Bulletin of Mathematical Biology (2022) 84:118
https://doi.org/10.1007/s11538-022-01072-w

Society for Mathematical Biology

ORIGINAL ARTICLE

The Shape of Phylogenies Under Phase-Type Distributed Times to Speciation and Extinction

Albert Ch. Soewongsono[1] · **Barbara R. Holland**[1] · **Małgorzata M. O'Reilly**[1]

Received: 13 October 2021 / Accepted: 29 August 2022 / Published online: 14 September 2022

Abstract

Phylogenetic trees describe relationships between extant species, but beyond that their shape and their relative branch lengths can provide information on broader evolutionary processes of speciation and extinction. However, currently many of the most widely used macro-evolutionary models make predictions about the shapes of phylogenetic trees that differ considerably from what is observed in empirical phylogenies. Here, we propose a flexible and biologically plausible macroevolutionary model for phylogenetic trees where times to speciation or extinction events are drawn from a Coxian phase-type (PH) distribution. First, we show that different choices of parameters in our model lead to a range of tree balances as measured by Aldous' β statistic. In particular, we demonstrate that it is possible to find parameters that correspond well to empirical tree balance. Next, we provide a natural extension of the β statistic to sets of trees. This extension produces less biased estimates of β compared to using the median β values from individual trees. Furthermore, we derive a likelihood expression for the probability of observing an edge-weighted tree under a model with speciation but no extinction. Finally, we illustrate the application of our model by performing both absolute and relative goodness-of-fit tests for two large empirical phylogenies (squamates and angiosperms) that compare models with Coxian PH distributed times to speciation with models that assume exponential or Weibull distributed waiting times. In our numerical analysis, we found that, in most cases, models assuming a Coxian PH distribution provided the best fit.

✉ Albert Ch. Soewongsono
albert.soewongsono@utas.edu.au

Barbara R. Holland
barbara.holland@utas.edu.au

Małgorzata M. O'Reilly
malgorzata.oreilly@utas.edu.au

[1] School of Natural Sciences (Discipline of Mathematics), University of Tasmania, Hobart 7005, Australia

Keywords Macro-evolutionary model · Diversification · Tree balance · Phase-type distribution

1 Introduction

Understanding how biodiversity is maintained and changed throughout time has been of long-standing interest in evolutionary biology (Quental and Marshall 2010; Morlon 2014). Fossil records are commonly used to make inferences about changes through time in speciation and extinction rates (Simpson 1944; Stanley 1998; Morlon et al. 2011). However, most clades do not possess sufficiently complete fossil records to make such inferences (Ricklefs 2007; Quental and Marshall 2010). In contrast, dated molecular trees are increasingly available; nevertheless, these "reconstructed phylogenies" only give relationships between extant species (Nee et al. 1992, 1994a; Stadler 2013b). These reconstructed phylogenies can also be used to study how diversification processes change throughout time (Nee et al. 1994a), although some have argued that the use of reconstructed phylogenies needs to be accompanied with availability of fossil records (Quental and Marshall 2010; Morlon 2014; Hagen et al. 2018). However, reconstructed phylogenies remain useful to study diversification and diversity dynamics when accompanied by biologically well-justified constraints (Louca and Pennell 2020).

Several mathematical models have been proposed for studying macroevolutionary processes. These range from the constant-rate birth and death (crBD) model where speciation and extinction rates are assumed to be constant through time (Nee et al. 1994b), to models where speciation and extinction rates change according to species age (Hagen et al. 2015), to models where an evolving trait can affect speciation and extinction rates (Maddison et al. 2007; FitzJohn 2012). For models under the general birth–death process, in which speciation and extinction rates can vary over time, a recent paper by Louca and Pennell (2020) shows that many parameter choices are indistinguishable as they generate the same expected lineage-through-time (LTT) plot. Despite the problems identified by Louca and Pennell (2020), these fitted parameters still provide some insight into speciation and extinction rates or structure of relationships between species through time (Harvey and Pagel 1991; Stadler 2013b).

