Predicting long pendant edges in model phylogenies, with applications to biodiversity and tree inference

Sergey Bocharov[2], Simon Harris[1], Emma Kominek[3], Arne Ø. Mooers[3] and Mike Steel[4*]

[1] *Department of Foundational Mathematics, Xian Jiaotong-Liverpool University, Suzhou, China*
[2] *Department of Statistics, University of Auckland, Auckland, New Zealand*
[3] *Biological Sciences, Simon Fraser University, 8888 Univ. Drive, Burnaby BC Canada V5A 1S6*
[4] *Biomathematics Research Centre, University of Canterbury, Christchurch, New Zealand*

**Corresponding author: mike.steel@canterbury.ac.nz*

Abstract

In the simplest phylogenetic diversification model (the pure-birth Yule process), lineages split independently at a constant rate λ for time t. The length of a randomly chosen edge (either interior or pendant) in the resulting tree has an expected value that rapidly converges to $\frac{1}{2\lambda}$ as t grows, and thus is essentially independent of t. However, the behaviour of the length L of the *longest* pendant edge reveals remarkably different behaviour: L converges to $t/2$ as the expected number of leaves grows. Extending this model to allow an extinction rate μ (where $\mu < \lambda$), we also establish a similar result for birth–death trees, except that $t/2$ is replaced by $t/2 \cdot (1 - \mu/\lambda)$. This 'complete' tree may contain subtrees that have died out before time t; for the 'reduced tree' that just involves the leaves present at time t and their direct ancestors, the longest pendant edge length L again converges to $t/2$. Thus, there is likely to be at least one extant species whose associated pendant branch attaches to the tree approximately half-way back in time to the origin of the entire clade. We also briefly consider the length of the shortest edges. Our results are relevant to phylogenetic diversity indices in biodiversity conservation, and to quantifying the length of aligned sequences required to correctly infer a tree. We compare our theoretical results with simulations, and with the branch lengths from a recent phylogenetic tree of all mammals.

Key words: Phylogenetic diversification models, birth–death process, phylogenetic diversity

Introduction

Stochastic birth–death models (of speciation and extinction) model the tree-like diversification of species over evolutionary time-scales and play an important role in systematic biology. These models trace back to a seminal paper of Yule (1925), and a rich literature of probabilistic modelling of birth–death processes has developed, from the 1940s to the present. These stochastic models lead to predictions concerning the 'shape' of phylogenetic trees, and thereby allow the testing of different speciation-extinction models. The models also allow the formulation of estimators for speciation and extinction rates (based on a given model) from large phylogenies (see e.g. Nee et al. (1994)) and provide priors for Bayesian phylogenetic methods. Importantly, the birth–death process tends to produce trees whose distribution of splitting times is intermediate between the deep splits expected in adaptive radiations (Gavrilets and Vose, 2005) and the shallow splits expected under density-dependent dynamics (Hey, 1992). Although more recent compilations are required, published trees tend to produce mildly deeper-than-expected splitting times (Morlon et al., 2010).

Mathematical investigations into phylogenetic diversification models have also led to new insights and predictions (e.g. Aldous (2001); Aldous and Popovic (2005); Lambert and Stadler (2013), and, most recently, Louca and Pennell (2020)). This last paper established an inherent limitation on the extent to which it is possible to identify an unknown diversification model from an observed phylogeny, regardless of its size.

One aspect of any birth–death process is the length it predicts for edges of the tree (note that 'edge' is synonymous with 'branch' and a 'pendant edge' is one that is incident with a tip (or leaf) at the present). The simplest such model, the *Yule process*, has a single constant speciation rate λ and the expected length of a randomly chosen edge in a Yule

tree grown for time t quickly converges to $\sim \frac{1}{2\lambda}$ (Steel and Mooers, 2010) as t grows; the factor of 2 in the denominator may at first appear surprising but is due to the process taking place on a binary tree rather than along a path). Similar results for the length of a randomly selected edge are known for birth–death trees (Mooers et al., 2012; Stadler and Steel, 2012).

In this paper, we focus on the length of the *longest* pendant edge in a tree generated by a birth–death process, for both the complete tree (which is relevant to studies involving total sampling through time, as with certain viral data-sets) and for the 'reduced tree', which corresponds to phylogenies reconstructed from genomic data sampled from individual taxa at the present.

For Yule trees (i.e. a birth process where extinction is set to 0), Gascuel and Steel (2010) showed that the expected length of the longest edge does not converge to 0 as $\lambda \to \infty$ (more precisely, the proof of their Proposition 2.2 showed that, with strictly positive probability, an edge of length at least $t/4$ exists). Here, we first establish a sharper result: the longest pendant edge for a large Yule tree grows linearly with t as the expected number of leaves grows, and it converges to $t/2$. In particular, it is independent of any fixed value of λ, in contrast to the 'average' pendant edge length, which essentially only depends on λ (and not t) for large trees.

We then extend this result to birth–death models (i.e. with nonzero extinction), for both the 'complete tree' and for the 'reduced tree'. For complete birth–death trees, the term $t/2$ is replaced by a term that is smaller but still proportional to t, whereas for reduced trees, the term $t/2$ again surfaces as the appropriate limit. Thus, in trees inferred from data at the present, the models predict that there is likely to be at least one extant species whose associated pendant branch attaches to the tree approximately half-way back in time to the origin of the entire clade.

We compare our theoretical results with simulations of Yule, and birth-death trees (both complete and reduced) and then describe some implications of this result for

systematic biology. First, we show that for large phylogenies generated under a birth–death model, the most 'evolutionary distinct' taxa are likely to be those at the end of the longest pendant edges. Second, we describe how our results bear on the question of how much genomic data (number of aligned sequence sites) are required in order to infer a fully resolved tree correctly.

Finally, we compare our expectations with recent inferred trees for 114 mammal families, and provide some concluding comments.

The Yule process

We first consider a simple pure-birth model, referred to as the *Yule process*. This process begins with a single species at time 0, which persists as a single lineage for a random time τ, where τ is exponentially distributed with rate parameter λ (i.e. $\tau \sim \mathrm{Exp}(\lambda)$). At time τ, this lineage splits into two lineages, representing a speciation event. These lineages then evolve under the same process and independently of each other and of all the previous history (i.e. they each persist for an independent exponential time and are replaced by two lineages at that point, and so on). We denote the resulting Yule tree at time t by $\mathcal{T}_t$, we let L_t be the length of the longest pendant edge of $\mathcal{T}_t$ and let S_t be the length of the shortest pendant edge of $\mathcal{T}_t$, as indicated in Fig. 1.

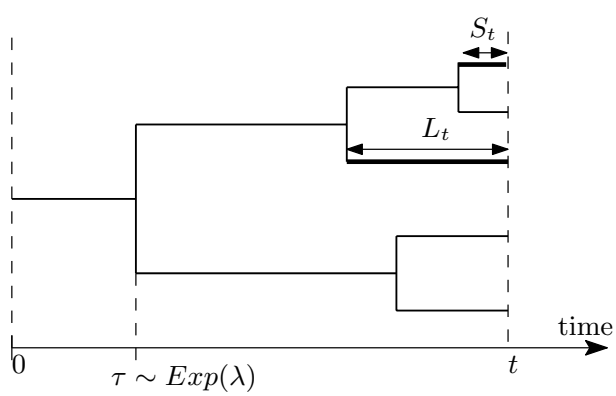


Fig. 1. A Yule (pure-birth) tree starts with a single lineage at time $t = 0$, and each lineage persists for an exponentially distributed time τ with mean $1/\lambda$. The four splitting events result in five leaf species at time t. The edge labels L_t and S_t denotes the length of the longest and shortest pendant edge of this tree, respectively.

The longest pendant edge in a Yule (pure-birth) tree

Proposition 1 Let L_t denote the length of the longest pendant edge in a Yule tree at time t with speciation rate λ. The following results hold:

(i)
$$\mathbb{P}(L_t \leqslant x) = \frac{1 - e^{-\lambda x}}{1 + e^{\lambda(t-2x)} - e^{-\lambda x}}\ , \quad (x < t). \tag{1}$$

(ii) The length of longest pendant edge when centred about $t/2$ (i.e., $\lambda(L_t - \frac{t}{2})$) converges in distribution to a logistic distribution as $\lambda t \to \infty$, where:
$$\lim_{\lambda t \to \infty} \mathbb{P}\big(\lambda(L_t - \frac{t}{2}) \leqslant y\big) = \frac{1}{1 + e^{-2y}}. \tag{2}$$

(iii) Furthermore, L_t/t converges to $\frac{1}{2}$ in mean and in probability as λt becomes large (i.e. as $\lambda t \to \infty$, $\frac{\mathbb{E}(L_t)}{t} \to \frac{1}{2}$, and for all $\epsilon > 0$, $\mathbb{P}\left(\left|\frac{L_t}{t} - \frac{1}{2}\right| < \epsilon\right) \to 1$).

Proposition 1 is a special case of the more general results stated later in Corollary 1 and Theorem 1 (the proofs for which are provided in the Appendix). For the Yule model, it is even possible to calculate the expected value of L_t exactly, as we now show.

The expected length of the longest pendant edge for Yule trees

For ease of presentation in what follows, let $\mu_t = \mathbb{E}(L_t)$, the expected length of the longest pendant edge in a Yule tree.

Proposition 2
$$\mu_t = \frac{t}{2} + \frac{e^{-\lambda t/2}}{\lambda\sqrt{4 - e^{-\lambda t}}}\left[\tan^{-1}\left(\frac{2e^{\lambda t/2} - e^{-\lambda t/2}}{\sqrt{4 - e^{-\lambda t}}}\right) - \tan^{-1}\left(\frac{e^{-\lambda t/2}}{\sqrt{4 - e^{-\lambda t}}}\right)\right]$$
and $\lim_{\lambda t \to \infty} \mu_t/t = \frac{1}{2}$.

Remarks:

(1) In Proposition 2, the (single) edge in a tree that has just one leaf is treated as a pendant edge. However, one might equally regard the stem edge as an interior edge, in which case a longest pendant edge would not exist in the one-leaf case (so one might then set $L_t = 0$ in that case). Proposition 2 is easily adjusted to accommodate this. Let

$$L'_t = \begin{cases} L_t, & \text{if } N_t > 1; \\ 0, & \text{if } N_t = 1. \end{cases}$$

In contrast to L_t, the probability distribution $\mathbb{P}(L'_t \leqslant x)$ is now *continuous* at $x = t$ (where it takes the value 1). It can be shown that $\mathbb{E}(L'_t) = \mu_t - te^{-\lambda t}$ and that $\mu_t/t = \mathbb{E}(L'_t/t)$ is monotone increasing, from 0 (at the limit $t \to 0^+$) to $\frac{1}{2}$ (as $\lambda t \to \infty$). By contrast, $\mu_t/t = \mathbb{E}(L_t/t)$ is monotone decreasing, from 1 (at the limit $t \to 0^+$) to $\frac{1}{2}$ (as $\lambda t \to \infty$).

(2) *The shortest pendant branch length.* The distribution of the length of the shortest pendant edge (we denote this by S_t) can also be exactly described. The condition that the shortest edge is longer than x is equivalent to the condition that each lineage present at time $t - x$ does not split over the last time x (these independent events each have probability $e^{-\lambda x}$). Since $N_{t-x} \sim \text{Geom}(e^{-\lambda(t-x)})$, we find, for $x \in [0, t]$:

$$\mathbb{P}(S_t \geqslant x) = \mathbb{E}(e^{-\lambda x N_{t-x}}) = \frac{e^{-\lambda t}}{1 + e^{-\lambda t} - e^{-\lambda x}}. \tag{3}$$

In particular, since $S_t \in [0, t]$, we have $\mathbb{P}(S_t = t) = \mathbb{P}(S_t \geqslant t) = e^{-\lambda t}$, corresponding to no death of the initial ancestor (and $\mathbb{P}(S_t \geqslant x) = 0$ for $x > t$) thus:

$$\mathbb{E}(S_t) = \int_0^t \mathbb{P}(S_t \geqslant x)dx = 2te^{-\lambda t}/(1 + e^{-\lambda t}).$$

Thus, as λt grows, $\mathbb{E}(S_t) \sim 2te^{-\lambda t}$. On the other hand, putting $x = ye^{-\lambda t}/\lambda$ in Eqn. (3) and noting that $1 - e^{-w} \sim w$ for w small, we see that $\lambda e^{\lambda t} S_t$ converges in distribution, with $\mathbb{P}(\lambda e^{\lambda t} S_t \geqslant y) \to \frac{1}{1+y}$ as $\lambda t \to \infty$. This further implies that the ratio $\frac{\ln(\lambda S_t)}{\lambda t}$ converges in probability to -1 as λt grows.

Long pendant edges in a (complete) birth–death phylogeny

The simple birth–death process generalises the Yule model from the previous section by allowing the extinction of lineages. In this model, each lineage (i.e. branch) present at any given time behaves independently, being replaced by two new lineages at a constant rate λ, and dying at a constant rate μ. Thus, each lineage persists for an independent random time τ which is exponentially distributed with rate $\lambda + \mu$. At the end of its lifetime, it is either replaced by two new lineages, with probability $\lambda/(\lambda + \mu)$, or by none, with probability $\mu/(\lambda + \mu)$. Lineages once alive behave independently of each other and of all the previous history, and in the same probabilistic manner as the parent (i.e. after an independent exponential time of rate $\lambda + \mu$, it will terminate and either be replaced by 2 or 0 lineages, etc.)

In this section, we will extend our results for the Yule process to obtain results for the longest pendant edges in a birth–death tree. In the Section *'Length of pendant edges in the reduced tree'*, we will then consider the pendant edges in the *reduced tree*, which is formed by tracing back the lineages that are alive at some fixed time t (i.e. we prune any lineages (subtrees) that have already died out before time t).

We start by recalling some well-known properties about birth–death processes.

Some properties of birth–death trees

The following results concerning birth-death processes are classical (see e.g. Kendall (1948) or Grimmett and Stirzaker (2001)). Recall that the probability-generating function of a discrete random variable X is the function $\varphi(\theta) = \sum_n \mathbb{P}(X = n) \cdot \theta^n = \mathbb{E}(\theta^X)$ where θ is a formal variable and n ranges over all possible values X can take.

Lemma 1 The probability-generating function $\varphi_t(\theta)$ of N_t (the number of leaves alive at time t in the complete tree $\mathcal{T}_t$) when $\lambda > \mu$ is given by:

$$\varphi_t(\theta) = \mathbb{E}(\theta^{N_t}) = p_t + (1 - p_t)\frac{q_t\theta}{1 - (1 - q_t)\theta} \tag{4}$$

where:

$$p_t = \frac{\mu - \mu \mathrm{e}^{-(\lambda-\mu)t}}{\lambda - \mu \mathrm{e}^{-(\lambda-\mu)t}} \tag{5}$$

and

$$q_t = \frac{(\lambda - \mu)\mathrm{e}^{-(\lambda-\mu)t}}{\lambda - \mu \mathrm{e}^{-(\lambda-\mu)t}}. \tag{6}$$

In particular, the probability that the process becomes extinct before time t is:

$$\mathbb{P}(N_t = 0) = p_t \tag{7}$$

and, conditional on non-extinction by time t, N_t has a geometric distribution on $\{1, 2, 3, \cdots\}$ with parameter q_t. Formally:

$$\mathbb{P}(N_t = k | N_t > 0) = q_t(1 - q_t)^{k-1} \qquad (k \geqslant 1).$$

N_t is said to have a *modified geometric* distribution written as $N_t \sim \mathrm{ModGeom}(p_t, q_t)$.

