infty$, we get

$$\left(m_j^+\right)^{-2} N(\tau_{m_j^+}) \to_{\mathrm{p}} \frac{c\gamma}{2}$$

and, therefore,

$$t_j^2 N(\tau_{m_j^+}) \to_{\mathrm{p}} \frac{2(1+\varepsilon)^2\gamma}{c}. \tag{3.15}$$

By the same reasoning,

$$t_j^2 N(\tau_{m_j^-}) \to_{\mathrm{p}} \frac{2(1-\varepsilon)^2\gamma}{c}. \tag{3.16}$$

By letting $\varepsilon \to 0$, we obtain the result from (3.14), (3.15) and (3.16). □

Acknowledgements. This project began while the authors were attending a Bath, UNAM, and CIMAT (BUC) workshop in Guanajuato, Mexico in May, 2016. The authors thank Andreas Kyprianou, Juan Carlos Pardo and Victor Rivero for their roles in organizing this workshop. ASJ would like to thank Amaury Lambert and Emmanuel Schertzer for insightful discussions about the nested coalescent.

REFERENCES

[1] ALDOUS, D. J. (1999). Deterministic and stochastic models for coalescence (aggregation and coagulation): A review of the mean-field theory for probabilists. *Bernoulli* **5** 3–48. MR1673235

[2] ALDOUS, D. J. and BANDYOPADHYAY, A. (2005). A survey of max-type recursive distributional equations. *Ann. Appl. Probab.* **15** 1047–1110. MR2134098

[3] BANSAYE, V., MÉLÉARD, S. and RICHARD, M. (2016). Speed of coming down from infinity for birth-and-death processes. *Adv. in Appl. Probab.* **48** 1183–1210. MR3595771

[4] BERESTYCKI, J. and BERESTYCKI, N. (2009). Kingman's coalescent and Brownian motion. *ALEA Lat. Am. J. Probab. Math. Stat.* **6** 239–259. MR2534485

[5] BERESTYCKI, J., BERESTYCKI, N. and LIMIC, V. (2010). The Λ-coalescent speed of coming down from infinity. *Ann. Probab.* **38** 207–233. MR2599198

[6] BERESTYCKI, N. (2009). *Recent Progress in Coalescent Theory*. *Ensaios Matemáticos* [*Mathematical Surveys*] **16**. Sociedade Brasileira de Matemática, Rio de Janeiro. MR2574323

[7] BLANCAS, A., DUCHAMPS, J.-J., LAMBERT, A. and SIRI-JÉGOUSSE, A. (2018). Trees within trees: Simple nested coalescents. *Electron. J. Probab.* **23** 1–27.

[8] CREER, S. et al. (2010). Ultrasequencing of the meiofaunal biosphere: Practice, pitfalls and promises. *Mol. Ecol.* **19** 4–20.

[9] DAWSON, D. A. (2018). Multilevel mutation-selection systems and set-valued duals. *J. Math. Biol.* **76** 295–378. MR3742789

[10] DHERSIN, J.-S., FREUND, F., SIRI-JÉGOUSSE, A. and YUAN, L. (2013). On the length of an external branch in the beta-coalescent. *Stochastic Process. Appl.* **123** 1691–1715. MR3027896

[11] EWENS, W. J. (1972). The sampling theory of selectively neutral alleles. *Theor. Popul. Biol.* **3** 87–112; erratum, ibid. **3** (1972), 240; erratum, ibid. **3** (1972), 376. MR0325177

[12] GIVENS, C. R. and SHORTT, R. M. (1984). A class of Wasserstein metrics for probability distributions. *Michigan Math. J.* **31** 231–240. MR0752258

[13] HARVEY, P. H., MAY, R. M. and NEE, S. (1994). Phylogenies without fossils. *Evolution* **48** 523–529.
[14] KANTOROVIČ, L. V. and RUBINŠTEĬN, G. Š. (1958). On a space of completely additive functions. *Vestnik Leningrad Univ. Math.* **13** 52–59. MR0102006
[15] KINGMAN, J. F. C. (1982). The coalescent. *Stochastic Process. Appl.* **13** 235–248. MR0671034
[16] LAMBERT, A. and SCHERTZER, E. (2018). Coagulation-transport equations and the nested coalescents. Available at arXiv:1807.09153.
[17] MADDISON, W. P. (1997). Gene trees in species trees. *Syst. Biol.* **46** 523–536.
[18] MÖHLE, M. (2000). Total variation distances and rates of convergence for ancestral coalescent processes in exchangeable population models. *Adv. in Appl. Probab.* **32** 983–993. MR1808909
[19] MOOERS, A. O. and HEARD, S. B. (1997). Inferring evolutionary process from phylogenetic tree shape. *Q. Rev. Biol.* **72** 31–54.
[20] MORAN, P. A. P. (1958). Random processes in genetics. *Proc. Camb. Philos. Soc.* **54** 60–71. MR0127989
[21] MORGAN, M. J., BASS, D., BIK, H., BIRKY, C. W., BLAXTER, M., CRISP, M. D., DERYCKE, S., FITCH, D., FONTANETO, D. et al. (2014). A critique of Rossberg et al.: Noise obscures the genetic signal of meiobiotal ecospecies in ecogenomic datasets. *Proc. Royal Soc., Biol. Sci.* **281** 20133076.
[22] ROSSBERG, A. G., ROGERS, T. and MCKANE, A. J. (2013). Are there species smaller than 1 mm? *Proc. Royal Soc. Biol. Sci.* **280** 20131248.
[23] ROSSBERG, A. G., ROGERS, T. and MCKANE, A. J. (2014). Current noise-removal methods can create false signals in ecogenomic data. *Proc. Royal Soc. Biol. Sci.* **281** 20140191.
[24] SZÖLLŐSI, G. J., TANNIER, E., DAUBIN, V. and BOUSSAU, B. (2014). The inference of gene trees with species trees. *Syst. Biol.* **64** e42–e62.

A. BLANCAS
INSTITUT FÜR MATHEMATIK
GOETHE UNIVERSITÄT FRANKFURT
ROBERT-MAYER-STR. 10,
60325 FRANKFURT AM MAIN
GERMANY
E-MAIL: airam.blancas@gmail.com

J. SCHWEINSBERG
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF CALIFORNIA, SAN DIEGO
9500 GILMAN DRIVE
LA JOLLA, CALIFORNIA 92093-0112
USA
E-MAIL: jschwein@math.ucsd.edu

T. ROGERS
DEPARTMENT OF MATHEMATICAL SCIENCES
UNIVERSITY OF BATH
BATH BA27AY
UNITED KINGDOM
E-MAIL: t.c.rogers@bath.ac.uk

A. SIRI-JÉGOUSSE
DEPARTAMENTO DE PROBABILIDAD Y ESTADÍSTICA
IIMAS
UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
APARTADO POSTAL 126
CIUDAD DE MÉXICO
CDMX, 01000 MÉXICO
MÉXICO
E-MAIL: arno@sigma.iimas.unam.mx