Given a choice of a model, various methods can be applied to use empirical (or simulated) data such as branch lengths from reconstructed trees to estimate the parameters of the model. For example, it is possible to derive an expression for the likelihood of observing these branch lengths and find the best-fitting parameters of the model using maximum-likelihood estimation (MLE) to make inference about the speciation and extinction rates (Morlon et al. 2011). In order to see which model fits empirical data best, we can assess models via the likelihood ratio test (LRT) or the Akaike's Information Criterion (AIC) (Anderson and Burnham 2004) or via the comparison of their simulated LTT plot, which counts the number of species that existed at each given time in the past, with an empirical LTT plot (Morlon 2014). Then, given a model with best choice of parameters, we can assess whether it fits well to the empirical data by comparing tree balance or tree topology and branch length distributions from empirical and simulated trees generated from the model.

The balance of a phylogenetic tree describes the branching pattern of the tree, ranging from imbalanced shape where sister clades tend to be very different in sizes to balanced shape where the clades are of similar sizes. Tree balance is important for understanding macroevolutionary dynamics on a tree (Hagen et al. 2015) as it gives indication of heterogeneity of diversification rate across the tree without requiring information on branch lengths. Several statistics for assessing tree balance have been proposed in the literature. These include the Colless index (Colless 1982), the Sackin index (Sackin 1972) and Aldous' β (Aldous 1996)—Section 3.3 of Steel (2016) gives a detailed description of all three measures. In this paper, we focus exclusively on the β statistic as, unlike the other two statistics, it is easily comparable between trees of different size. The β statistic arises as a parameter of the Aldous' β-splitting model; in this model β is in the range $[-2, \infty)$ where values close to -2 mean that taxa are likely to split into unbalanced subsets and large values mean that splits are likely to be balanced. Many models in phylogenetics fail to resemble empirical datasets which often have β value around -1 (Aldous 1996). For example, the simplest macroevolutionary model is the pure birth model, also known as the Yule–Harding (YH) model (Yule 1925), where each species is equally likely to speciate. It has been shown that trees under this model have the expected value $\beta = 0$ (Aldous 1996; Hagen et al. 2015). In other words, the YH model predicts trees that are too balanced compared to empirical data (Aldous 1996, 2001). Likewise, models that include diversity-dependent (Etienne et al. 2012) and time-dependent speciation and extinction have been shown to produce the same expected tree balance as the YH model (Lambert and Stadler 2013). These models fall under a general class of species-speciation-exchangeable models as described in Stadler (2013b). This suggests that this class of models is not adequate to explain the macroevolutionary dynamics that has produced empirical trees.

Another statistic that has been widely used to compare empirical trees with macroevolutionary models is the γ statistic. The γ statistic was introduced in Pybus and Harvey (2000) and unlike the tree balance statistics it makes use of the branch lengths. The statistic is designed to have a zero mean standard normal distribution under a pure birth model. Negative values of γ mean that more diversification has occurred earlier in the tree than expected under a pure birth model, i.e., the edges nearer the root tend to be shorter relative to the other edges. Correspondingly, positive values of γ mean that more diversification has occurred later in the tree and that edges nearer the root tend to be relatively longer. It has been shown that γ values for empirical phylogenies tend to be below 0, which has sometimes been taken to indicate a slowdown in the diversification rate (Phillimore and Price 2008; Rabosky and Lovette 2008; Morlon et al. 2010).

In this paper, we construct a stochastic model for generating species phylogenies in which we apply Coxian PH distributions (Neuts 1981; Marshall and McClean 2004) for times to speciation and times to extinction. PH distributions describe the time to absorption in a continuous-time Markov chain (CTMC) with a single absorbing state and a finite number of non-absorbing states. Biologically, this could be thought of as a species passing through different phases where it may be more or less likely to speciate depending on a current underlying phase (Fig. 1). While these phases need not represent any particular biological state, the PH distribution gives great flexibility to model different ways that rates of speciation may depend on a species' age. Similarly,