Throughout this paper, we will let $\rho = \frac{\mu}{\lambda}$, which is sometimes called the 'turnover rate' in phylogenetic diversification models and plays a fundamental role. For example, in the *supercritical* case ($\lambda > \mu$), it follows from the formulae above (Eqns. (5) and (7)), that

$$\mathbb{P}(\text{eventual extinction}) = \lim_{t\to\infty} \mathbb{P}(N_t = 0) = \rho.$$

We assume throughout this paper that $\lambda > \mu \geqslant 0$, and so $0 \leqslant \rho < 1$. Thus the birth–death process is supercritical with a strictly positive probability of $1 - \rho$ of surviving forever.

Length of pendant edges in complete trees

For any $t \geqslant x \geqslant 0$, define the random variable N_t^x as follows:

$$\begin{aligned} N_t^x &:= \text{number of lineages alive at time } t \text{ of age at least } x \\ &= \text{number of pendant edges in the tree at time } t \text{ with length } \geqslant x. \end{aligned}$$

For the birth–death tree, we can find the distribution for the number of pendant edges greater than length x at time t explicitly. We will then be easily able to deduce the

distributions for the sizes of the longest pendant edges at time t. For $x \in [0, t]$, let:

$$p_{t,x,\lambda,\mu} := \frac{1 - \rho - e^{-(\lambda+\mu)x} + \rho e^{(\lambda-\mu)t-2\lambda x}}{1 - \rho - e^{-(\lambda+\mu)x} + e^{(\lambda-\mu)t-2\lambda x}} \tag{8}$$

and

$$q_{t,x,\lambda,\mu} := \frac{1 - \rho}{1 - \rho - e^{-(\lambda+\mu)x} + e^{(\lambda-\mu)t-2\lambda x}} \tag{9}$$

The proof of the following fundamental result is provided in the Appendix.

Lemma 2 (Distribution of the number of long pendant edges in a birth–death process) The number of pendant edges of length at least x at time t, N_t^x, has a modified geometric distribution with $N_t^x \sim \mathrm{ModGeom}(p_{t,x,\lambda,\mu}, q_{t,x,\lambda,\mu})$. Formally:

$$\mathbb{P}(N_t^x = k) = \begin{cases} p & (k = 0) \\ (1-p)q(1-q)^{k-1} & (k \geqslant 1) \end{cases} \tag{10}$$

where $p = p_{t,x,\lambda,\mu}$ and $q = q_{t,x,\lambda,\mu}$, as given in Eqns. (8) and (9). In particular,

$$\mathbb{E}(N_t^x) = (1-p)/q = e^{(\lambda-\mu)t-2\lambda x} \tag{11}$$

$$\mathbb{V}ar(N_t^x) = (1-p)(1-q+p)/q^2 \tag{12}$$

Let L_t denote the length of the longest pendant edge that is still alive at time t in a birth–death tree at time t.

Corollary 1 (Distribution of the longest pendant edges in a birth–death process)

$$\mathbb{P}(L_t \leqslant x) = \begin{cases} p_{t,x,\lambda,\mu} & (x < t), \\ 1 & (x = t). \end{cases} \tag{13}$$

In particular, $\mathbb{P}(L_t = t) = e^{-(\lambda+\mu)t}$, which corresponds to the initial ancestor neither branching nor dying over the entire time period t. Moreover, if $L_t^{(k)}$ denotes the length of the kth longest pendant edge at time t, then:

$$\mathbb{P}(L_t^{(k)} \leqslant x) = 1 - (1-p)(1-q)^{k-1} \tag{14}$$

where $p = p_{t,x,\lambda,\mu}$ and $q = q_{t,x,\lambda,\mu}$, as given in Eqns. (8) and (9).

Proof of Corollary 1. This follows directly from the distribution of N_t^x given in Lemma 2 since $\mathbb{P}(L_t \leqslant x) = \mathbb{P}(N_t^x = 0)$ and $\mathbb{P}(L_t^{(k)} \leqslant x) = \mathbb{P}(N_t^x \leqslant k-1)$. □

Remark: There may be extinction by time t in the birth–death process when $\mu > 0$, in which case $L_t = 0$ and this event has probability:

$$\mathbb{P}(L_t = 0) = p_{t,0,\lambda,\mu} = \rho(e^{(\lambda-\mu)t} - 1)/(e^{(\lambda-\mu)t} - \rho).$$

Similarly, there may not always be a kth longest pendant edge (even for the Yule process), as only a finite number of lineages are present at time t. In such cases, by convention, we set the random variable $L_t^{(k)} = 0$ whenever $N_t < k$, and we note that $\mathbb{P}(L_t^{(k)} = 0) = \mathbb{P}(N_t < k) > 0$ whenever $k \geqslant 2$.

Distribution of the longest pendant edge in large complete trees

We now consider what happens for 'large' birth–death trees; by 'large', we mean the expected number of leaves in a birth–death tree (i.e. $e^{(\lambda-\mu)t}$) is large. If we regard ρ $(= \mu/\lambda)$ as a fixed value, then the expected number of leaves in a birth tree grows as a function of λt. Thus, in the following theorem, we consider what happens as λt grows (note that this may be due to λ becoming large even if t is not (for instance, a rapid species radiation in short time) or due to t becoming large (for instance, a tree that traces back deep into the past).

In this subsection, we assume that the birth–death process is supercritical with $\lambda > \mu$ and ρ is fixed. The expected number of leaves in the tree at time t can then be written as $e^{(1-\rho)\lambda t}$, which grows as a function of λt. Indeed, the number of leaves conditional on survival will also grow asymptotically at the rate $e^{(1-\rho)\lambda t}$ whenever $\lambda t \to \infty$. We can now state our main result for large complete birth–death trees, the formal proof of which is provided in the Appendix.

Theorem 1 (Longest pendant edges in the birth–death process as λt grows) Let $\lambda > \mu$ with $\rho := \mu/\lambda$ being fixed. Conditional on the survival of the tree at time t, the following results hold:

(i) For any $y \in \mathbb{R}$, the number of pendant edges at time t that are longer than $t(1-\rho)/2 + y/\lambda$ converges in distribution as $\lambda t \to \infty$ to a geometric distribution supported on $\{0, 1, 2, \dots\}$ as given by:

$$\lim_{\lambda t \to \infty} \mathbb{P}\left(N_t^{\frac{t}{2}(1-\rho)+\frac{y}{\lambda}} = k \,\middle|\, N_t > 0 \right) = \frac{1-\rho}{1-\rho+e^{-2y}} \left(\frac{e^{-2y}}{1-\rho+e^{-2y}} \right)^k \quad (k \geqslant 0). \tag{15}$$

(ii) In particular, the length of longest pendant edge when centred about $t(1-\rho)/2$ (i.e. $\lambda(L_t - \frac{t}{2}(1-\rho))$) converges in distribution as $\lambda t \to \infty$ to a logistic distribution, where

$$\lim_{\lambda t \to \infty} \mathbb{P}\left(\lambda(L_t - \frac{t}{2}(1-\rho)) \leqslant y \,\middle|\, N_t > 0 \right) = \frac{1-\rho}{1-\rho+e^{-2y}}. \tag{16}$$

Furthermore, given survival up until time t, L_t/t converges in probability to the constant $(1-\rho)/2$, and in addition the average value of L_t/t also tends to $(1-\rho)/2$ (ie. $\mathbb{E}(L_t/t | N_t > 0) \to (1-\rho)/2$).

(iii) More generally, for each $k \geqslant 1$, the length of the kth longest pendant edge when centred about $t(1-\rho)/2$, that is, $\lambda(L_t^{(k)} - \frac{t}{2}(1-\rho))$, converges in distribution for large trees where

$$\lim_{\lambda t \to \infty} \mathbb{P}\left(\lambda(L_t^{(k)} - \frac{t}{2}(1-\rho)) \leqslant y \,\middle|\, N_t > 0 \right) = 1 - \left(\frac{e^{-2y}}{1-\rho+e^{-2y}} \right)^k \tag{17}$$

Furthermore, given survival up until time t, $L_t^{(k)}/t$ converges in probability to the constant $(1-\rho)/2$, and in addition the average value of $L_t^{(k)}/t$ also tends to $(1-\rho)/2$ (ie. $\mathbb{E}(L_t^{(k)}/t | N_t > 0) \to (1-\rho)/2$).

Length of pendant edges in the reduced tree

We now turn from complete trees to reduced trees based just on the leaves extant at the present time t. For this latter class of trees, we will show that the length of the longest pendant edge turns out to be (asymptotically) independent of the rates λ and μ, in contrast to the situation with complete trees, for which the asymptotic value was shown to be $t/2 \cdot (1-\rho)$, and also in contrast to a randomly selected pendant edge from a reduced tree, for which the expected length does depend on ρ (Stadler and Steel, 2012).

More precisely, the *reduced tree* of the birth–death process, is the genealogical tree constructed only from the ancestors of the lineages still alive at time t. In other words, starting with the complete birth–death tree, we prune away any subtrees that died out before time t (or, equivalently, only keeping a lineage alive at time $s < t$ if it has at least one descendant still alive at time t). This is illustrated in Fig. 2. The reduced tree is essentially the same as what is referred to as the 'reconstructed tree', except that in the latter the stem edge is often not present.

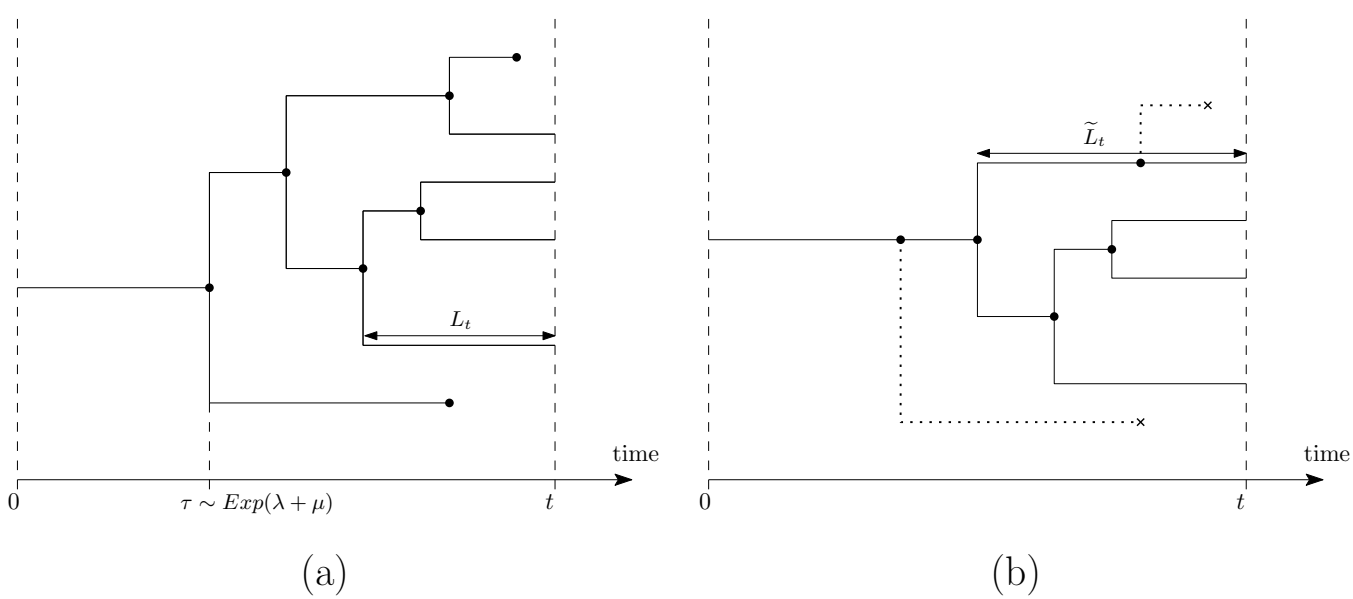


Fig. 2. (a) A complete birth–death tree and (b) the corresponding reduced tree, where the dashed lineages have been pruned off. In (a), the edge of length L_t is the longest pendant edge that is still alive at time t (the longer pendant edge below it dies before time t). Notice also that the length of the longest pendant edge in the reduced tree ($\widetilde{L}_t$) is greater than the length of the longest pendant edge of the complete tree (L_t).

As any branches ending in a death are pruned away, a reduced tree for the birth–death process looks similar to a Yule tree. However, the probabilistic behaviour of the reduced tree is more complicated than that of a Yule tree. In a Yule tree, each individual branches into two at some constant rate λ. In the reduced tree, however, the rate of branching of each individual becomes time-dependant at rate $\lambda(1 - p_{t-s})$ for any lineage alive at time s, where $1 - p_{t-s} := \mathbb{P}(N_{t-s} > 0)$ corresponds to the probability that at least one lineage born at time s will still be alive at time t.

Although the number of pendant edges that persist to time t remains unchanged as we prune away any sub-trees that died out in the birth–death tree, it is straightforward to see that these pendant edges can either remain the same or increase their length in the reduced tree (see Fig. 2). As a consequence, the behaviour of the long pendant edges in the reduced tree turns out to be somewhat different than in the corresponding complete birth–death tree when extinctions can occur ($\mu > 0$).

Length of pendant edges in a reduced tree

Lemma 3 (Distribution for the number of long pendant edges in the reduced birth–death tree) The number of pendant edges of length at least x at time t, $\widetilde{N}_t^x$, has a modified geometric distribution with $\widetilde{N}_t^x \sim \mathrm{ModGeom}(\widetilde{p}_{t,x,\lambda,\mu}, \widetilde{q}_{t,x,\lambda,\mu})$. Formally:

$$\mathbb{P}(\widetilde{N}_t^x = k) = \begin{cases} \widetilde{p} & (k = 0) \\ (1 - \widetilde{p})\widetilde{q}(1 - \widetilde{q})^{k-1} & (k \geqslant 1) \end{cases} \tag{18}$$

where $\widetilde{p} = \widetilde{p}_{t,x,\lambda,\mu}$ and $\widetilde{q} = \widetilde{q}_{t,x,\lambda,\mu}$ are given in the Appendix as Eqns. ((37)) and ((38)). In particular,

$$\mathbb{E}(\widetilde{N}_t^x) = (1 - \widetilde{p})/\widetilde{q} = e^{(\lambda-\mu)(t-x)}(1 - p_x)q_x \tag{19}$$

$$\mathbb{V}ar(\widetilde{N}_t^x) = (1 - \widetilde{p})(1 - \widetilde{q} + \widetilde{p})/\widetilde{q}^2 \tag{20}$$

Let $\widetilde{L}_t$ denote the length of the longest pendant edge in a reduced birth-death tree at time t.

Corollary 2 (Distribution of the longest pendant edges in the reduced birth–death tree for fixed values of λ and t)

$$\mathbb{P}(\widetilde{L}_t \leqslant x) = \begin{cases} \widetilde{p}_{t,x,\lambda,\mu} & (x < t), \\ 1 & (x \geqslant t). \end{cases} \tag{21}$$

In particular, $\mathbb{P}(\widetilde{L}_t = 0) = p_t$, which corresponds to extinction by time t, and $\mathbb{P}(\widetilde{L}_t = t) = (1 - p_t)q_t$, which corresponds to the initial ancestor having exactly one descendant alive at time t.

