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 phylodynamic 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 questions concerning 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: Phylodynamics, speciation–extinction 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 '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 deeper-than-expected splitting times (Morlon et al., 2010).

Mathematical investigations into phylodynamic 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' (which is relevant to data sets that have complete sampling through time, as with certain viral data sets, or fossil-rich phylogenies) and for the 'reduced tree' (relevant to reconstructed phylogenies built from genomic sequence data at the present). 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 latest 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.

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)

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

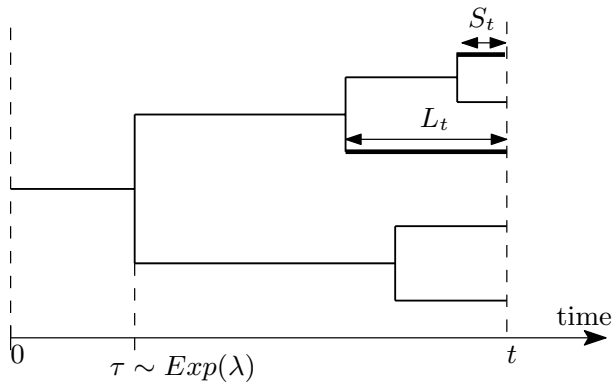

Fig. 1. A Yule (pure-birth) tree starts with a single lineage at time $t = 0$, with each lineage persisting for an exponentially distributed time τ with mean $1/\lambda$. The four splitting events result in five leaf species at time t. The edge label L_t refers to the length of the longest pendant edge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

(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 a 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 phylodynamics 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, L_t/t converges to $(1-\rho)/2$ in mean (i.e. $\mathbb{E}(\frac{L_t}{t}|N_t > 0) \to (1-\rho)/2$) as well as in probability, whenever $\lambda t \to \infty$.

(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, $L_t^{(k)}/t$ converges in mean and in probability to $(1-\rho)/2$ as $\lambda t \to \infty$.

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.