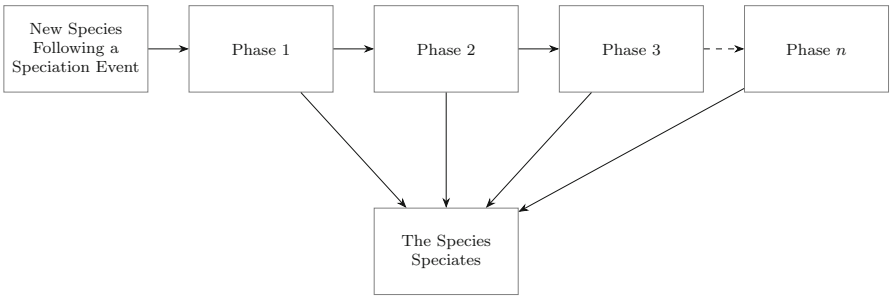

Fig. 1 A new species passes through different phases during its 'lifetime' until the next speciation event. Each phase corresponds to a non-absorbing state in a CTMC and speciation corresponds to the single absorbing state. At the start, the species directly goes to phase 1 where it can either undergo speciation or move to the next phase with certain rates. The process can continue up to a finite number of n phases, each corresponding to a different rate of speciation

times to extinction can also be modeled using PH distributions. We show that different parameter choices for age-dependent speciation rates produce phylogenetic trees that can range from highly balanced to highly unbalanced. In particular, we find parameters that give similar tree balance statistics to empirical trees.

An additional contribution of the paper is that we develop a new approach for computing the β statistic based on a set of trees rather than computing β from a single tree. We suggest that this approach leads to more accurate estimates of the β statistic compared to computing β for single trees and then taking an average and that this is particularly true for trees with fewer extant species.

For a special case of our model, in which only speciation (and not extinction) occurs, we derive a likelihood expression for the probability of observing any edge-weighted tree. For two very large phylogenies—squamates (Zheng and Wiens 2016) and angiosperms (Zanne et al. 2014)—we perform model selection for different clades of both trees to compare our Coxian PH model for the speciation process to the exponential and Weibull distributions.

The rest of our paper is structured as follows. In the mathematical methods section we: (1) summarize the key properties of the PH distribution, (2) introduce some examples of Coxian PH distributions, (3) present our method for calculating the β statistic for a set of trees, (4) and derive a likelihood expression based on our model for fitting empirical branch length data. The next section contains simulations that: (1) demonstrate the use of treeset β, (2) show that the model can produce trees with a wide range of tree shapes, (3) examine how well fitted models do in recovering the speciation process in scenarios with and without extinction. In the section on empirical data we apply our model to two large published phylogenies—squamates (Zheng and Wiens 2016) and angiosperms (Zanne et al. 2014). In summary, we find that Coxian PH distributions are a useful tool for studying macroevolutionary dynamics.

2 Mathematical Methods

2.1 PH Distribution and Relevant Properties

In this section, we introduce the PH distribution and some of its key properties.

Definition 1 (*Continuous PH distributions*) Let $\{X(t) : t \geq 0\}$ be a continuous time Markov chain defined on state space $S = \hat{S} \bigcup \{n+1\}$, where $\hat{S} = \{1, 2, \ldots, n\}$ is the set of non-absorbing states and $n+1$ is an absorbing state, initial distribution vector $\boldsymbol{\alpha} = [\alpha_i]_{i \in \hat{S}}$, and generator matrix

$$\mathbf{Q}^* = [Q^*_{i,j}]_{i,j \in S} = \begin{bmatrix} \mathbf{Q} & \boldsymbol{q} \\ \mathbf{0} & 0 \end{bmatrix}, \tag{1}$$

where $\mathbf{Q} = [Q_{i,j}]_{i,j \in \hat{S}}$ is a square matrix with dimension n that records the transition rates between non-absorbing states $i, j \in \hat{S}$, $\boldsymbol{q} = [Q_{i,n+1}]_{i \in \hat{S}}$ is a column vector that records the transition rates from non-absorbing states $i \in \hat{S}$ to the absorbing state $n+1$, and $\mathbf{0}$ is the row vector with corresponding dimension. By the definition of generator matrix $\mathbf{Q}$, we have $Q_{i,i} < 0$, for all i, $Q_{i,j} \geq 0$ for $i \neq j$, and $\mathbf{Q1} + \boldsymbol{q} = \mathbf{0}$, where $\boldsymbol{q}$ is the exit rate vector.