Moreover, if $\widetilde{L}_t^{(k)}$ is the length of the kth longest pendant edge at time t, then:

$$\mathbb{P}(\widetilde{L}_t^{(k)} \leqslant x) = 1 - (1 - \widetilde{p})(1 - \widetilde{q})^{k-1} \tag{22}$$

where $\widetilde{p} = \widetilde{p}_{t,x,\lambda,\mu}$ and $\widetilde{q} = \widetilde{q}_{t,x,\lambda,\mu}$, as in Eqns. (37) and (38).

Length of the longest pendant edges in large reduced trees

When $\lambda > \mu > 0$, conditional on survival, the reduced tree will become an increasingly large tree as λt grows. Note that from Eqns. (5) and (6), as $\lambda t \to \infty$:

$$p_t = \mathbb{P}(N_t = 0) \to \rho, \qquad q_t \sim (1 - \rho)e^{-\lambda t(1-\rho)}. \tag{23}$$

Using the formulae for $\widetilde{p}_{t,x,\lambda,\mu}$ and $\widetilde{q}_{t,x,\lambda,\mu}$ given by Eqns. (37) and (38) respectively in the Appendix, and setting $x = \frac{t}{2} + \frac{y}{\lambda}$, in those equations, we find that, as $\lambda t \to \infty$,

$$\widetilde{q}_{t,\frac{t}{2}+\frac{y}{\lambda},\lambda,\mu} \to \widetilde{q}_* := \frac{1}{1 + (1 - \rho)e^{-2(1-\rho)y}} \tag{24}$$

and

$$\widetilde{p}_{t,\frac{t}{2}+\frac{y}{\lambda},\lambda,\mu} \to \widetilde{p}_* := \frac{1 + \rho(1 - \rho)e^{-2(1-\rho)y}}{1 + (1 - \rho)e^{-2(1-\rho)y}} = \rho + (1 - \rho)\widetilde{q}_*. \tag{25}$$

Theorem 2 (Longest pendant edges in the reduced tree as λt grows) Let $\lambda > \mu$ with $\rho := \mu/\lambda$ being fixed. Conditional on the survival of the tree at time t, the following results hold:

(i) For any $y \in \mathbb{R}$, the number of pendant edges in the reduced tree at time t that are longer than $t/2 + y/\lambda$ converges in distribution as $\lambda t \to \infty$ to a geometric

distribution supported on $\{0, 1, 2, \dots\}$ as given by:

$$\lim_{\lambda t \to \infty} \mathbb{P}\left(\widetilde{N}_t^{\frac{t}{2}+\frac{y}{\lambda}} = k \,\middle|\, N_t > 0\right) = \widetilde{q}_*(1-\widetilde{q}_*)^k \quad (k \geqslant 0). \tag{26}$$

(ii) In particular, the length of the longest pendant edge in the reduced tree when centred about $t/2$ (i.e. $\lambda(\widetilde{L}_t - \frac{t}{2})$) converges in distribution as $\lambda t \to \infty$ to a logistic distribution, where

$$\lim_{\lambda t \to \infty} \mathbb{P}\left(\widetilde{L}_t - \frac{t}{2} \leqslant \frac{y}{\lambda} \,\middle|\, N_t > 0\right) = \widetilde{q}_* = \frac{1}{1 + (1-\rho)e^{-2(1-\rho)y}} \tag{27}$$

Furthermore, given survival up until time t, $\widetilde{L}_t/t$ converges in probability to the constant $\frac{1}{2}$, and in addition the average value of $\widetilde{L}_t/t$ also tends to $\frac{1}{2}$ (ie. $\mathbb{E}(\widetilde{L}_t/t | N_t > 0) \to \frac{1}{2}$.

(iii) More generally, for each $k \geqslant 1$, the length of the kth longest pendant edge in the reduced tree when centred about $t/2$ (i.e. $\lambda(\widetilde{L}_t^{(k)} - \frac{t}{2})$) converges in distribution for large trees, where:

$$\lim_{\lambda t \to \infty} \mathbb{P}\left(\lambda(\widetilde{L}_t^{(k)} - \frac{t}{2}) \leqslant y \,\middle|\, N_t > 0\right) = 1 - (1-\widetilde{q}_*)^k \tag{28}$$

Furthermore, given survival up until time t, $\widetilde{L}_t^{(k)}/t$ converges in probability to the constant $\frac{1}{2}$, and in addition the average value of $\widetilde{L}_t^{(k)}/t$ also tends to $\frac{1}{2}$ (ie. $\mathbb{E}(\widetilde{L}_t^{(k)}/t | N_t > 0) \to \frac{1}{2}$.

Initially, it may seem surprising that the longest pendant edge in a large complete birth–death tree is roughly $t(1-\rho)/2$ and this changes to roughly $t/2$ for the reduced tree, just like for Yule trees. This is because, if we condition a birth–death process to survive, its reduced tree will look very close to being a Yule tree with a branching rate $\lambda(1-\rho)$, at least until near to the end time t (in the reduced tree, lineages undergo binary branching at a rate of $\lambda(1-p_{t-s})$ at time s, but $p_{t-s} \to \rho$ whenever $\lambda(t-s) \to \infty$).

Sampling at the present

Suppose that, in addition to a birth–death process (with rates λ and μ respectively), a proportion σ of the leaves are randomly sampled at the present. From Stadler (2009), the reduced tree on this pruned leaf set has the same distribution as a reduced birth–death tree with modified birth and death rates, as given by the following relationships:

$$\lambda' = \sigma\lambda \text{ and } \mu' = \mu - \lambda(1-\sigma). \tag{29}$$

In order that $\mu' \geqslant 0$ one requires that $\mu \geqslant \lambda(1-\sigma)$. Thus the results on long edges in birth–death trees can be extended to certain settings that involve sampling at the present.

Simulation results

We used the *Treesim R* package (Stadler (2011); sim.bd.age function) to simulate 500 Yule trees (with $\lambda = 0.416, \mu = 0$, $t = 10$) giving a target size of $n = 64$ extant species. We also simulated 500 complete birth-death trees (with $\lambda = 1.11$, $\mu = 0.5\lambda = 0.555$, $t = 10$) retaining only trees with 1 or more extant species, giving a target size of $n = 513$ (since N_t, conditioned on non-extinction, has a geometric distribution with expected value $1/q_t$). We created a third set of 500 reduced birth-death trees by pruning all extinct species from the second set of complete birth-death trees using the *geiger* package (Pennell et al. (2014); drop.tip function).

We then calculated pendant edge lengths for all trees in each set using the ape, picante, phytools and geiger packages (Kembel et al., 2010; Revell, 2012; Pennell et al., 2014; Paradis and Schliep, 2019) and identified the longest pendant edge for each tree (denoted in this section by $\hat{L}_t$). These observed $\hat{L}_t$ values were then compared to the theoretical expected values of L_t/t (namely, $\frac{1}{2}$, $\frac{1-\mu/\lambda}{2}$ and $\frac{1}{2}$) for the Yule, complete and reduced birth–death trees, respectively. The results are presented in Fig. 3.

Overall, $\hat{L}_t/t$ values are large, and in line with the theoretical predictions (note that the dashed lines refer to expected values of $\hat{L}_t/t$ in the limit as the expected size of the

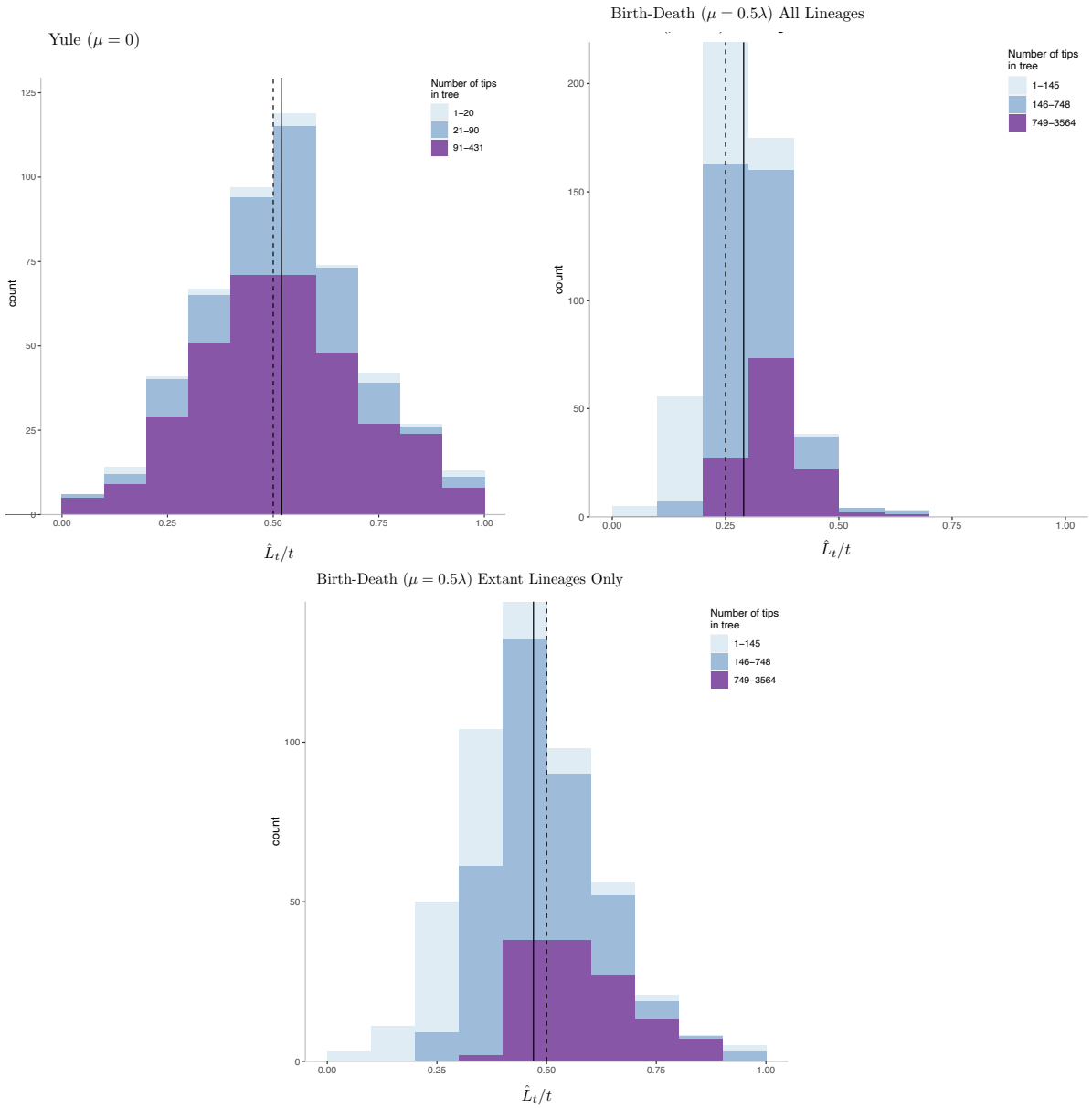


Fig. 3. The longest pendant edges (that are still alive at time t) on simulated birth-death trees, with $\hat{L}_t/t$ on the horizontal axis, and with trees sorted into three bins of small, medium and large trees. *Top Left:* The longest pendant edges (as $\hat{L}_t/t$) for 500 Yule trees ($\lambda = 0.416, \mu = 0$) and depth $t = 10$. The dotted vertical line at 0.5 indicates the expected value of $\hat{L}_t/t$ in the large tree limit (i.e. as $\lambda t \to \infty$). *Top Right:* $\hat{L}_t/t$ for complete birth-death trees ($\lambda = 1.11$, $\mu = 0.5\lambda = 0.555$, $t = 10$). The dotted vertical line at 0.25 ($= (1-\rho)/2$) indicates the expected value of $\hat{L}_t/t$ in the large tree limit and the solid line at 0.29 is the average value of $\hat{L}_t/t$. *Bottom:* $\hat{L}_t/t$ for reduced birth-death trees ($\lambda = 1.11, \mu = 0.555$, $t = 10$). The dotted vertical line at 0.5 indicates the expected value of $\hat{L}_t/t$ in the large tree limit and the solid line at 0.47 indicates the average value of $\hat{L}_t/t$.

trees tends to infinity (which, for ρ fixed, is equivalent to the limit as $\lambda t \to \infty$)). Moreover, larger trees (having more leaves than expected) tended to have even longer-than-expected longest pendant edges, especially for the complete trees with extinction. This observation may reflect the "push of the past" phenomenon (Phillimore and Price, 2008) whereby many early splits occur in clades that persist and become large, a phenomenon that will both produce short stem edges and more room for an early lineage to persist to become a pendant edge. To the extent that the push of the past is real, this might make our expectations somewhat conservative (though see below for data from mammal families).

Relevance to phylogenetic diversity

In biodiversity conservation, given a phylogenetic tree on a leaf set X of extant species, the *Fair Proportion (FP) index* (often called the 'evolutionary distinctiveness' index) is a way to assign the total sum of edge lengths across the tree 'fairly' to each of the extant species (Isaac et al., 2007; Redding, 2003; Redding et al., 2008). More precisely, suppose we have a phylogenetic tree T on the leaf set X, with edge lengths. We will assume (as in the rest of this paper and in most applications of FP), that these edge lengths correspond to (or are proportional to) time, and thus the sum of the lengths from the root to each leaf is the same (the so-called 'ultrametric' condition).

For a leaf x of T, let $e_x^i, i = 1, \ldots, k_x$ denote the edges on the directed path from x back to the root, let ℓ_x^i be the length of edge e_x^i, and let $n_i(x)$ be the number of leaves of T that are separated from the root of T by e_x^i (i.e. the number of leaves of T descended from e_x^i). For a leaf x of a phylogenetic tree T with edge lengths ℓ, the FP index of x, denoted $FP_{(T,\ell)}(x)$ (or, more briefly, $FP(x)$), is then defined as follows:

$$FP_{(T,\ell)}(x) := \sum_{i=1}^{k_x} \ell_x^i / n_i(x). \tag{30}$$

Since $n_1(x)$ is always equal to 1, we can re-write Eqn. (30) as follows:

$$FP(x) = \ell_x^1 + \epsilon(x), \tag{31}$$

where:

$$\epsilon(x) = \sum_{i>1}^{k_x} \ell_x^i / n_i(x).$$

Two key features of the FP index are that (i) summing $FP(x)$ over all the species x in X gives the total sum of edge lengths of the tree, and (ii) $FP(x)$ has an equivalent description in terms of the Shapley value in cooperative game theory (Fuchs and Jin, 2015).

Fair proportion is a measure of evolutionary non-redundancy. More precisely, it provides a measure of the extent of sharing the products of evolution across species (a more formal justification of this statement under a model in which features arise at most once and are retained is described in Wicke et al. (2021)). An early empirical observation from (Redding et al., 2008) is that, on average, $\frac{1}{2}$ of the value of FP comes from the pendant edge (i.e.the value ℓ_x^1) for Yule trees. Thus, species on very long pendant edges would be expected to be those with more truly unique features.