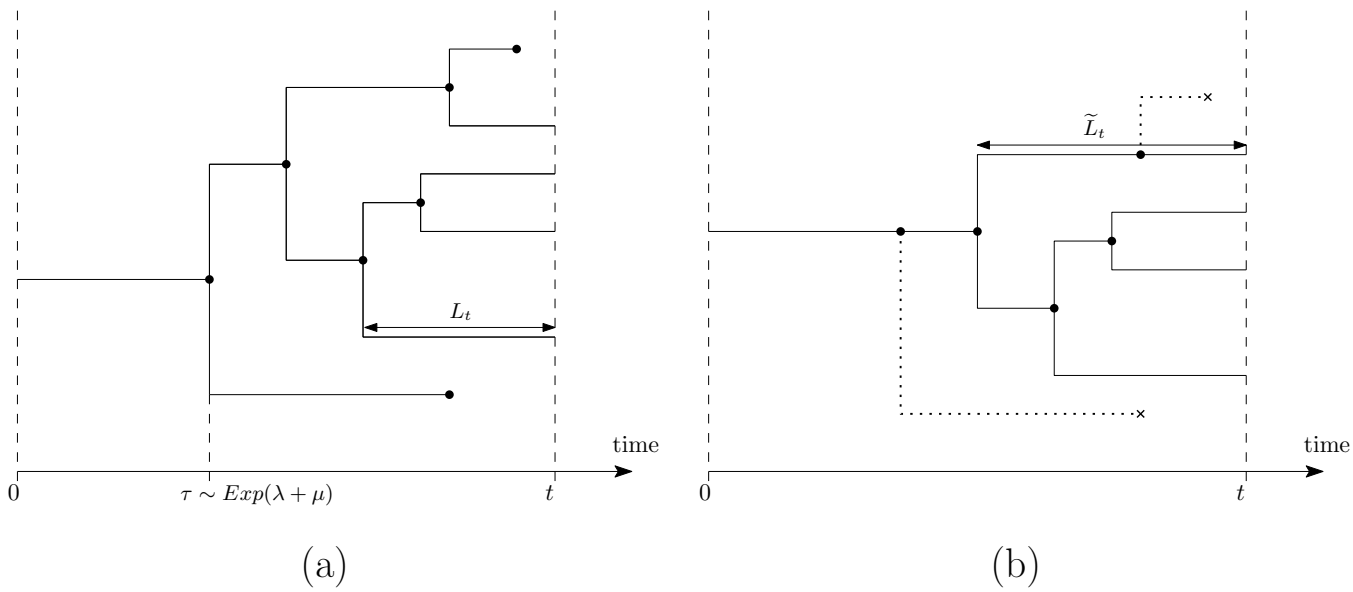

Fig. 2. (a) A complete birth–death tree and (b) the corresponding reduced tree, where the dashed lineages have been pruned off. In (a), the edge of length L_t is the longest pendant edge that is still alive at time t (the longer pendant edge below it dies before time t). Notice also that the length of 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 $1 - \lambda p_{t-s}$ 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. ((38)) and ((39)). 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. (38) and (39).

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. (38) and (39) 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, $\widetilde{L}_t/t$ converges to $\frac{1}{2}$ in mean (i.e. $\mathbb{E}(\frac{\widetilde{L}_t}{t}|N_t > 0) \to \frac{1}{2}$) as well as in probability as $\lambda t \to \infty$.

(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, $\widetilde{L}_t^{(k)}/t$ converges in mean and in probability to $\frac{1}{2}$ as $\lambda t \to \infty$.

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 two sets of 500 complete birth-death phylogenies (with $\lambda = 0.416, \mu = 0$ and with $\lambda = 0.832, \mu = \lambda/2 = 0.416$), all with a fixed age of $t = 10$ and so a target size of $n = 64$ extant species, retaining only trees with 1 or more extant species. 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, 2018) 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, $\frac{1-\mu/\lambda}{2}$ and $\frac{1}{2}$) for 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 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).