Let $Z = \inf\{t \geq 0 : X(t) = n+1\}$ be the random variable recording the time until absorption, then Z is said to be continuous PH distributed with parameters $\boldsymbol{\alpha}$ and $\mathbf{Q}$, which we denote $Z \sim \text{PH}(\boldsymbol{\alpha}, \mathbf{Q})$.

Theorem 1 (The cumulative distribution and density functions of continuous PH distribution) *Suppose* $Z \sim \text{PH}(\boldsymbol{\alpha}, \mathbf{Q})$*, then the cumulative distribution and the probability density function of* Z *are given, respectively, by*

$$F_Z(z) = 1 - \boldsymbol{\alpha} e^{\mathbf{Q}z} \mathbf{1}, \tag{2}$$

$$f_Z(z) = \boldsymbol{\alpha} e^{\mathbf{Q}z} \boldsymbol{q}, \tag{3}$$

and its mean and variance are given by

$$E(Z) = -\boldsymbol{\alpha} \mathbf{Q}^{-1} \mathbf{1}, \tag{4}$$

$$Var(Z) = 2\boldsymbol{\alpha} \mathbf{Q}^{-2} \mathbf{1} - \left(\boldsymbol{\alpha} \mathbf{Q}^{-1} \mathbf{1}\right)^2. \tag{5}$$

Proof of this theorem is originally given in Neuts (1975)*, and a clear exposition is given in* Verbelen (2013). □

Definition 2 (*Coxian PH distribution*) If $\boldsymbol{\alpha}$ and $\mathbf{Q}$ are defined as

$$\boldsymbol{\alpha} = [1, 0, \ldots, 0], \tag{6}$$

$$\mathbf{Q} = \begin{bmatrix} -\lambda_1 & p_1\lambda_1 & 0 & \ldots & 0 & 0 \\ 0 & -\lambda_2 & p_2\lambda_2 & \ddots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \ddots & -\lambda_{n-2} & p_{n-2}\lambda_{n-2} & 0 \\ 0 & 0 & \ldots & 0 & -\lambda_{n-1} & p_{n-1}\lambda_{n-1} \\ 0 & 0 & \ldots & 0 & 0 & -\lambda_n \end{bmatrix}, \tag{7}$$

where $0 < p_i \leq 1$ and $\lambda_1, \ldots, \lambda_n > 0$ for all $i = 1, 2, \ldots n-1$, then we say that the random variable $T \sim \text{PH}(\boldsymbol{\alpha}, \mathbf{Q})$ follows Coxian PH distribution.

Cumani (1982) showed that any acyclic PH (APH) distribution (including Coxian PH distributions), that is, a distribution with an upper triangular generator matrix (Asmussen et al. 1996), can be restructured to a canonical form such as shown above and thus only requires $2n$ parameters as opposed to $n^2 + n$ parameters for a general PH distribution. This reduction in the number of parameters makes it computationally simpler to fit parameters (Thummler et al. 2006). Further, Cumani (1982) and Dehon and Latouche (1982) showed that for any APH distribution, there exists an equivalent representation as a Coxian PH distribution with $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$.

To fit a PH distribution to data it is necessary to fix the number of non-absorbing states. Thummler et al. (2006) stated that it is difficult to fit general PH distributions if the number of non-absorbing states is larger than four, due to the increased computational cost and the dependence on the initial values. They also state that having a PH distribution of low order (less than four non-absorbing states) is not sufficient to get parameter values that correspond to small coefficients of variation (CV).

In Sects. 2.3 and 3.2 where we simulate data under different conditions, we focus solely on PH distributions with four non-absorbing states. In Sect. 4, where we fit models to empirical data, we explore a wider range of options for the number of non-absorbing states.