We can formalise this earlier empirical finding as follows. For a Yule tree T grown for time t at rate λ and a leaf x selected uniformly at random from those present at time t the following equation holds:

$$\mathbb{E}\left[\ell_x^1\right] = \frac{1}{2}(1 + e^{-\lambda t}) \cdot \mathbb{E}\left[FP_{(T,\ell)}(x)\right]. \tag{32}$$

In particular, as λt grows, the ratio of the expected pendant edge length to expected FP value for leaf x (i.e. $\mathbb{E}\left[\ell_x^1\right] / \mathbb{E}\left[FP_{(T,\ell)}(x)\right]$) converges rapidly towards $\frac{1}{2}$. The proof of Eqn. (32) is provided in the Appendix.

We now present a result that states, roughly speaking, that in large reduced birth–death trees with $\lambda > \mu$, a species with the highest FP score is expected to lie at the end of a pendant edge that has length close to $t/2$. One can easily construct trees where this expectation does not hold, so this prediction is not for trees in general, but rather for trees that have shapes captured by the birth–death model. The proof of the following theorem is provided in the Appendix.

Theorem 3 For any $\epsilon > 0$, and λ, μ fixed ($\lambda > \mu$), consider a reduced birth-death tree T grown for time t. Then with probability tending to 1 as λt grows (with $\rho = \mu/\lambda$ fixed), any species x that maximises $FP(x)$ is the endpoint of a pendant edge of length at least $(1 - \epsilon) \cdot \frac{t}{2}$.

Mammal families conform to edge length predictions

We identified the longest pendant edges across 114 mammal families with $\geqslant 3$ species (Burgin et al., 2018; Upham et al., 2018). A cursory look suggests that several of the results above are predictive for empirical trees.

Firstly, the length of the longest pendant edge is a significant proportion of t (the time back to the origin of the clade) across a wide range of tree sizes, as summarised in Table 1 and the accompanying histogram. The values of $\hat{L}_t/t$ are less than 0.5 but appear to be increasing toward this value with increasing tree sizes, and the variation in $\hat{L}_t/t$ is also declining with tree sizes, both of which are consistent with Theorem 2.

These data also support a strong connection between pendant edge length and the FP index of evolutionary distinctiveness: fully 95 (>80%) have the highest-ranking FP species also being the species on the longest pendant edge. Importantly, this pattern is not driven by small-sized families; 26/30 clades with 3 to 8 species, 46/57 clades with 9 to 38 species, and 23/27 families with 39–767 species show this pattern.

We make no claims that taxa generally diversify under simple constant-rate birth-death processes – for instance, the full mammal phylogeny we use shows evidence consistent with multiple diversification rate shifts within Families (Upham et al., 2021). However, and interestingly, Morlon et al. (2010) found that fully 30% of 289 sampled phylogenies had shapes consistent with Yule expectations, and a further 35% had shapes consistent with a slowing diversification rate through time, an expectation consistent with both tree-inference biases (Revell et al., 2005) and speciation theory (see, e.g. Etienne and Rosindell (2012)). Taken together, this suggests that both real and expected trees are

likely to contain particularly long pendant edges.

Table 1. The ratio $\hat{L}_t/t$ across phylogenetic trees for 114 mammal families

Number of tips	Number of trees	mean of $\hat{L}_t/t$	median of $\hat{L}_t/t$	coefficient of variation of $\hat{L}_t/t$
$3-7$ species	30	0.253	0.180	0.714
$8-38$ species	57	0.340	0.313	0.539
$39-767$ species	27	0.464	0.435	0.435

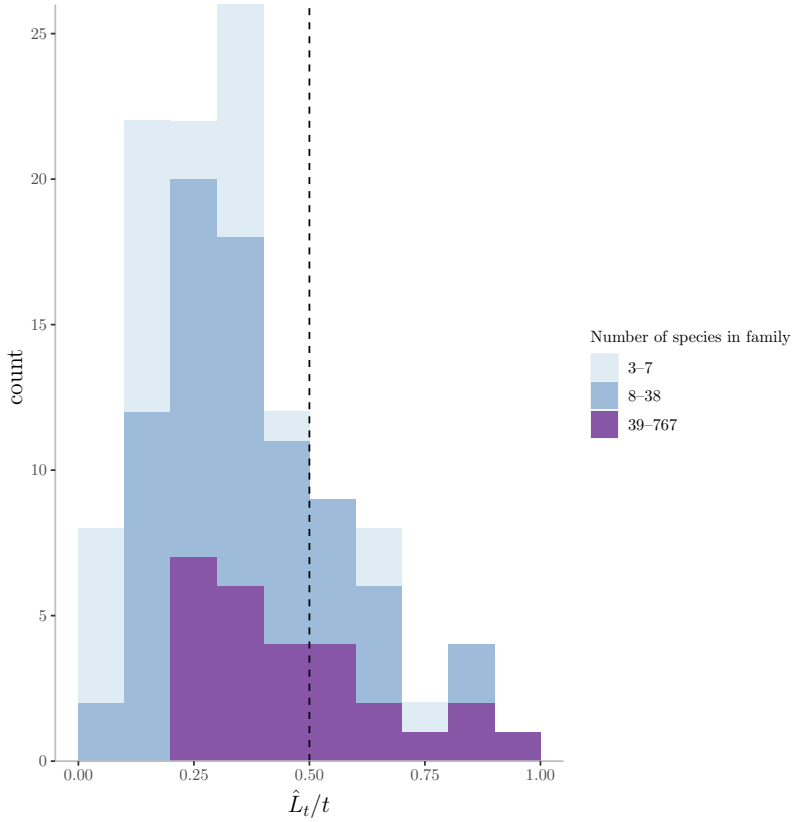


Fig. 4. A histogram of the values $\hat{L}_t/t$ across 114 mammal families, broken up into the three groups (3-7 species, 8-38 species and 39+ species) distinguished by separate colours (stacked on each other) and with further statistical details provided in Table 1.

Relevance for the amount of sequence data required to accurately infer a fully resolved tree

Both too-short and too-long edges can decrease the probability of correctly inferring a phylogenetic tree. In this section, we describe the impact of the interplay of long and short edges on the number of aligned DNA sequence sites required to accurately infer a phylogenetic tree from sequence data. The evolution of an aligned DNA sequence site is typically modelled by a continuous-time Markov process on a finite state space (typically the 4-element nucleotide set $\{A, C, G, T\}$) operating along the edges of the tree. Suppose that each site evolves independently along the edges of the tree with a fixed substitution rate ν per site (where μ is constant across the edges and sites). We wish to consider the number of sites K required to infer T correctly with a given high probability.

Note that K depends on the tree, the edge lengths, the model of site substitution and the desired accuracy of tree reconstruction. K also depends on the tree reconstruction method applied, and so we consider here the usual form of maximum likelihood estimation without rate variation across sites (i.e. the substitution rate can vary across the tree but not across sites, and these along with the edge lengths are treated as nuisance parameters for the reconstruction). In particular, the inference of the tree does not include specifying the location of the root vertex.

It has been shown that for any binary phylogenetic tree with n leaves that has 'carefully chosen' edge lengths, K can grow as slowly as $\log(n)$ at least for simple site substitution models; a somewhat surprising and nontrivial result due to Daskalakis et al. (2011) (see also (Mossel et al., 2011; Mossel and Steel, 2004)). However, as noted by Felsenstein (2004) (pp. 173–174), edge lengths are likely to be variable and depend on the number of leaves of a tree (n) so a logarithmic dependence of K on n should lead to a faster growth function with n. Here, we describe a lower bound on K that is a positive power of n (rather than logarithmic in n), when the tree is generated by the Yule model (thus, there are now two random processes at play – the generation of the tree and its edge

lengths, and the evolutions of sites on that tree).

Proposition 3 Consider a Yule tree T grown for time t and a Markovian site-substitution model with site-substitution rate ν operating on this tree. The sequence length K required to accurately reconstruct T from its associated sequence of sites at the leaves is bounded below by a term of order N_t^γ as $\lambda t \to \infty$, where $\gamma > 0$ and N_t is the number of leaves of the tree. Moreover, a lower bound of this type can be made independent of the value of ν (for a possibly smaller but strictly positive value of γ).

The proof of Proposition 3, given in the Appendix, combines our earlier result on the longest pendant edge length in a Yule tree T with the following additional result concerning the length of the *shortest* interior edge in T.

Proposition 4 Let S_t^o denote the length of the shortest interior edge in a Yule tree grown for time t.

$$\mathbb{P}\left(S_t^o \geqslant xe^{-\lambda t}\right) \to \frac{1}{1+2\lambda x}, \text{ as } \lambda t \to \infty. \tag{33}$$

The proof of Proposition 4 is provided in the online Supplementary Material.

Concluding comments

The simplest constant-rate birth-death process continues to offer up surprising tree shapes. While we know that any particular tree shape can be created under an infinite number of time-variant histories (Louca and Pennell, 2020), the constant-rate model is the default for species-level phylogenetic inference (e.g. in the popular RevBayes and BEAST packages); simple tree statistics based on the birth-death process (e.g. the methods of moments estimator of diversification rate; Magallón and Sanderson (2001)) have predictive power (see, e.g. Greenberg et al. (2021)); and limited surveys of the shapes of inferred trees and alternative models of diversification suggest the process we model here might

underestimate pendant edge lengths in particular (viz. Morlon et al. (2010)). Remarkably old lineages (e.g. living fossils such as the tuatara, bichir, *Welwitschia* and *Amborella*) are well-known. The stochastic nature of speciation and extinction as we investigated here suggests some of these may (just) be expected statistical outliers (viz. Liow (2007)). And to the extent that the models capture the variation in edge lengths expected in large clades, we can expect our phylogenetic inferences to only slowly converge on the underlying true tree.

We also note that in the simulations (Fig. 3) as well as the mammal data (Table 1 and Fig. 4), trees with larger numbers of leaves tend to have larger values of $\hat{L}_t/t$ than trees with fewer leaves. In the theoretical results in our paper we are not conditioning on n (the number of leaves) apart from insisting that $n \geqslant 1$. Now, it might be expected that, for fixed times t, trees that have more leaves than the expected value (e.g. $e^{\lambda t}$ for the Yule process) should have pendant edges that are slightly shorter on average, since the height of the tree is fixed, but the number of edges is increased (*cf.* Theorem 2 and Table 2 of Mooers et al. (2012)). However, our focus in this paper has been on the 'longest' pendant edge length (rather than the average), and increasing n may allow for more opportunity of an 'outlier' (extra-long) pendant edge to be present. For the mammal families, of course, t is not fully fixed and diversification rates vary – larger families may either have stochastically larger pushes of the past, or be governed by lower μ or higher λt values.

Finally, it would be of interest to investigate the length of the longest (and shortest) edges in phylogenetic diversification models in which the birth and death rates (λ, μ) are not treated as constants, but allowed to depend on time t, or perhaps on other (stochastic) aspects of the branching process; for example the number of other lineages present in the tree at the given time, or the 'age' of the lineage (the time back to when it first split off from another lineage).

Supplementary Material

The comparative data (mammal family trees) used in this study were retrieved from previously published phylogenies (Upham et al., 2018). A set of 500 global mammal phylogenies was downloaded from VertLife.org: http://vertlife.org/phylosubsets/. Additional mathematical results are available in the online Supplementary Material file.

Acknowledgements

AOM and EK thank the Natural Sciences and Engineering Research Council (NSERC) Discovery Grant (AOM), NSERC PGSM scholarship (EK), and the NSERC CREATE program ('RenewZoo' training grant; AOM and EK). SB is supported by NSFC grant (No.11731012) and MS thanks the NZ Marsden Fund (MFP-UOC2005) for funding support. We thank Dr Paul Lewis and a second (anonymous) reviewer for several helpful comments and suggestions on an earlier version of this manuscript.

References

Aldous, D. 1996. Probability distributions on cladograms. Pages 1–18 *in* Random Discrete Processes (Eds. D Aldous and R. Pemantle) vol. 76 of *IMA volumes in Mathematics and its Applications*. Springer.

Aldous, D. and L. Popovic. 2005. A critical branching process model for biodiversity. Adv. Appl. Probab. 37:1094–1115.

Aldous, D. J. 2001. Stochastic models and descriptive statistics for phylogenetic trees, from Yule to today. Statist. Sci. 16:23–34.

Burgin, C. J., J. P. Colella, P. L. Kahn, and N. S. Upham. 2018. How many species of mammals are there? J. Mammal. 99:1–14.

Daskalakis, C., E. Mossel, and S. Roch. 2011. Evolutionary trees and the Ising model on the Bethe Lattice: A proof of Steel's conjecture. Prob. Theor. Rel. Fields 149:149–189.

Etienne, R. S. and J. Rosindell. 2012. Prolonging the past counteracts the pull of the present: Protracted speciation can explain observed slowdowns in diversification. Systematic Biology 61:204–213.

Felsenstein, J. 2004. Inferring Phylogenies. Sinauer Associates, Sunderland, MA.

Fuchs, M. and E. Y. Jin. 2015. Equality of Shapley value and fair proportion index in phylogenetic trees. J. Math. Biol. 71:1133–1147.

Gascuel, O. and M. Steel. 2010. Inferring ancestral sequences in taxon-rich phylogenies. Math. Biosci. 227:125–135.

Gavrilets, S. and A. Vose. 2005. Dynamics of adaptive radiation. Proc. Natl. Acad. Sci. U.S.A. 102:18040–18045.

Greenberg, D. A., R. A. Pyron, L. G. W. Johnson, N. S. Upham, W. Jetz, and A. O. Mooers. 2021. Evolutionary legacies in contemporary tetrapod imperilment. Ecol. Lett. 24:2464–2476.

Grimmett, G. and D. Stirzaker. 2001. Probability and Random Processes (3rd ed.). Oxford University Press.

Hey, J. 1992. Using phylogenetic trees to study speciation and extinction. Evolution 46:627–640.

Isaac, N. J. B., S. T. Turvey, B. Collen, C. Waterman, and J. E. M. Baillie. 2007. Mammals on the EDGE: Conservation priorities based on threat and phylogeny. PLoS ONE 2:e296.

Kembel, S. W., P. D. Cowan, M. R. Helmus, W. K. Cornwell, H. Morlon, D. D. Ackerly, S. P. Blomberg, and C. O. Webb. 2010. Picante: R tools for integrating phylogenies and ecology. Bioinformatics 26:1463–1464.

Kendall, D. G. 1948. On the generalized ‘birth-and-death’ process. Ann. Math. Statist. 19:1–15.

Lambert, A. and T. Stadler. 2013. Birth–death models and coalescent point processes: the shape and probability of reconstructed phylogenies. Theor. Popul. Biol. 90:113–128.

Liow, L. H. 2007. Lineages with long durations are old and morphologically average: an analysis using multiple datasets. Evolution 61:885–901.

Louca, S. and M. W. Pennell. 2020. Extant timetrees are consistent with a myriad of diversification histories. Nature 580:502–505.

Magallón, S. and M. J. Sanderson. 2001. Absolute diversification rates in angiosperm clades. Evolution 55:1762–1780.

Mooers, A. O., O. Gascuel, T. Stadler, H. Li, and M. Steel. 2012. Branch lengths on birth–death trees and the expected loss of phylogenetic diversity. Syst. Biol. 61:195–203.

Morlon, H., M. D. Potts, and J. B. Plotkin. 2010. Inferring the dynamics of diversification: a coalescent approach. PLOS Biol. 8:e1000493.

Mossel, E., S. Roch, and A. Sly. 2011. On the inference of large phylogenies with long branches: How long is too long? Bull. Math. Biol. 73:1627–1644.