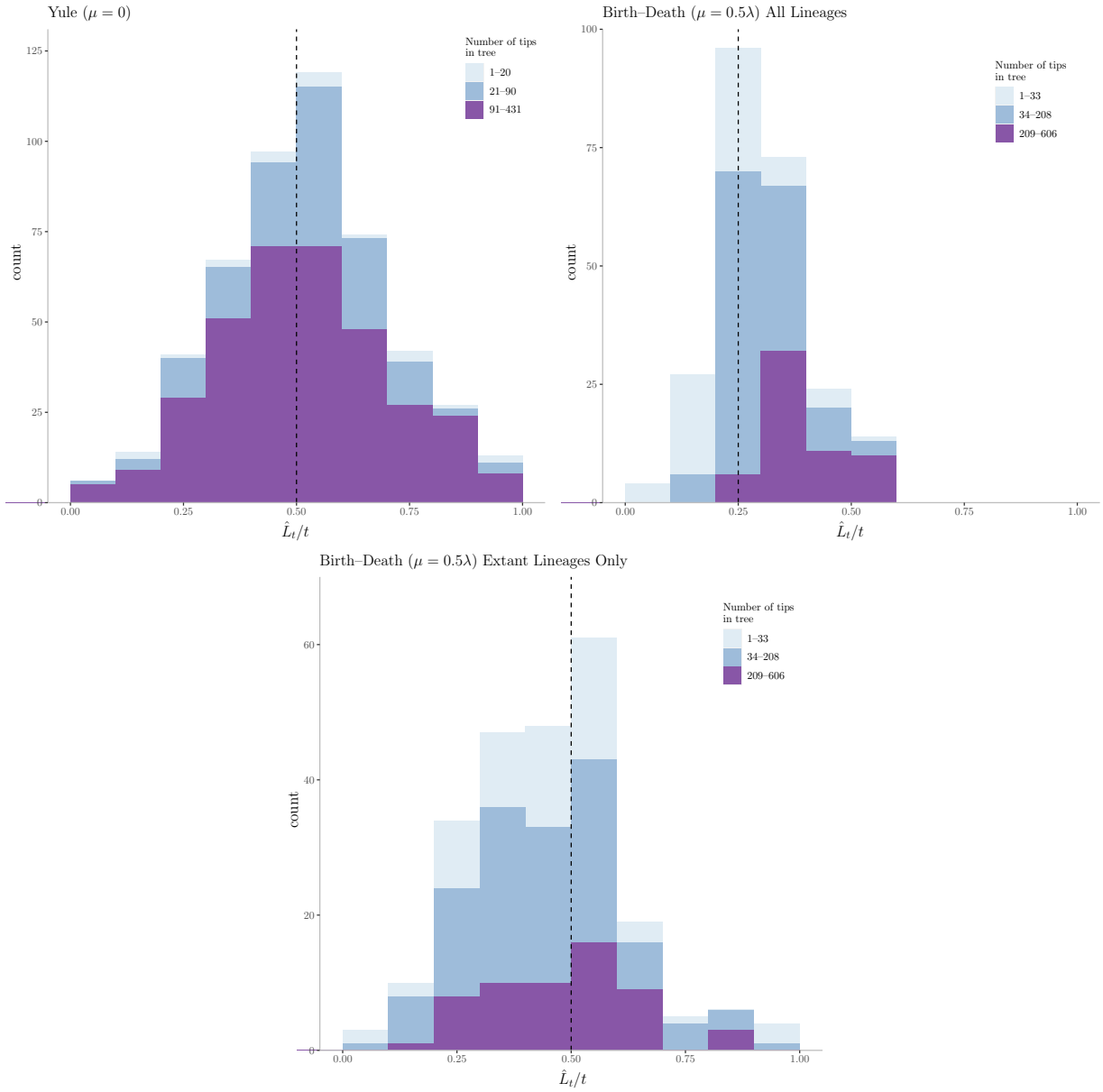

Fig. 3. The longest pendant edges (that are still alive at time t) on simulated birth-death trees, with $\hat{L}_t/t$ on the horizontal axis, and with trees sorted into three bins of small, medium and large trees. *Top Left:* The longest pendant edges (as $\hat{L}_t/t$) for 500 Yule trees ($\lambda = 0.416, \mu = 0$) 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 ($\lambda = 0.832$, $\mu = 0.416$, $t = 10$). The dotted vertical line at 0.25 ($= (1-\rho)/2$) indicates the expected value of $\hat{L}_t/t$ in the large tree limit. *Bottom:* $\hat{L}_t/t$ for reduced birth-death trees ($\lambda = 0.832, \mu = 0.416, t = 10$). The dotted vertical line at 0.5 again indicates the expected value of $\hat{L}_t/t$ in the large tree limit.

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:

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

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).

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:

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

Notice that if $n_2(x) \geqslant M$, then:

$$0 \leqslant \epsilon(x) \leqslant \ell_x^2/M + \ell_x^3/(M+1) + \cdots + \ell_x^{k_x}/(M+k_x-2) \leqslant (t-\ell_x^1)k_x/M. \tag{32}$$

Suppose that x is a tip of a longest pendant edge in T. By Theorem 2, we expect that $\ell_x^1 \sim t/2$. Now consider the subtree of T descending from the other endpoint of edge e_1 to x. If the tree 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 $\ell_x^1 \sim t/2$, we see that $n_2(x)$ is at least $e^{(\lambda-\mu)t/4}$ with high probability. 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 can also be assumed to be of order $\log(n)$ which, by Jensen's inequality, grows at most linearly in expectation with λt.

Consequently, from Inequality (32), $\epsilon(x)$ is (with high probability as λt grows) bounded above by a term of order $\lambda t^2 e^{-\lambda t/4}$. Consequently, for the longest pendant edge (incident with leaf x), we have:

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

where $o(1)$ refers to a positive term that tends to 0 in expectation as λt grows (provided that t does not converge to 0).

Note that for a randomly selected extant species x in a birth–death tree, the equation $FP(x) = \ell_x^1 + o(1)$ 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:

$$\sum_{k=2}^{n-1} \frac{4}{k^2(k^2-1)} + \frac{2}{n^2(n-1)}. \tag{33}$$

By using partial fractions, and the identity $\sum_{k=1}^{\infty} k^{-2} = \pi^2/6$, the expression in (33) 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:

$$\ell_x^1 + \beta \cdot \ell_x^2 + \cdots .$$

In particular, for this randomly selected species x, the size of ℓ_x^2 has a non-vanishing influence on the FP score of species x. These considerations lead to the following prediction:

> *In large birth–death trees with $\lambda > \mu$, a species with the highest FP score is typically a species at the end of a longest pendant edge.*

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.