2.2 Coxian-Based Macro-Evolutionary Model

Now, we develop a stochastic model for generating species phylogenies, in which we assume that the time spent by each newly formed lineage before the next speciation or extinction event is drawn from a Coxian PH distribution. Our model is a special case of the well-studied Bellman–Harris model which allows any distribution of waiting times to extinction or speciation (Bellman and Harris 1948). This model is discussed in Hagen and Stadler (2018) and they provide an *R* package (Hagen and Stadler 2018) that allows users to simulate trees under a general Bellman–Harris model. However, while it is possible to simulate trees under this very general class of models, it is not possible to fit parameters of a general Bellman–Harris distribution to empirical data. A novelty of our approach is that we are able derive a likelihood expression for the

probability of observing a reconstructed phylogeny under our model in the case with no extinction and that we can therefore fit parameters.

In our model, we primarily focus on symmetric speciation. This means that after a speciation event two "child" species are created that are identical and of age 0. Thus, each branch length on a given tree can be thought of as an independent random variable drawn from the imposed Coxian PH distribution. We also consider asymmetric speciation in which the "parent" species is considered to continue and one new "child" species is created with age 0. Both symmetric and asymmetric speciation modes are supported by the *R* package *TreeSimGM* (Hagen and Stadler 2018).

We also construct two examples of the Coxian PH distribution as given in Definition 2. We parameterize the two examples so as to enforce either monotonically increasing or monotonically decreasing rates of absorption. In Example 1, the rate of speciation (or extinction) decreases as species get older, and in Example 2 the rate of speciation (or extinction) increases as species get older. We chose a parameterization with three free variables (x, y and z), as this gives flexibility to pick instances of each example with a given mean and variance, while at the same time reducing the number of free parameters for faster computational time (Okamura and Dohi 2016). Moreover, these two examples follow canonical form 3 of an APH distribution as stated in Okamura and Dohi (2016) (see also the derivation of the form by Cumani 1982). Note that there are different parameterizations that can be derived from the general Coxian PH distribution defined in Definition 2 which have either decreasing or increasing rate. However, these particular examples still provide some flexibility to choose different parameter values that give a wide range of coefficients of variation (CV) needed in Sect. 3.2.

Example 1 (*Coxian PH Distributed Model for Decreasing Rate*)

$$\mathbf{Q} = \begin{bmatrix} -z & (1-y)z & 0 & 0 \\ 0 & -(1+x) & \left(1-y^2\right)(1+x) & 0 \\ 0 & 0 & -\left(1+x^2\right) & \left(1-y^3\right)\left(1+x^2\right) \\ 0 & 0 & 0 & -x^3 \end{bmatrix}, \boldsymbol{q} = \begin{bmatrix} yz \\ y^2(1+x) \\ y^3\left(1+x^2\right) \\ x^3 \end{bmatrix}, \tag{8}$$

where $0 < x \leq 1$, $0 < y < 1$, $z \geq 2$ and $\boldsymbol{q}$ is the exit rate vector.

The restrictions on x and y imply that each entry of the exit rate vector $\boldsymbol{q}$ is less than the preceding entry.

Example 2 (*Coxian PH Distributed Model for Increasing Rate*)

$$\mathbf{Q} = \begin{bmatrix} -\left(1+x^3\right) & \left(1-y^4\right)\left(1+x^3\right) & 0 & 0 \\ 0 & -\left(1+x^2\right) & \left(1-y^3\right)\left(1+x^2\right) & 0 \\ 0 & 0 & -(1+x) & \left(1-y^2\right)(1+x) \\ 0 & 0 & 0 & -z \end{bmatrix},$$
$$\boldsymbol{q} = \begin{bmatrix} y^4\left(1+x^3\right) \\ y^3\left(1+x^2\right) \\ y^2(1+x) \\ z \end{bmatrix}, \tag{9}$$

where $0 < x \leq 1$, $0 < y < 1$, $z \geq 2$ and $\boldsymbol{q}$ is the exit rate vector.

Here, the restrictions on x and y imply that each entry of the exit rate vector $\boldsymbol{q}$ is greater than the preceding entry.