Mossel, E. and M. Steel. 2004. A phase transition for a random cluster model on phylogenetic trees. Math. Biosci. 187:189–203.

Nee, S., E. C. Holmes, R. M. May, and P. H. Harvey. 1994. Extinction rates can be estimated from molecular phylogenies. Philos. Trans. R. Soc. Lond. B Biol. Sci. 344:77–82.

Paradis, E. and K. Schliep. 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses and evolutionary analyses in R. Bioinformatics 35:526–528.

Pennell, M. W., J. M. Eastman, G. J. Slater, J. W. Brown, J. C. Uyeda, R. G. FitzJohn, M. E. Alfaro, and L. J. Harmon. 2014. geiger v2.0: an expanded suite of methods for fitting macroevolutionary models to phylogenetic trees. Bioinformatics 30:2216–2218.

Phillimore, A. B. and T. D. Price. 2008. Density-dependent cladogenesis in birds. PLOS Biol. 6:483–489.

Redding, D. W. 2003. Incorporating genetic distinctness and reserve occupancy into a conservation prioritisation approach. Masters Thesis, University Of East Anglia, Norwich, UK .

Redding, D. W., K. Hartmann, A. Mimoto, D. Bokal, M. Devos, and A. O. Mooers. 2008. Evolutionarily distinctive species often capture more phylogenetic diversity than expected. J. Theor. Biol. 251:606–615.

Revell, L. J. 2012. phytools: an R package for phylogenetic comparative biology (and other things). Methods Ecol. Evol. 3:217–223.

Revell, L. J., L. J. Harmon, and R. E. Glor. 2005. Underparameterized model of sequence evolution leads to bias in the estimation of diversification rates from molecular phylogenies. Systematic Biology 54:973–983.

Stadler, T. 2009. On incomplete sampling under birth–death models and connections to the sampling-based coalescent. J. Theor. Biol. 261:58–66.

Stadler, T. 2011. Simulating trees with a fixed number of extant species. Syst. Biol 60:676–684.

Stadler, T. and M. Steel. 2012. Distribution of branch lengths and phylogenetic diversity under homogeneous speciation models. J. Theor. Biol. 297:33–40.

Steel, M. 2016. Phylogeny: Discrete and Random Processes in Evolution. Society for Industrial and Applied Mathematics, Philadelphia PA.

Steel, M. and A. Mooers. 2010. The expected length of pendant and interior edges of a Yule tree. Appl. Math. Lett. 23:1315–1319.

Upham, N. S., J. A. Esselstyn, and W. Jetz. 2018. Inferring the mammal tree: Species-level sets of phylogenies for questions in ecology, evolution, and conservation. PLoS Biol. 17:e3000494.

Upham, N. S., J. A. Esselstyn, and W. Jetz. 2021. Molecules and fossils tell distinct yet complementary stories of mammal diversification. Curr. Biol. 31:4195–4206.e3.

Wicke, K., A. Mooers, and M. Steel. 2021. Formal links between feature diversity and phylogenetic diversity. Syst. Biol. 70:480–490.

Yule, G. U. 1925. A mathematical theory of evolution: Based on the conclusions of Dr. J.C. Willis F.R.S. Philos. Trans. R. Soc. Lond. B 213:21–87.

Appendix: Mathematical details and proofs

Part One: Yule tree

Proof of Proposition 2. We have:

$$\mu_t = \int_0^t \mathbb{P}(L_t > x)dx.$$

Thus, by Proposition 1 (Eqn. (1)):

$$\mu_t = \int_0^t \frac{e^{\lambda(t-2x)}}{1 - e^{-\lambda x} + e^{\lambda(t-2x)}} dx, \tag{34}$$

Making the substitution $\zeta = \lambda(x - t/2)$, this last integral can be written as:

$$\frac{1}{\lambda} \int_{-\lambda t/2}^{\lambda t/2} \frac{e^{-2\zeta}}{1 - e^{-\lambda t/2} e^{-\zeta} + e^{-2\zeta}} d\zeta.$$

By making the further substitution $\xi = e^{-\zeta}$, we obtain:

$$\frac{1}{\lambda} \int_{e^{-\lambda t/2}}^{e^{\lambda t/2}} \frac{\xi}{1 - e^{-\lambda t/2}\xi + \xi^2} d\xi. \tag{35}$$

We now apply a standard integral result:

$$\int \frac{xdx}{ax^2 + bx + c} = \frac{1}{2a} \ln|ax^2 + bx + c| - \frac{b}{a\sqrt{4ac - b^2}} \tan^{-1} \frac{2ax + b}{\sqrt{4ac - b^2}},$$

provided that $4ac - b^2 > 0$. Setting $a = c = 1, b = -e^{-\lambda t/2}$ and $x = \xi$, Expression (35) becomes:

$$\frac{1}{2\lambda}\left[\ln(e^{\lambda t}) - \ln(1)\right] + \frac{e^{-\lambda t/2}}{\lambda\sqrt{4 - e^{-\lambda t}}}\left[\tan^{-1}\left(\frac{2e^{\lambda t/2} - e^{-\lambda t/2}}{\sqrt{4 - e^{-\lambda t}}}\right) - \tan^{-1}\left(\frac{e^{-\lambda t/2}}{\sqrt{4 - e^{-\lambda t}}}\right)\right]$$

Finally, since $\frac{1}{2\lambda}\left[\ln(e^{\lambda t}) - \ln(1)\right] = t/2$, we obtain the claimed expression. The claimed limit as $\lambda t \to \infty$ now follows.

□

Part two: Complete birth–death tree

Proof of Lemma 2. Let N_t be the number of lineages alive at time t in the birth–death process. By Lemma 1, N_t has a *modified geometric distribution* on $\{0, 1, 2, 3, \dots\}$, where:

$$\mathbb{P}(N_t = i) = \begin{cases} p_t & (k = 0) \\ (1 - p_t)q_t(1 - q_t)^{k-1} & (k \geqslant 1) \end{cases}$$

where p_t and q_t are as given in Lemma 1. We write $N_t \sim \mathrm{ModGeom}(p_t, q_t)$. In particular, the average number alive at time t is $\mathbb{E}N_t = e^{(\lambda-\mu)t}$, and the survival probability is

$$\mathbb{P}(N_t > 0) = \frac{(1-\rho)}{1-\rho e^{-(\lambda-\mu)t}}.$$

Note, the supercritical case corresponds to $\rho \in [0, 1)$ and in that case $\mathbb{P}(N_t > 0) \to 1 - \rho$ as $t \to \infty$.

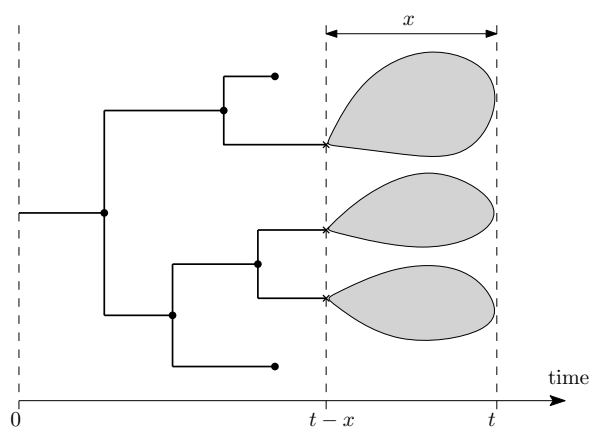


Fig. 5. Subtrees of the birth–death tree $\mathcal{T}_t$ initiated at time $t - x$ behave like N_{t-x} independent birth–death trees each distributed like $\mathcal{T}_x$. A pendant edge at time t of length greater than x must have been present at time $t - x$ and not have any other birth or death event occur over remaining time of length x, that is, a subtree consisting only of a single edge.

Observe that every pendant edge at time t that has length greater than x must have both been present at time $t - x$ and also not have split into two lineages during the remaining time period $(t - x, t]$. Since each of the N_{t-x} lineages alive at time $t - x$ evolve forward in time independently, and (using the memoryless property of exponential lifetimes) each has an independent probability of success of $e^{-(\lambda+\mu)x}$ to give rise to a pendant edge of length at least x by not having any split or extinction event over the remaining time. Thus, conditional on N_{t-x}, we have $N_t^x \sim \mathrm{Bin}(N_{t-x}, e^{-(\lambda+\mu)x})$.

A binomial random variable $Z \sim \mathrm{Bin}(N, r)$, where each of the N independent trials has a probability r of success but the total number of trials $N \sim \mathrm{ModGeom}(p, q)$ is an independent random variable, also gives rise to a modified geometric distribution for the

total number of successes, where:

$$\mathbb{P}(Z=0) = p + (1-p)\frac{(1-r)q}{1-(1-q)(1-r)} = \frac{pr+(1-r)q}{r+(1-r)q}$$
$$\mathbb{P}(Z=k|Z>0) = s(1-s)^{k-1} \quad (k \geqslant 1)$$

where $s := q/(1-(1-q)(1-r))$.

Simplifying these parameters above for the special case when $q = q_{t,x,\lambda,\mu}$, $p = p_{t,x,\lambda,\mu}$, and $r = e^{-(\lambda+\mu)x}$ gives the claimed distribution for N_t^x. □

Proof of Theorem 1. This result largely follows as a corollary to the explicit distribution for N_t^x given in Lemma 2. Recalling the formulae for $p_{t,x,\lambda,\mu}$ and $q_{t,x,\lambda,\mu}$ given in Eqns. (8) and (9) respectively, setting $x = \frac{t}{2}(1-\rho) + \frac{y}{\lambda}$ and then letting $\lambda t \to \infty$, we find:

$$p_{t,\frac{t}{2}(1-\rho)+\frac{y}{\lambda},\lambda,\mu} \to p^* := \frac{1-\rho+\rho e^{-2y}}{1-\rho+e^{-2y}}, \quad q_{t,\frac{t}{2}(1-\rho)+\frac{y}{\lambda},\lambda,\mu} \to q^* := \frac{1-\rho}{1-\rho+e^{-2y}}.$$

We also observe that $(1-p^*)/(1-\rho) = 1-q^*$ and recall that $\mathbb{P}(N_t > 0) = 1 - p_t \to 1-\rho$ as $\lambda t \to \infty$. Then, for $k \geqslant 1$:

$$\mathbb{P}\left(N_t^{\frac{t}{2}(1-\rho)+\frac{y}{\lambda}} = k \,\middle|\, N_t > 0\right) = \frac{\mathbb{P}\left(N_t^{\frac{t}{2}(1-\rho)+\frac{y}{\lambda}} = k\right)}{\mathbb{P}(N_t > 0)} \to \frac{(1-p^*)q^*(1-q^*)^{k-1}}{1-\rho} = q^*(1-q^*)^k$$

by using (10), conditional probability, and noting that $N_t > 0$ is guaranteed whenever $N_t^x \geqslant 1$. Similarly, also noting $\{N_t > 0\}^c = \{N_t = 0\} \subseteq \{N_t^x = 0\}$, we find

$$\begin{aligned}\mathbb{P}\left(N_t^{\frac{t}{2}(1-\rho)+\frac{y}{\lambda}} = 0 \,\middle|\, N_t > 0\right) &= \frac{\mathbb{P}\left(N_t^{\frac{t}{2}(1-\rho)+\frac{y}{\lambda}} = 0; N_t > 0\right)}{\mathbb{P}(N_t > 0)} \\ &= \frac{\mathbb{P}\left(N_t^{\frac{t}{2}(1-\rho)+\frac{y}{\lambda}} = 0\right) - \mathbb{P}\left(N_t = 0\right)}{\mathbb{P}(N_t > 0)} \\ &\to \frac{p^*-\rho}{1-\rho} = q^*,\end{aligned}$$

yielding Part (i). For Part (ii), we need simply note that:

$$\mathbb{P}\left(L_t - \frac{t}{2}(1-\rho) \leqslant \frac{y}{\lambda} \,\middle|\, N_t > 0\right) = \mathbb{P}\left(N_t^{\frac{t}{2}(1-\rho)+\frac{y}{\lambda}} = 0 \,\middle|\, N_t > 0\right). \tag{36}$$

For Part (iii), we similarly observe that for $k \geqslant 1$,

$$\mathbb{P}\left(L_t^{(k)} - \frac{t}{2}(1-\rho) > \frac{y}{\lambda} \,\middle|\, N_t > 0\right) = \mathbb{P}\left(N_t^{\frac{t}{2}(1-\rho)+y} > k-1 \,\middle|\, N_t > 0\right) \to (1-q^*)^k$$

It remains to show the convergence in means. The arguments for Parts (ii)-(iii) are identical, so we discuss only (ii). As a simple consequence of (36), L_t/t converges in distribution to the constant $(1-\rho)/2$ whenever $\lambda t \to \infty$, that is,

$$\mathbb{P}\left(\frac{L_t}{t} \leqslant \frac{1}{2}(1-\rho) + x \,\middle|\, N_t > 0\right) \to \begin{cases} 1 & (x > 0) \\ 0 & (x < 0). \end{cases}$$

Now, convergence in distribution to a *constant* also implies convergence in probability. In other words, for all $\epsilon > 0$:

$$\mathbb{P}\left(\left|\frac{L_t}{t} - \frac{1}{2}(1-\rho)\right| < \epsilon \,\middle|\, N_t > 0\right) \to 1 \quad \text{as } \lambda t \to \infty.$$

Finally, since $L_t/t \in [0,1]$, we can use the bounded convergence theorem to deduce that $\mathbb{E}(\frac{L_t}{t}|N_t > 0) \to (1-\rho)/2$ as $\lambda t \to \infty$.

In addition to the convergence in means in the theorem above, we note that some finer almost sure convergence results also hold for $L_t^{(k)}/t$. In particular, $\frac{L_t}{t} \to \frac{1}{2}(1-\rho)$ as $t \to \infty$ almost surely (i.e. $\mathbb{P}\left(\frac{L_t}{t} \to \frac{1}{2}(1-\rho) \text{ as } t \to \infty \,\middle|\, N_t > 0 \text{ for all } t \geqslant 0\right) = 1$). However, these are omitted as they would require a substantial additional analysis beyond the scope of the present article.

□

Part 3: Reduced tree

Definition of $\widetilde{p}_{t,x,\lambda,\mu}$ and $\widetilde{q}_{t,x,\lambda,\mu}$

For $x \in [0,t]$, let:

$$\begin{aligned} \widetilde{p}_{t,x,\lambda,\mu} &:= \frac{p_{t-x}(1-p_x)q_x + (1-(1-p_x)q_x)q_{t-x}}{q_{t-x} + (1-q_{t-x})(1-p_x)q_x} \\ &= \frac{(1-\rho e^{-(\lambda-\mu)x})^2 + (1-\rho)(1-2\rho)e^{-(\lambda-\mu)x} + \rho(1-\rho)e^{(\lambda-\mu)(t-2x)}}{(1-\rho e^{-(\lambda-\mu)x})^2 - (1-\rho)e^{-(\lambda-\mu)x} + (1-\rho)e^{(\lambda-\mu)(t-2x)}} \end{aligned} \tag{37}$$

and

$$\begin{aligned}\widetilde{q}_{t,x,\lambda,\mu} :&= \frac{q_{t-x}}{q_{t-x} + (1-q_{t-x})(1-p_x)q_x} \\ &= \frac{(1-\rho e^{-(\lambda-\mu)x})^2}{(1-\rho e^{-(\lambda-\mu)x})^2 - (1-\rho)e^{-(\lambda-\mu)x} + (1-\rho)e^{(\lambda-\mu)(t-2x)}}\end{aligned} \tag{38}$$

Proof of Lemma 3. We proceed by modifying the proof of Lemma 2. Firstly, recall that $N_t \sim \mathrm{ModGeom}(p_t, q_t)$. Observe also that every pendant edge in the reduced tree at time t that has length greater than x must have come from some individual present at time $t-x$ that has *exactly* one descendant alive after additional time x elapses. Now, each of the N_{t-x} lineages alive at time $t-x$ evolve forward in time independently, and each has a probability of success of $\mathbb{P}(N_x = 1) = (1-p_x)q_x$ of giving rise to a pendant edge in the reduced tree at time t having length at least x. Thus, conditional on N_{t-x}, we have $\widetilde{N}_t^x \sim \mathrm{Bin}(N_{t-x}, (1-p_x)q_x)$.