Technically, evolutionary distinctiveness measures evolutionary non-redundancy. However, this result, coupled with an early observation that about $\frac{1}{2}$ of the value of FP comes from the pendant edge (Redding et al., 2008), supports the commonsense interpretation of evolutionary distinctness as a measure of feature uniqueness, namely as products of evolution found nowhere else in the Tree of Life. Species on very long pendant edges would be expected to be those with more truly unique features.

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 confirms 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 supports 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.

Although we make no claims that mammal families have diversified via a Yule or birth–death process (*cf.* (Louca and Pennell, 2020)), this suggests that our results have a direct practical application.

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

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

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 these symbols rather than L_t and S_t to distinguish between actual edge lengths on a given tree versus (random) edge lengths for a tree generated by a model). Under standard finite-state models of site substitution, with substitution rate ν per cite, consider the number of sites required to infer T correctly with a given high probability; we will call this quantity K.

Note that K depends on the tree, the edge lengths, the model of site substitution

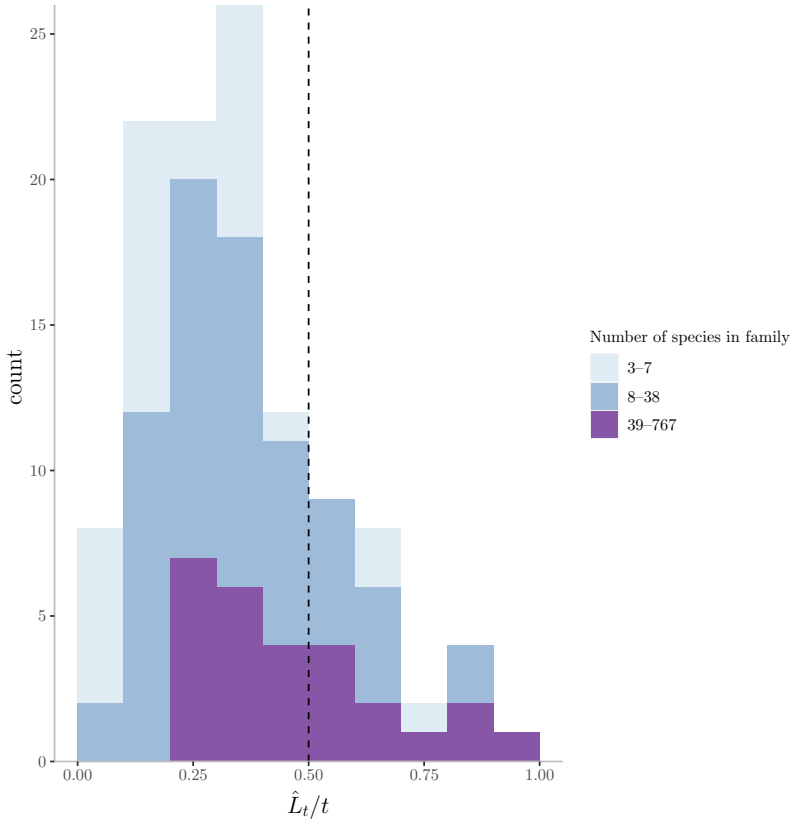

Fig. 4. A histogram of the values $\hat{L}_t/t$ across 114 mammal families, broken up into the three groups (3-7 species, 8-38 species and 39+ species) distinguished by separate colours (stacked on each other) and with further statistical details provided in Table 1.

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 (which treats the edge lengths on the inferred tree as nuisance parameters that are unconstrained and unlinked (in particular a molecular clock assumption is not imposed for the reconstruction), and so the inference of the tree does not include specifying the location of the root vertex.

Under these assumptions it can be shown that K grows as the inverse square of $\nu \cdot \ell_-$ as $\ell_- \to 0$, moreover, K grows exponentially with $\nu \cdot \ell_+$ as $\ell_+ \to \infty$ (for details, see Section 8.2.1 of Steel (2016)). Thus, short interior edges and/or long pendant edges in T require large amounts of sequence data for accurate tree reconstruction. Moreover, in some

special cases these lower bounds combine together multiplicatively to give a stronger lower bound (Martyn and Steel, 2012).