From now on, we refer Examples 1 and 2 as PH_{Dec} and PH_{Inc}, respectively. By standard theory of the PH distribution, the first and second moments of the Coxian PH distribution in PH_{Dec} and PH_{Inc} are given by

$$\begin{aligned} \mathbb{E}_{\text{PH}}(X) &= \frac{1}{z} + (1-y)\left(\frac{1}{1+x} + \left(1-y^2\right)\left(\frac{1}{1+x^2} + \frac{1-y^3}{x^3}\right)\right), \\ \mathbb{E}_{\text{PH}}\left(X^2\right) &= \frac{2}{z^2} + \frac{2(1-y)}{1+x} \\ &\quad \left(\frac{1}{z} + \frac{1}{1+x}\right) + \frac{2(1-y)\left(1-y^2\right)}{1+x^2}\left(\frac{1}{z} + \frac{1}{1+x} + \frac{1}{1+x^2}\right) \\ &\quad + \frac{2(1-y)\left(1-y^2\right)\left(1-y^3\right)}{x^3}\left(\frac{1}{z} + \frac{1}{1+x} + \frac{1}{1+x^2} + \frac{1}{x^3}\right), \end{aligned} \tag{10}$$

and

$$\begin{aligned} \mathbb{E}_{\text{PH}}(X) &= \frac{1}{1+x^3} + \left(1-y^4\right)\left(\frac{1}{1+x^2} + \left(1-y^3\right)\left(\frac{1}{1+x} + \frac{1-y^2}{z}\right)\right), \\ \mathbb{E}_{\text{PH}}\left(X^2\right) &= \frac{2}{\left(1+x^3\right)^2} + \frac{2\left(1-y^4\right)}{1+x^2}\left(\frac{1}{1+x^3} + \frac{1}{1+x^2}\right) + \frac{2\left(1-y^4\right)\left(1-y^3\right)}{1+x} \\ &\quad \left(\frac{1}{1+x^3} + \frac{1}{1+x^2} + \frac{1}{1+x}\right) + \frac{2\left(1-y^4\right)\left(1-y^3\right)\left(1-y^2\right)}{z} \\ &\quad \left(\frac{1}{1+x^3} + \frac{1}{1+x^2} + \frac{1}{1+x} + \frac{1}{z}\right), \end{aligned} \tag{11}$$

respectively. The derivations of Eqs. 10 and 11 are shown in "Appendix."

2.3 Computing β for a Set of Trees

We propose a new approach for estimating the tree-balance statistic β from a set of rooted trees $\{T_1, \ldots, T_M\}$, which can be either empirical trees or simulated trees under some model of interest. For each subtree with four or more tips in each tree in $\{T_1, \ldots, T_M\}$ we compute the probability $q_n(i, \beta)$ of observing i tips on the left out of the n tips of that subtree. This is done using Eq. 4 from Aldous (1996),

$$q_n(i, \beta) = \frac{1}{a_n(\beta)} \frac{\Gamma(\beta + i + 1)\Gamma(\beta + n - i + 1)}{\Gamma(i + 1)\Gamma(n - i + 1)}, 1 \leq i \leq n - 1, \tag{12}$$

where $a_n(\beta)$ is the normalizing constant. We note that subtrees of size 2 or 3 are not of interest as there is only one possible division of the tips. In the case where the tree size is too large, the above expression is not numerically tractable, so we use the following approximation instead (which is also used in the *apTreeShape* package (Bortolussi et al. 2006)), given by

$$q_n(i, \beta) = \frac{1}{\hat{a}_n(\beta)} \left(\frac{i}{n}\right)^{\beta} \left(1 - \frac{i}{n}\right)^{\beta}, \tag{13}$$

where $\hat{a}_n(\beta)$ is the normalizing constant. (Justification for the approximation in Eq. 13 is given in "Appendix.")

We then use numerical optimization to find the value of β in the range $[-2, 10]$ which maximizes the product of all the $q_n(i, \beta)$ values. This is the maximum likelihood estimate of β for the set of trees. Our custom *R* script, based on *maxlik.betasplit* function from the *apTreeShape* package (Bortolussi et al. 2006) to estimate β from sets of trees, is available as a Supplementary Material on Dryad (https://doi.org/10.5061/dryad.w9ghx3fpk).