In general, if $N \sim \mathrm{ModGeom}(p,q)$ and, conditional on N, we have $Z \sim \mathrm{Bin}(N, r)$, then we know that $Z \sim \mathrm{ModGeom}(\widetilde{p}, \widetilde{q})$, where:

$$\widetilde{p} = p + (1-p)\left(\frac{(1-r)q}{1-(1-q)(1-r)}\right) = \frac{pr + (1-r)q}{q + (1-q)r} \tag{39}$$

$$\widetilde{q} = \frac{q}{1-(1-q)(1-r)} = \frac{q}{q+(1-q)r}, \tag{40}$$

Simplifying these parameters above for the special case when $p = p_{t-x}$, $q = q_{t-x}$ and $r = (1-p_x)q_x$ gives the claimed distribution for $Z = \widetilde{N}_t^x$. □

Proof of Corollary 2. This follows directly from the distribution of $\widetilde{N}_t^x$ given in Lemma 3, since $\mathbb{P}(\widetilde{L}_t \leqslant x) = \mathbb{P}(\widetilde{N}_t^x = 0)$ and $\mathbb{P}(\widetilde{L}_t^{(k)} \leqslant x) = \mathbb{P}(\widetilde{N}_t^x \leqslant k-1)$. □

Proof of Theorem 2. The proof of this result is essentially the same as the proof of Theorem 1 for the complete tree. It follows directly from the explicit distribution for $\widetilde{N}_t^x$ given in Lemma 3 (instead of Lemma 2 in the complete tree analogue), combined with the limits given in (24), (25), (23), together with the simple observation that $(1-\widetilde{p}_*)/(1-\rho) = 1 - \widetilde{q}_*$. □

Proof of Eqn. (32)

Let $\mathcal{I}(t)$ (resp. $\mathcal{P}(t)$) denote the sum of the interior edges (resp. pendant edges) in a Yule tree grown for time t. Summing over the leaves of the tree that are present at time t we have $\sum_y FP_{(T,\ell)}(y) = \mathcal{I}(t) + \mathcal{P}(t)$ (since the sum of FP over all leaves is the total sum of edge lengths in the tree) and $\sum_y \ell_y^1 = \mathcal{P}(t)$. By Theorem 4 of Steel and Mooers (2010), $\mathbb{E}[\mathcal{I}(t)] = i(t)$ and $\mathbb{E}[\mathcal{P}(t)] = p(t)$, where $i(t) = \frac{1}{\lambda}(e^{\lambda t} + e^{-\lambda t} - 2)$ and $p(t) = \frac{1}{\lambda}(e^{\lambda t} - e^{-\lambda t})$. Observe that $p(t) = \frac{1}{2}(1 + e^{-\lambda t})(i(t) + p(t))$, and so

$$\mathbb{E}\left[\sum_y \ell_y^1\right] = \frac{1}{2}(1 + e^{-\lambda t}) \cdot \mathbb{E}\left[\sum_y FP_{(T,\ell)}(y)\right]. \tag{41}$$

Now consider a leaf x selected uniformly at random from the N leaves present at time t (note that N is a random variable). We have $\mathbb{E}[\ell_x^1|N] = \frac{1}{N}\mathbb{E}\left[\sum_y \ell_y^1\right]$ and $\mathbb{E}\left[FP_{(T,\ell)}(x)|N\right] = \frac{1}{N}\mathbb{E}\left[\sum_y FP_{(T,\ell)}(y)\right]$, and so, from Eqn. (41),

$$\mathbb{E}[\ell_x^1|N] = \frac{1}{2}(1 + e^{-\lambda t}) \cdot \mathbb{E}[FP_{(T,\ell)}(x)|N].$$

This equality holds for all values of N and so taking a further expectation (over N) yields Eqn. (32).

Proof of Theorem 3

The proof of Theorem 3 relies on the following two claims

(i) With probability tending to 1 as λt grows (with $\rho = \mu/\lambda$ fixed), every leaf x of the reduced birth-death tree T that has incident pendant edge length $\leqslant \frac{1}{16} \cdot (t/2)$ satisfies the inequality: $FP(x) \leqslant \frac{11}{12} \cdot (t/2)$.

(ii) Let $\delta \in (0, 1)$ be fixed, and consider a leaf x of the reduced birth-death tree T for which the incident pendant edge has length $\ell_x^1 \geqslant \delta t/2$. Then,

$$FP(x) = \ell_x^1 + o(1), \tag{42}$$

where $o(1)$ is a positive term that converges in probability to 0 as λt grows (with $\rho = \mu/\lambda$ fixed).

The proofs of Claims (i) and (ii) are provided shortly. First, we show how Theorem 3 follows from them. We may assume, without loss of generality, that $0 < \epsilon < \frac{1}{12}$. Let e' be a longest pendant edge in T and let x' be its end leaf. By Theorem 2, e' has length at least $(1-\epsilon)(t/2)$ with probability $1-o(1)$ (as t grows), and since $FP(x') \geqslant \ell^1_{x'}$ it follows that $FP(x') \geqslant (1-\epsilon)(t/2)$ with probability $1-o(1)$.

Next, consider any leaf x of T. If $\ell^1_x < \frac{1}{16}(t/2)$ then, by Claim (i), $FP(x) \leqslant \frac{11}{12}(t/2)$ with probability $1-o(1)$, and since $\epsilon < \frac{1}{12}$ it follows that $FP(x) < FP(x')$ so x cannot be a leaf that maximises FP. Thus any leaf that maximises FP satisfies $\ell_x \geqslant \delta(t/2)$ for $\delta = \frac{1}{16}$. It then follows from Claim (ii) that (for any such leaf x satisfying this last inequality) we have: $FP(x) = \ell^1_x + o(1)$, and so any leaf with maximal FP value will have $\ell^1_x \geqslant (1-\epsilon)(t/2)$ as λt grows (such a leaf x' exists as shown at the start of this proof). This completes the proof of Theorem 3 modulo verifying the two claims.

Proof of Claim (i): We first state a lemma, the proof of which is provided in the online Supplementary Material.

Lemma 4 For any $\epsilon > 0$, the probability the reconstructed birth-death tree has an interior edges of length greater than $(\frac{1}{2}+\epsilon)t$ decays exponentially fast to 0 as $\lambda t \to \infty$ (with $\rho = \mu/\lambda$ fixed).

Next, observe that for any leaf x, $FP(x) \leqslant \ell^1_x + \frac{1}{2}\ell^2_x + \frac{1}{3}\sum_{j>2}\ell^j_x$, and since $\sum_{j>2}\ell^j_x = t - \ell^1_x - \ell^2_x$ it follows that $FP(x) \leqslant \frac{2}{3}\ell^1_x + \frac{1}{6}\ell^2_x + \frac{1}{3}t$. In particular, if $\ell^1_x \leqslant \alpha \cdot \frac{t}{2}$ and $\ell^2_x \leqslant (1+\epsilon')\frac{t}{2}$ then:

$$FP(x) \leqslant \left(\frac{2}{3}(1+\alpha) + \frac{1}{6}(1+\epsilon')\right) \cdot \left(\frac{t}{2}\right). \tag{43}$$

Note that ℓ^2_x is an interior edge, and so the condition that $\ell^2_x \leqslant \frac{t}{2}(1+\epsilon')$ holds with probability $1-o(1)$ by Lemma 4. Taking $\alpha = 1/16$ and $\epsilon' = 1/4$ and we see that the right-hand side Inequality (43) is 11/12, which establishes Claim (i).

Proof of Claim (ii): Notice that if $n_2(x) \geqslant M$, then:

$$0 \leqslant \epsilon(x) \leqslant \ell^2_x/M + \ell^3_x/(M+1) + \cdots + \ell^{k_x}_x/(M+k_x-2) \leqslant (t-\ell^1_x)k_x/M. \tag{44}$$

Suppose that x is a tip of a pendant edge e of a reduced birth-death tree T grown for time t, and that e has length $\ell_x \geqslant \delta t/2$ for some constant $\delta > 0$. Consider the subtree of T descending from the other endpoint of edge e_1 to x. By the assumption that T has evolved according to a birth–death model, the tree topology is described by the Yule–Harding model, and thus the number of leaves of this subtree (i.e. $n_2(x) - 1$) is geometrically distributed with parameter $p = e^{-\ell_x^1}$. In particular, since $\ell_x^1 \geqslant \delta t/2$, it follows that $n_2(x)$ is at least $e^{(\lambda-\mu)\delta t/4}$ probability converging to 1 as t grows. Furthermore, since the topology of a reduced birth–death tree with n leaves is described by the Yule-Harding distribution (i.e. the β–splitting model with $\beta = 0$) the number of edges on the path from the root to a most distant leaf is concentrated around a term of order $\log(n)$ (Proposition 4 of Aldous (1996)) and thus k_x is of order $\log(n)$ which, by Jensen's inequality, grows at most linearly in expectation with λt. Consequently, from Inequality (44), $\epsilon(x)$ is (with high probability as λt grows) bounded above by a term of order $\lambda t^2 e^{-\lambda t/4}$, and so for an edge of length $\geqslant \delta t/2$,

$$FP(x) = \ell_x^1 + o(1),$$

where $o(1)$ refers to a positive term that converges in probability to 0 as t grows. This establishes Claim (ii).

We end this section by noting that for a randomly selected extant species x in a reconstructed birth–death tree, the equation $FP(x) = \ell_x^1 + o(1)$ (Eqn. 42) fails to hold. This is because the expected value of $\frac{1}{n_2(x)}$ for the Yule-Harding distribution (e.g. a birth–death tree) with n leaves is given (via Eqn. (3.8) in Steel (2016)) by:

$$\sum_{k=2}^{n-1} \frac{4}{k^2(k^2-1)} + \frac{2}{n^2(n-1)}. \tag{45}$$

By using partial fractions, and the identity $\sum_{k=1}^{\infty} k^{-2} = \pi^2/6$, the expression in (45) simplifies to $\beta + o(1)$, where $\beta = 7 - 2\pi^2/3 \approx 0.42$. Thus, up to a term of vanishing order $o(1)$, the expected value of $FP(x)$ for such a tree (where expectation refers to the tree shape produced by the birth–death process) is: $\ell_x^1 + \beta \cdot \ell_x^2 + \cdots$. Thus, for a randomly selected species x, the size of ℓ_x^2 has a non-vanishing influence on the FP score of x.

Proof of Proposition 3

Given a fully-resolved (i.e. binary) tree T with edge lengths, let ℓ_+ denote the length of the longest pendant edge and let ℓ_- denote the length of the shortest interior edge (we use ℓ_+ rather than L_t to distinguish between actual edge lengths on a given tree versus (random) edge lengths for a tree generated by a model).

The proof of Proposition 3 combines two results:

(i) K is bounded below by a term that grows at the rate $e^{c\nu\ell_+}$ as $\ell_+ \to \infty$, and

(ii) K is bounded below by a term of order $\frac{1}{\nu \cdot \ell_-}$ as $\ell_- \to 0$.

(for details, see Section 8.2.1 of Steel (2016); note that in the constant c is strictly positive and depends (only) on the particular model of site substitution). The intuition behind Result (i) is that for a very long edge e, incident with leaf x, sites evolved along e are likely to have undergone multiple substitutions and this erases the signal in the data concerning where leaf x attaches to the rest of the tree. The intuition behind Result (ii) is that if none of the sites in the data have evolved with a substitution on internal edge e then e cannot be detected from the data.

By Proposition 1(iii), $\ell_+ \sim t/2$ as λt grows. Consequently, Result (i) from the previous paragraph implies that K is bounded below by $e^{c\nu t/2}$ as $\lambda t \to \infty$. Since N_t has a geometric distribution with mean $e^{\lambda t}$, as $\lambda t \to \infty$:

$$\mathbb{P}(N_t < \lambda t e^{\lambda t}) \to 1 \tag{46}$$

So, with probability 1 as $\lambda t \to \infty$,

$$e^{c\nu t/2} = (e^{\lambda t})^{c\nu/2\lambda} \geqslant \left(\frac{e^{\lambda t}}{\lambda t}\right)^{c\nu/2\lambda} \geqslant N_t^{c\nu/2\lambda}.$$

To explore the impact of Result (ii) we take $x = t$ in Proposition 4 to obtain $\mathbb{P}(S_t^o < te^{-\lambda t}) \to 1$ as $\lambda t \to \infty$, and so Result (ii) implies that K is bounded below by a

term of order $e^{\lambda t}/(\nu t)$ as λt grows. Moreover, by Eqn. (46), and any $\epsilon > 0$, with probability tending to 1 as $\lambda t \to \infty$:

$$e^{\lambda t}/(\nu t) \geqslant \frac{N_t}{\lambda t} \cdot \frac{1}{t} \cdot \frac{1}{\nu} \geqslant N_t^{1-\epsilon}/\nu.$$

Comparing these two lower bounds on K (namely, $N_t^{c\nu/2\lambda}$ and $N_t^{1-\epsilon}/\nu$) reveals that if $\nu \geqslant 1$ then we can take $\gamma = c/2\lambda$, while if $\nu < 1$ we can take $\gamma = 1 - \epsilon$ (for any $\epsilon \in (0, 1)$), justifying the second claim of a lower bound on K that is a positive power of N_t and independent of ν.

SUPPLEMENTARY MATERIAL

Further mathematical details for the paper 'Predicting long pendant edges in model phylogenies, with applications to biodiversity and tree inference'

Sergey Bocharov, Simon C. Harris, Emma Kominek, Arne Ø. Mooers and Mike Steel

April 1, 2022

Abstract

We provide some supplementary material for the article [1] that concerns (a) shortest interior edges in Yule trees (proving Proposition 4 in [1]), and (b) long interior edges in time inhomogeneous Yule trees (which includes reconstructed Birth-Death trees, proving Lemma 4 in [1]). Throughout, we will use models and notation given in [1].

1 Short interior edges

1.1 Shortest interior edge of a Yule process

Let S_t° be the *shortest interior edge* up to time t in the Yule process with birth rate λ, that is, the shortest of all edges excluding any pendant edges that are alive at time t. Then, $\lambda e^{\lambda t} S_t^\circ$ converges in distribution to W where

$$\mathbb{P}(\lambda e^{\lambda t} S_t^\circ \geq y | N_t > 0) \to \mathbb{P}(W \geq y) = \frac{1}{1+2y} \tag{1.1}$$

as $\lambda t \to \infty$.