From our theoretical results, we have $\ell_+ \sim t/2$ (for a Yule tree or a reduced birth–death tree) and $\ell_+ \sim t/2 \cdot (1 - \mu/\lambda)$ (for a complete tree). Consequently, for these phylodynamic models, and a fixed site substitution rate (ν), the first result in the previous paragraph (concerning the growth of K) entails the following:

> *the number of aligned sequence sites required to accurately infer a phylogenetic tree generated by a birth–death process grows exponentially with the depth of the tree (the time from the present back to the origin of the clade).*

Notice that the number of leaves of the tree (under the Yule and birth-death processes we consider in this paper) also grows exponentially with the depth of the tree. Thus, a related question is the following: *How does K grow in terms of the expected number of leaves of the tree?* Consider, for simplicity, a Yule tree $\mathcal{T}_t$ grown for time t with speciation rate λ, and we let $\overline{N}_t$ denote the expected number of leaves of $\mathcal{T}_t$, namely, $e^{\lambda t}$. The justification of the following result is provided in the Appendix.

Proposition 3 With probability $1 - o(1)$, K is bounded below by terms of order:

$$\left(\frac{\overline{N}_t}{\ln(\overline{N}_t)(\nu/\lambda)}\right)^2 \text{ and } \overline{N}_t^{c\cdot\nu/\lambda}, \tag{34}$$

where c is a model-dependent constant.

Notice that first term in (34) is (almost) quadratic in $\overline{N}_t$, and it also increases quadratically with $\frac{1}{\nu}$. Thus, using faster evolving sequences (i.e. increasing ν) will reduce the size of this first lower bound on K (this first term in (34) reflects the difficulty of resolving short interior edges of the tree correctly). However, the second term in (34) grows exponentially with ν (this second term comes the difficulty of accurately placing a leaf that is distantly related to the rest of the tree). Consequently, the two terms of (34) in concert place a fundamental lower bound on how much data is required to reconstruct a large Yule

tree, regardless of what value ν takes.

It is interesting to note 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 variable and likely to depend on the number of leaves of a tree (n) so a logarithmic dependence of K on n would be replaced by a faster growth function, and our results above provide a more formal verification of this claim.

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; Magallon and Sanderson (2001)) have predictive power (see, e.g. Greenberg et al. (in press)); 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, Weltwitschia 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.

Finally, it would be of interest to investigate the length of the longest (and shortest) edges in phylodynamic trees 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).

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.

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. Stat. 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.

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

Fischer, M. and M. Steel. 2009. Sequence length bounds for resolving a deep phylogenetic divergence. J. Theor. Biol. 256:157–163.

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. Upham, W. Jetz, and A. O. Mooers. in press. Evolutionary legacies in contemporary tetrapod imperilment. Ecol. Lett. .

Grimmett, G. and D. Stirzaker. 2001. Probability and Random Processes. 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. Stat. 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.

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

Martyn, I. and M. Steel. 2012. The impact and interplay of long and short branches on phylogenetic information content. J. Theor. Biol. 314:157–163.

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. Evolutionary trees and the Ising model on the Bethe lattice: A proof of Steel's conjecture. 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 344:77—82.

Paradis, E. and K. Schliep. 2018. ape 5.0: an environment for modern phylogenetics and evolutionary analyses. 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 priorisation 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 distinct species 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.

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.

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)):

$$\mu_t = \int_0^t \frac{e^{\lambda(t-2x)}}{1 - e^{-\lambda x} + e^{\lambda(t-2x)}} dx, \tag{35}$$

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:

$$\frac{1}{\lambda}\int_{e^{-\lambda t/2}}^{e^{\lambda t/2}} \frac{\xi}{1 - e^{-\lambda t/2}\xi + \xi^2} d\xi. \tag{36}$$

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 (36) 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 $t \to \infty$.