Proof of (1.1). A rigorous proof for the shortest interior edge in the Yule tree can be obtained by calculating the first and second moments of the number of short interior edges at time t of length less than x, $N^\circ_{t,[0,x]}$. This will requires some quite involved calculations.

Recall that, in the Yule process, we have binary branching at rate λ (and no deaths). We fix an edge length $x > 0$. For times $t > 0$, we define the number of edges of length at most x given by

$$N_t^x := \#\{\text{interior edges of length} \leq x\}$$

We will make good use of some asymptotics for the first and second moments ($m_t := \mathbb{E}N_t^x$ and $v_t := \mathbb{E}\left((N_t^x)^2\right)$) as in the following result:

Proposition 1.1.

$$m_t := \mathbb{E}(N_t^x) = \begin{cases} (1-\mathrm{e}^{-2\lambda x})\mathrm{e}^{\lambda t} - (1-\mathrm{e}^{-\lambda x}) & \textit{for } x < t, \\ e^{\lambda t} - 1 & \textit{for } x \geq t. \end{cases} \tag{1.2}$$

In particular, whenever $\lambda t \to \infty$ and $\lambda x \to 0$, we have $m_t \sim 2\lambda x \mathrm{e}^{\lambda t}$. Further, we find that the first and second moments are asymptotically equivalent, that is,

$$v_t = \mathbb{E}\Big[\big(N_t^x\big)^2\Big] \sim \mathbb{E}(N_t^x) = m_t, \tag{1.3}$$

whenever x depends on t such that $x = o(\mathrm{e}^{-\lambda t})$ as $\lambda t \to \infty$.

We prove this result shortly, but first we show how it can be used to establish the main result of this section, namely Eqn. (1.1).

Let us now fix any $y > 0$ and $\delta \in (0,1)$ and throughout we fix $x := y\mathrm{e}^{-\lambda t}$ (in this way, $x = o(\mathrm{e}^{-\lambda(1-\delta)t})$). We let S_t° be the length of the shortest interior edge at time t so that

$$\mathbb{P}(S_t^\circ > x) = \mathbb{P}(N_t^x = 0).$$

We consider the events

$$\begin{aligned} A_t =& \Big\{N_{\delta t}^x = 0\Big\}, \\ B_t =& \bigcap_{u \in N_{\delta t}} \Big\{T(u) > x\Big\}, \\ C_t =& \bigcap_{u \in N_{\delta t}} \Big\{N_{t-\delta t}^x(u) = 0\Big\}, \end{aligned}$$

where for every $u \in N_{\delta t}$, $T(u)$ is the first branching time in the subtree initiated by particle u at time δt, and $N_{t-\delta t}^x(u)$ is the size of the population after time $t - \delta t$ in the subtree initiated by particle u at time δt. We observe that

$$\mathbb{P}\Big(A_t \cap C_t\Big) \leq \mathbb{P}\Big(N_t^x = 0\Big) \quad \text{and} \quad \mathbb{P}\Big(\big\{N_t^x = 0\big\} \cap B_t\Big) \leq \mathbb{P}(C_t) \tag{1.4}$$

(Note, any edges that span time δt will be considered as two shorter edges in terms of events A_t and C_t, and event B_t implies none of the edges alive at time δt can be shorter than x.) The idea is to show that, as $\lambda t \to \infty$, $\mathbb{P}(A_t) \to 1$, $\mathbb{P}(B_t) \to 1$, and $\mathbb{P}(C_t) \to \sigma > 0$ where we can readily calculate this limit, and hence, by the bounds in (1.4), this necessarily implies $\mathbb{P}(N_t^x = 0) \to \sigma$.

To that end, since

$$\mathbb{P}(N_{\delta t}^x > 0) \leq \mathbb{E}N_{\delta t}^x = m_{\delta t} \to 0,$$

it follows that

$$\mathbb{P}(A_t) \to 1 \text{ as } \lambda t \to \infty. \tag{1.5}$$

Also, since

$$\mathbb{E}\theta^{N_t} = \frac{\mathrm{e}^{-\lambda t}\theta}{1 - (1-\mathrm{e}^{-\lambda t})\theta} \tag{1.6}$$

and as subtrees initiated by particles $u \in N_{\delta t}$ are independent copies of the original tree, it follows that

$$\mathbb{P}(B_t) = \mathbb{E}\Big[\big(\mathbb{P}(T > x)\big)^{N_{\delta t}}\Big] = \frac{\mathrm{e}^{-\lambda\delta t}\mathrm{e}^{-\lambda x}}{1 - (1-\mathrm{e}^{-\lambda\delta t})\mathrm{e}^{-\lambda x}} = \frac{\mathrm{e}^{-\lambda\delta t}\mathrm{e}^{-\lambda x}}{\lambda x + o(x) + \mathrm{e}^{-\lambda\delta t}\mathrm{e}^{-\lambda x}} \to 1 \tag{1.7}$$

as $\lambda t \to \infty$ (recall, $x = y\mathrm{e}^{-\lambda t} \ll \mathrm{e}^{-\lambda\delta t}$).

Finally, to estimate $\mathbb{P}(C_t)$, recalling the moment asymptotics of Proposition 1.1, we note that

$$\mathbb{P}\Big(N^x_{t-\delta t} > 0\Big) \le \mathbb{E}N^x_{t-\delta t} = m_{(1-\delta)t} \sim 2\lambda x \mathrm{e}^{\lambda(1-\delta)t} = 2\lambda y \mathrm{e}^{-\lambda\delta t}$$

and, using the Paley-Zygmund inequality,

$$\mathbb{P}\Big(N^x_{t-\delta t} > 0\Big) \ge \frac{\Big[\mathbb{E}N^x_{t-\delta t}\Big]^2}{\mathbb{E}\Big[\big(N^x_{t-\delta t}\big)^2\Big]} = \frac{m^2_{(1-\delta)t}}{v_{(1-\delta)t}} \sim \frac{(2\lambda x \mathrm{e}^{\lambda(1-\delta)t})^2}{2\lambda x \mathrm{e}^{\lambda(1-\delta)t}} = 2\lambda y \mathrm{e}^{-\lambda\delta t}.$$

Thus, these asymptotic upper and lower bounds agree, so we have

$$\mathbb{P}\Big(N^x_{t-\delta t} > 0\Big) \sim 2\lambda y \mathrm{e}^{-\lambda\delta t} \quad \text{as } \lambda t \to \infty. \tag{1.8}$$

Then, using (1.6) again, we find that

$$\begin{aligned}
\mathbb{P}(C_t) = \mathbb{E}\Big[\big(\mathbb{P}(N^x_{t-\delta t} = 0)\big)^{N_{\delta t}}\Big] &= \frac{\mathrm{e}^{-\lambda\delta t}\mathbb{P}(N^x_{t-\delta t} = 0)}{1 - (1 - \mathrm{e}^{-\lambda\delta t})\mathbb{P}(N^x_{t-\delta t} = 0)} \\
&= \frac{\mathrm{e}^{-\lambda\delta t}\mathbb{P}(N^x_{t-\delta t} = 0)}{\mathbb{P}(N^x_{t-\delta t} > 0) + \mathrm{e}^{-\lambda\delta t}\mathbb{P}(N^x_{t-\delta t} = 0)} \\
&= \frac{\mathrm{e}^{-\lambda\delta t}(1 + o(1))}{2\lambda y \mathrm{e}^{-\lambda\delta t} + o(\mathrm{e}^{-\lambda\delta t}) + \mathrm{e}^{-\lambda\delta t}(1 + o(1))} \qquad (1.9) \\
&\to \frac{1}{1 + 2\lambda y} \qquad \text{as } \lambda t \to \infty \qquad (1.10)
\end{aligned}$$

In conclusion, combining (1.5), (1.7) and (1.9) with inequalities (1.4), we find

$$\mathbb{P}(\mathrm{e}^{\lambda t} S^\circ_t > y) = \mathbb{P}(S^\circ_t > x) = \mathbb{P}(N^x_t = 0) \to \frac{1}{1 + 2\lambda y} \quad \text{as } \lambda t \to \infty,$$

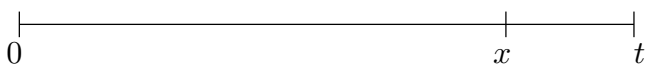
as required. □

It only remains to confirm the moment results for the number of interior edges less than length x at time t.

Proof of Proposition 1.1. (a) First moment. If $t \in (0, x]$, then every interior edge is $\le x$. That is, $N^x_t = N_t - 1$, where N_t is the population size at time t. Thus

$$m_t = \mathrm{e}^{\lambda t} - 1. \tag{1.11}$$

If $t > x$, then by considering when the first birth $T \sim \mathrm{Exp}(\lambda)$ occurs, we have



$$\begin{aligned}
m_t =& 0 \cdot \mathbb{P}(T > t) + \int_x^t 2m_{t-s}\lambda \mathrm{e}^{-\lambda s}\mathrm{d}s + \int_0^x (1 + 2m_{t-s})\lambda \mathrm{e}^{-\lambda s}\mathrm{d}s \\
=& 1 - \mathrm{e}^{-\lambda x} + 2\int_0^t m_{t-s}\lambda \mathrm{e}^{-\lambda s}\mathrm{d}s \\
=& 1 - \mathrm{e}^{-\lambda x} + 2\mathrm{e}^{-\lambda t}\int_0^t m_u \lambda \mathrm{e}^{\lambda u}\mathrm{d}u. \qquad (1.12)
\end{aligned}$$

From (1.12) and (1.11) it follows that

$$m_{x^+} = 1 - \mathrm{e}^{-\lambda x} + 2\mathrm{e}^{-\lambda x} \int_0^x (\mathrm{e}^{\lambda u} - 1)\lambda \mathrm{e}^{\lambda u} \mathrm{d}u = \mathrm{e}^{\lambda x} - 1.$$

Furthermore, from the integral equation (1.12) we see that m is continuous and differentiable on (x, ∞). Differentiating (1.11) yields

$$m_t' = \lambda m_t + \lambda(1 - \mathrm{e}^{-\lambda x}) \tag{1.13}$$

Multiplying (1.13) by the integrating factor $\mathrm{e}^{-\lambda t}$ gives

$$\frac{\mathrm{d}}{\mathrm{d}t}\Big(m_t \mathrm{e}^{-\lambda t}\Big) = \lambda(1 - \mathrm{e}^{-\lambda x})\mathrm{e}^{-\lambda t}.$$

Integrating the above equation between x and t gives

$$m_t \mathrm{e}^{-\lambda t} - m_{x^+}\mathrm{e}^{-\lambda x} = (1 - \mathrm{e}^{-\lambda x})(\mathrm{e}^{-\lambda x} - \mathrm{e}^{-\lambda t})$$

and hence

$$m_t = (1 - \mathrm{e}^{-2\lambda x})\mathrm{e}^{\lambda t} - (1 - \mathrm{e}^{-\lambda x}) \sim 2\lambda x \mathrm{e}^{\lambda t} \tag{1.14}$$

whenever $\lambda x \to 0$ and $\lambda t \to \infty$ since $e^y = 1 + y + o(y)$ for $y \to 0$.
(b) Second moment. If $t \in (0, x]$ then, as before, every interior edge is at most length x. That is, $N_t^x = N_t - 1$ and hence

$$v_t = \mathbb{E}\big[N_t^2\big] - 2\mathbb{E}N_t + 1 = 2\mathrm{e}^{2\lambda t} - 3\mathrm{e}^{\lambda t} + 1 \tag{1.15}$$

If $t > x$, then once again by considering when the first birth $T \sim \mathrm{Exp}(\lambda)$ occurs, we find

$$\begin{aligned} v_t &= 0 \cdot \mathbb{P}(T > t) + \mathbb{E}\Big[(\hat{N}_{t-T}^x + \tilde{N}_{t-T}^x)^2 \mathbf{1}_{\{x < T \le t\}}\Big] + \mathbb{E}\Big[(1 + \hat{N}_{t-T}^x + \tilde{N}_{t-T}^x)^2 \mathbf{1}_{\{0 \le T \le x\}}\Big] \\ &= \int_x^t (2v_{t-s} + 2m_{t-s}^2)\lambda \mathrm{e}^{-\lambda s}\mathrm{d}s + \int_0^x (1 + 2v_{t-s} + 4m_{t-s} + 2m_{t-s}^2)\lambda \mathrm{e}^{-\lambda s}\mathrm{d}s \\ &= (1 - \mathrm{e}^{-\lambda x}) + 2\mathrm{e}^{-\lambda t}\int_0^t v_u \lambda \mathrm{e}^{\lambda u}\mathrm{d}u + 2\mathrm{e}^{-\lambda t}\int_0^t m_u^2 \lambda \mathrm{e}^{\lambda u}\mathrm{d}u + 4\mathrm{e}^{-\lambda t}\int_{t-x}^t m_u \lambda \mathrm{e}^{\lambda u}\mathrm{d}u \end{aligned} \tag{1.16}$$

In particular, v_t is continuous on (x, ∞) and hence also differentiable on (x, ∞). We may also find from the integral equation (1.16) and also (1.11) and (1.15) that

$$\begin{aligned} v_{x^+} &= (1 - \mathrm{e}^{-\lambda x}) + 2\mathrm{e}^{-\lambda x}\int_0^x v_u \lambda \mathrm{e}^{\lambda u}\mathrm{d}u + 2\mathrm{e}^{-\lambda x}\int_0^x m_u^2 \lambda \mathrm{e}^{\lambda u}\mathrm{d}u + 4\mathrm{e}^{-\lambda x}\int_0^x m_u \lambda \mathrm{e}^{\lambda u}\mathrm{d}u \\ &= 2\mathrm{e}^{2\lambda x} - 3\mathrm{e}^{\lambda x} + 1 \end{aligned}$$

In particular, $v_{x^+} \sim \lambda x$ as $x \to 0$. In order to help find its solution, the integral equation (1.16) can be converted to the following differential equation

$$v_t' - \lambda v_t = 2\lambda m_t^2 + 4\lambda m_t - 4\lambda \mathrm{e}^{-\lambda x} m_{t-x} + \lambda(1 - \mathrm{e}^{-\lambda x}). \tag{1.17}$$

Equivalently,

$$\frac{\mathrm{d}}{\mathrm{d}t}\big(\mathrm{e}^{-\lambda t} v_t\big) = \Big(2\lambda m_t^2 + 4\lambda m_t - 4\lambda \mathrm{e}^{-\lambda x} m_{t-x} + \lambda(1 - \mathrm{e}^{-\lambda x})\Big)\mathrm{e}^{-\lambda t}$$

which needs to be solved separately in intervals $(x, 2x)$ and $(2x, \infty)$ due to the fact that m_{t-x} has a different form in each of these two intervals. We find that

$$v_t = \begin{cases} \mathrm{e}^{\lambda t}\left\{\mathrm{e}^{-\lambda x} v_{x^+} + \int_x^t \left(2\lambda m_s^2 + 4\lambda m_s - 4\lambda \mathrm{e}^{-\lambda x} m_{s-x} + \lambda(1-\mathrm{e}^{-\lambda x})\right)\mathrm{e}^{-\lambda s}\mathrm{d}s\right\} & \text{if } t \in (x, 2x], \\ \mathrm{e}^{\lambda t}\left\{\mathrm{e}^{-2\lambda x} v_{2x} + \int_{2x}^t \left(2\lambda m_s^2 + 4\lambda m_s - 4\lambda \mathrm{e}^{-\lambda x} m_{s-x} + \lambda(1-\mathrm{e}^{-\lambda x})\right)\mathrm{e}^{-\lambda s}\mathrm{d}s\right\} & \text{if } t > 2x. \end{cases} \tag{1.18}$$

Rather than evaluating this exactly, which is unnecessary for our purposes, we find from the first line of (1.18) that

$$\begin{aligned} v_{2x} =& \mathrm{e}^{2\lambda x}\Big[\mathrm{e}^{-\lambda x} v_{x^+} + \int_x^{2x} \Big(2\lambda m_s^2 + 4\lambda m_s - 4\lambda \mathrm{e}^{-\lambda x} m_{s-x} + \lambda(1-\mathrm{e}^{-\lambda x})\Big)\mathrm{e}^{-\lambda s}\mathrm{d}s\Big] \\ =& \big(1+o(1)\big) v_{x^+} + o(x) \sim \lambda x \text{ as } \lambda x \to 0 \end{aligned} \tag{1.19}$$

(In the integral, one should use (1.14) for m_s and (1.11) for m_{s-x}.)

The second line of (1.18) says that

$$v_t = \mathrm{e}^{\lambda t}\mathrm{e}^{-2\lambda x} v_{2x} + \mathrm{e}^{\lambda t}\int_{2x}^t \Big(2\lambda m_s^2 + 4\lambda m_s - 4\lambda \mathrm{e}^{-\lambda x} m_{s-x}\Big)\mathrm{e}^{-\lambda s}\mathrm{d}s + \mathrm{e}^{\lambda t}\big(1-\mathrm{e}^{-\lambda x}\big)\int_{2x}^t \lambda \mathrm{e}^{-\lambda s}\mathrm{d}s \tag{1.20}$$

If we now choose x such that $x = o(\mathrm{e}^{-\lambda t})$ as $\lambda t \to \infty$ and use (1.19) for v_{2x}, (1.14) for terms m_s and m_{s-x} inside the integral then (1.20) becomes

$$v_t = \lambda x \mathrm{e}^{\lambda t}\big(1+o(1)\big) + o(x\mathrm{e}^{\lambda t}) + \lambda x \mathrm{e}^{\lambda t}\big(1+o(1)\big) \sim 2\lambda x \mathrm{e}^{\lambda t} \tag{1.21}$$

as required. □

2 Long interior edges in time inhomogeneous Yule processes

Consider a *time inhomogeneous Yule process* where each individual alive at time s independently splits into 2 individuals at (time dependent) instantaneous rate λ_s. That is, for example, for small times $h > 0$,

$$\mathbb{P}(N_{t+h} = k+1 | N_t = k) = k\lambda_t h + o(h), \quad \mathbb{P}(N_{t+h} = k | N_t = k) = 1 - k\lambda_t h + o(h),$$

and the probability there has been no birth event by time t is given by

$$\mathbb{P}(N(t) = 1 | N(0) = 1) = \exp\left(-\int_0^t \lambda_s \, ds\right).$$

We will assume throughout that λ_s is a continuous function in time parameter s.

Example 2.1. *The following are special cases of time inhomogeneous Yule processes:*

(i) The classical Yule process with constant branching rate has $\lambda_s \equiv \lambda$.

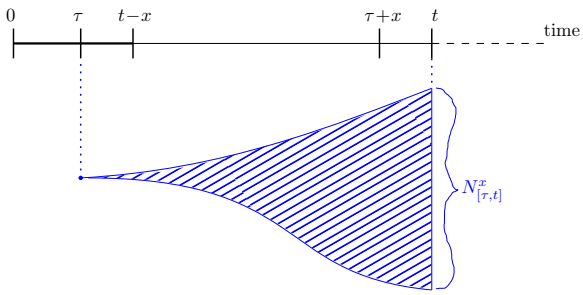
(ii) The reconstructed (or reduced) tree derived from the birth-death process with birth rate λ and death rate μ has instantaneous branching rate

$$\lambda_s = \lambda(1 - p_{t-s}) := \lambda \cdot \frac{\lambda - \mu}{\lambda - \mu \mathrm{e}^{-(\lambda-\mu)(t-s)}} = \lambda - \mu + \frac{\mu(\lambda-\mu)\mathrm{e}^{-(\lambda-\mu)(t-s)}}{\lambda - \mu \mathrm{e}^{-(\lambda-\mu)(t-s)}}.$$

We can interpret this rate intuitively as follows. Recall, any individual that remains in the reduced tree must have at least one descendant still alive at time t (ie. any family lines that go extinct before time t are removed). Now, any individual alive at time s that is part of the reduced tree at time t will give birth to another individual at rate λ but this individual will only be in the reduced tree if its descendants survive to time t from its birth at time s, where this has probability $1 - p_{t-s}$ (as defined above). Thus, the effective rate of births in the reduced tree becomes $\lambda(1 - p_{t-s})$.

We wish to find the expected number of interior edges in an inhomogeneous Yule process at time t that are of length at least x when starting from a single ancestor at time 0.

- Fix a time $t > 0$ and an edge length $x \in \left(\frac{1}{2}t, t\right)$.
- For any time $\tau \in [0, t - x]$ we let $N^x_{[\tau,t]}$ be the number of interior edges of length at least x at time t in Yule process with instantaneous branching rate λ_s initiated at time τ. (Here, interior edges at time t means all edges up to time t except for pendant edges alive at t.)
- Define $m_\tau := \mathbb{E}(N^x_{[\tau,t]} \,|\, N_\tau = 1)$
- Note, whenever $t > 0$, $x \in (t/2/t)$, and $\tau \in [0, t - x]$, we have $\tau \leq t - x \leq \tau + x \leq t$.



Lemma 2.2. *For $t > 0$ and $x \in (t/2, t)$,*

$$m_0 = \mathbb{E}(N^x_{[0,t]} \,|\, N_0 = 1) = \int_0^{t-x} \lambda_{\tau+x} \exp\left(\int_0^\tau \lambda_u \mathrm{d}u - \int_\tau^{\tau+x} \lambda_u \mathrm{d}u\right) \mathrm{d}\tau.$$

Proof of Lemma 2.2. Let T be the first branching time in the process initiated from a single particle at time τ. Then, for $s > \tau \geq 0$,

$$\mathbb{P}(T > s \,|\, N_\tau = 1) = \mathrm{e}^{-\int_\tau^s \lambda_u \mathrm{d}u} \tag{2.1}$$

and we find that m_τ satisfies the following integral equation

$$
\begin{aligned}
m_\tau &= \mathbb{E}(N^x_{[\tau,t]} \,|\, N_\tau = 1) \\
&= \mathbb{E}\Big[N^x_{[\tau,t]}\mathbf{1}_{\{\tau\le T<t-x\}} \,|\, N_\tau = 1\Big] + \mathbb{E}\Big[N^x_{[\tau,t]}\mathbf{1}_{\{t-x\le T<\tau+x\}} \,|\, N_\tau = 1\Big] + \\
&\qquad\qquad \mathbb{E}\Big[N^x_{[\tau,t]}\mathbf{1}_{\{\tau+x\le T\le t\}} \,|\, N_\tau = 1\Big] + \mathbb{E}\Big[N^x_{[\tau,t]}\mathbf{1}_{\{T>t\}} \,|\, N_\tau = 1\Big] \\
&= \int_\tau^{t-x} 2m_s \cdot \lambda_s \mathrm{e}^{-\int_\tau^s \lambda_u \mathrm{d}u}\mathrm{d}s + 0 \cdot \mathbb{P}(t-x \le T \le \tau + x \,|\, N_\tau = 1) + \\
&\qquad\qquad 1 \cdot \mathbb{P}(\tau + x \le T \le t \,|\, N_\tau = 1) + 0 \cdot \mathbb{P}(T > t \,|\, N_\tau = 1) \\
&= 2\int_\tau^{t-x} m_s \lambda_s \mathrm{e}^{-\int_\tau^s \lambda_u \mathrm{d}u}\mathrm{d}s + \Big[\mathrm{e}^{-\int_\tau^{\tau+x}\lambda_u \mathrm{d}u} - \mathrm{e}^{-\int_\tau^t \lambda_u \mathrm{d}u}\Big].
\end{aligned}
$$

In the above decomposition, the first term covers the first birth event occurring at any times $s \le t - x$, so the first interior edge is not longer that x (ie. too short to be counted) and we add up edges from the 2 independent offspring starting at time s. The second term covers branching events at times $s \in [t-x, \tau+x]$, so this first interior edge is shorter than x (ie. no contribution as occurs too early before time $\tau+x$) and neither offspring can result in any interior edges longer than x (since they start too late after time $t-x$). In the third term, the first birth event occurs at time $s \in [\tau+x, t]$ so this first interior edge is longer than x but the offspring cannot have any interior edges longer than x, hence exactly 1 interior edge results. Finally, the last term covers when there is no birth event before time t, hence no interior edges at all.

Differentiating this integral equation with respect to τ and also taking $\tau = t - x$ for the initial condition (where the first birth would have to occur *exactly* at time t to give 1 interior edge of length at least x, and this has zero probability), we derive the following first-order linear ordinary differential equation.

$$
\begin{cases} m'_\tau &= -\lambda_\tau m_\tau - \nu_\tau \ , \qquad \tau \in (0, t-x) \\ m_{t-x} &= 0 \end{cases}
$$

where

$$
\nu_\tau := \lambda_{\tau+x}\mathrm{e}^{-\int_\tau^{\tau+x}\lambda_u \mathrm{d}u}.
$$

To solve this differential equation we multiply it by the integrating factor $\exp\Big(-\int_\tau^{t-x}\lambda_u \mathrm{d}u\Big)$, so

$$
\frac{\mathrm{d}}{\mathrm{d}\tau}\Big(m_\tau \mathrm{e}^{-\int_\tau^{t-x}\lambda_u \mathrm{d}u}\Big) = m'_\tau \mathrm{e}^{-\int_\tau^{t-x}\lambda_u \mathrm{d}u} + \lambda_\tau m_\tau \mathrm{e}^{-\int_\tau^{t-x}\lambda_u \mathrm{d}u} = -\nu_\tau \mathrm{e}^{-\int_\tau^{t-x}\lambda_u \mathrm{d}u}\,.
$$

Then integrating over τ from 0 to $t-x$ reveals that

$$
m_{t-x} - m_0 \mathrm{e}^{-\int_0^{t-x}\lambda_u \mathrm{d}u} = \int_0^{t-x} \frac{\mathrm{d}}{\mathrm{d}\tau}\Big(m_\tau \mathrm{e}^{-\int_\tau^{t-x}\lambda_u \mathrm{d}u}\Big)\mathrm{d}\tau = -\int_0^{t-x} \nu_\tau \mathrm{e}^{-\int_\tau^{t-x}\lambda_u \mathrm{d}u}\mathrm{d}\tau
$$

Finally, using boundary condition $m_{t-x} = 0$, we obtain

$$
m_0 = \int_0^{t-x} \nu_\tau \mathrm{e}^{\int_0^\tau \lambda_u \mathrm{d}u}\mathrm{d}\tau = \int_0^{t-x} \lambda_{\tau+x}\mathrm{e}^{\int_0^\tau \lambda_u \mathrm{d}u - \int_\tau^{\tau+x}\lambda_u \mathrm{d}u}\mathrm{d}\tau.
$$

□

Example 2.3 (Classical Yule process)**.** *In the special case of constant branching where $\lambda_s \equiv \lambda$,*

$$
m_0 = \int_0^{t-x} \lambda \mathrm{e}^{\lambda(\tau - x)}\mathrm{d}\tau = \mathrm{e}^{\lambda(t-2x)} - \mathrm{e}^{-\lambda x}.
$$

Corollary 2.4 (Atypically long interior edges in reconstructed birth-death tree)**.** *For some constant C and any sufficiently small $\epsilon > 0$,*

$$\mathbb{P}\left(\exists \text{ interior edge of length } \geq (\frac{1}{2}+\epsilon)t\right) \leq Ce^{-2\epsilon(1-\rho)\lambda t}$$

That is, the probability of finding an interior edge longer than $(1/2+\epsilon)t$ in the reconstructed birth-death tree decays exponentially fast to zero as $\lambda t \to \infty$ (where $\rho := \mu/\lambda$ is fixed)

Proof of Corollary 2.4. It is known that the reduced (reconstructed) tree of the birth-death process with birth rate λ and death rate μ is a time inhomogeneous Yule tree with instantaneous branching rate

$$\lambda_s = \lambda(1-p_{t-s}) = \lambda \cdot \frac{\lambda-\mu}{\lambda-\mu \mathrm{e}^{-(\lambda-\mu)(t-s)}} = \lambda - \mu + \frac{\mu(\lambda-\mu)\mathrm{e}^{-(\lambda-\mu)(t-s)}}{\lambda-\mu\mathrm{e}^{-(\lambda-\mu)(t-s)}}$$

We note that for all s,

$$\lambda - \mu \leq \lambda_s \leq \lambda - \mu + \mu \mathrm{e}^{-(\lambda-\mu)(t-s)}$$

and trivially $\lambda_s \leq \lambda$. Combining these simple bounds with Lemma 2.2, we find

$$\begin{aligned}
\mathbb{E}(N^x_{[0,t]} \mid N_0 = 1) &= \int_0^{t-x} \lambda_{\tau+x} \mathrm{e}^{\int_0^\tau \lambda_u \mathrm{d}u - \int_\tau^{\tau+x} \lambda_u \mathrm{d}u} \mathrm{d}\tau \\
&\leq \int_0^{t-x} \lambda \mathrm{e}^{\int_0^\tau (\lambda-\mu+\mu\mathrm{e}^{-(\lambda-\mu)(t-u)})\mathrm{d}u - \int_\tau^{\tau+x}(\lambda-\mu)\mathrm{d}u} \mathrm{d}\tau \\
&\leq C\int_0^{t-x} (\lambda-\mu)\mathrm{e}^{(\lambda-\mu)(\tau-x)}\mathrm{d}\tau = C\Big(\mathrm{e}^{(\lambda-\mu)(t-2x)} - \mathrm{e}^{-(\lambda-\mu)x}\Big)
\end{aligned}$$

for some constant C.

Thus, for small $\epsilon > 0$, we have $\mathbb{E}(N^{(1/2+\epsilon)t}_{[0,t]} \mid N_0 = 1) \leq Ce^{-(\lambda-\mu)2\epsilon t} = e^{-2\epsilon(1-\rho)\lambda t}$. Further, by applying Markov's inequality,

$$\begin{aligned}
\mathbb{P}\left(\exists \text{ interior edge of length } \geq (\frac{1}{2}+\epsilon)t\right) &= \mathbb{P}\big(N^{(\frac{1}{2}+\epsilon)t}_{[0,t]} > 0 \,\big|\, N_0 = 1\big) \\
&\leq \mathbb{E}\big(N^{(\frac{1}{2}+\epsilon)t}_{[0,t]} \,\big|\, N_0 = 1\big) \leq Ce^{-2\epsilon(1-\rho)\lambda t}
\end{aligned}$$

which is decaying exponentially fast to zero as $\lambda t \to \infty$ (where $\rho = \mu/\lambda$ is fixed), as required. □

References

[1] Sergey Bocharov, Simon Harris, Emma Kominek, Arne Ø. Mooers, and Mike Steel. Predicting long pendant edges in model phylogenies, with applications to biodiversity and tree inference. (2022)