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

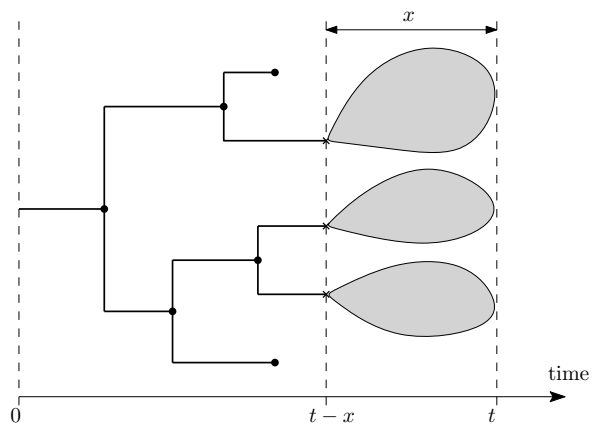

Fig. 5. Subtrees of the birth–death tree $\mathcal{T}_t$ initiated at time $t-x$ behave like N_{t-x} independent birth–death trees each distributed like $\mathcal{T}_x$. A pendant edge at time t of length greater than x must have been present at time $t-x$ and not have any other birth or death event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

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 (18), 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:

$$\mathbb{P}\left(L_t - \frac{t}{2}(1-\rho) \leqslant \frac{y}{\lambda} \,\middle|\, N_t > 0\right) = \mathbb{P}\left(N_t^{\frac{t}{2}(1-\rho)+\frac{y}{\lambda}} = 0 \,\middle|\, N_t > 0\right). \tag{37}$$

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 (37), 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:

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

and

$$\begin{aligned}\widetilde{q}_{t,x,\lambda,\mu} :&= \frac{q_{t-x}}{q_{t-x} + (1-q_{t-x})(1-p_x)q_x} \\ &= \frac{(1-\rho e^{-(\lambda-\mu)x})^2}{(1-\rho e^{-(\lambda-\mu)x})^2 - (1-\rho)e^{-(\lambda-\mu)x} + (1-\rho)e^{(\lambda-\mu)(t-2x)}}\end{aligned} \tag{39}$$

Proof of Lemma 3. We proceed by modifying the proof of Lemma 2. Firstly, recall that $N_t \sim \text{ModGeom}(p_t, q_t)$. 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 $\widetilde{N}_t^x \sim \text{Bin}(N_{t-x}, (1-p_x)q_x)$.

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

$$\widetilde{p} = p + (1-p)\left(\frac{(1-r)q}{1-(1-q)(1-r)}\right) = \frac{pr+(1-r)q}{q+(1-q)r} \tag{40}$$

$$\widetilde{q} = \frac{q}{1-(1-q)(1-r)} = \frac{q}{q+(1-q)r}, \tag{41}$$

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 $(1-\widetilde{p}_*)/(1-\rho) = 1-\widetilde{q}_*$. □

Justification of Proposition 3

If S'_t denotes the length of the shortest interior edge in a Yule tree $\mathcal{T}_t$ then it can be shown that S'_t converges in distribution, with:

$$\mathbb{P}\left(S'_t \geqslant y \cdot \frac{e^{-\lambda t}}{\lambda}\right) \to \frac{1}{1+2y}, \text{ as } \lambda t \to \infty. \tag{42}$$

If we now set $y = \lambda t$ in (42) we obtain $\mathbb{P}(S'_t \leqslant te^{-\lambda t}) \to 1$, as $\lambda t \to \infty$, and so $\frac{1}{\nu \cdot S'_t} \geqslant \frac{e^{\lambda t}}{\lambda t(\nu/\lambda)} = \frac{\overline{N}_t}{\ln(N_t)(\nu/\lambda)}$ with probability $1-o(1)$. Since K grows at least as fast as $(\nu \cdot \ell_-)^{-2}$ (Fischer and Steel, 2009) we obtain the claimed bound by taking $\ell_- = S'_t$.

The second lower bound in (34) holds since (as noted earlier in that section) K grows exponentially with time at a rate that is proportional to $\nu \cdot \ell_+$ and for Yule trees we have $\ell_+ = t/2 + o(1)$. Finally, notice that $t/2 = \frac{\ln \overline{N}_t}{2\lambda}$ and so $e^{2c \cdot \ell_+} = \overline{N}_t^{c(\nu/\lambda)+o(1)}$.