2.4 Fitting PH Distributions to Branch Length Data

In this section, we propose a method for finding parameters of a PH distribution using branch length data from a phylogenetic tree. We assume that the time until a speciation event on a branch follows a PH distribution and that there is no extinction. We write the likelihood expression using parameters from the PH distribution to calculate the probability of observing a tree with a given number of extant species.

Assuming that a tree evolves under a symmetric speciation mode, and that times to speciation events are drawn from a PH distribution, we can treat each branch length on the tree as independently drawn from the same PH distribution. We illustrate this in Fig. 2, in which the lengths of internal branches and pendant branches are denoted by $\{b_1, b_2, b_3, b_4\}$ and $\{\tilde{b}_1, \tilde{b}_2, \tilde{b}_3, \tilde{b}_4, \tilde{b}_5\}$, respectively.

In general, we denote the lengths of internal and pendant branches by b_i, for $i = 1, \ldots, k$, and $\tilde{b}_j$, for $j = 1, \ldots, \ell$, where the total number of internal branches and pendant branches is denoted by k and ℓ, respectively. Here, because we consider the root branch, we note that $k = \ell - 1$. Both internal and pendant branches follow a

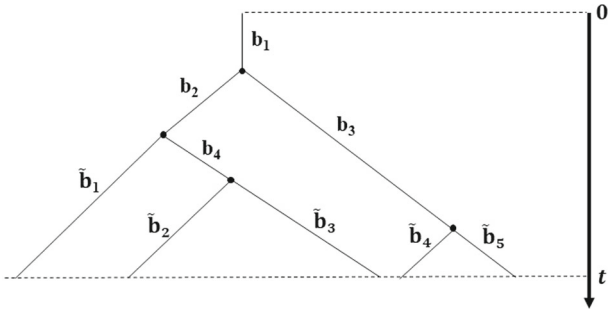

Fig. 2 Phylogenetic tree with five extant tips evolving under a symmetric speciation mode. Branch lengths are independent and drawn from the same PH distribution PH ($\boldsymbol{\alpha}$, $\mathbf{Q}$)

PH distribution with parameter $\boldsymbol{\alpha}$ and rate matrix $\mathbf{Q}$, that is, $b_i, \tilde{b}_j \sim \text{PH}(\boldsymbol{\alpha}, \mathbf{Q})$. It follows from the properties of the PH distribution (Neuts 1981), that the likelihood of observing an internal branch of length b_i is the probability density of the distribution along the branch given by $\boldsymbol{\alpha} e^{\mathbf{Q}b_i}\boldsymbol{q}$ and the likelihood of observing a pendant branch of length $\tilde{b}_j$ is the probability that the branch has survived until time t (i.e., one minus the cumulative probability of the distribution) given by $\boldsymbol{\alpha} e^{\mathbf{Q}\tilde{b}_j}\mathbf{1}$, where $\mathbf{1}$ is a column vector of ones. Therefore, by independence of the branch lengths, the likelihood of observing tree T can be written as,

$$\mathcal{L}(T \mid \boldsymbol{\alpha}, \mathbf{Q}) = \prod_{i=1}^{k} \left(\boldsymbol{\alpha} e^{\mathbf{Q}b_i}\boldsymbol{q}\right) \times \prod_{j=1}^{\ell} \left(\boldsymbol{\alpha} e^{\mathbf{Q}\tilde{b}_j}\mathbf{1}\right), \tag{14}$$

with $\boldsymbol{\alpha} = [1, 0, \ldots, 0]$, since we apply Coxian PH distribution. Note that if we consider all the possible permutations on the tips of the tree, then the likelihood becomes,

$$\mathcal{L}(T \mid \boldsymbol{\alpha}, \mathbf{Q}) = (\ell - 1)! \times \prod_{i=1}^{k} \left(\boldsymbol{\alpha} e^{\mathbf{Q}b_i}\boldsymbol{q}\right) \times \prod_{j=1}^{\ell} \left(\boldsymbol{\alpha} e^{\mathbf{Q}\tilde{b}_j}\mathbf{1}\right). \tag{15}$$

Given the branch lengths of a *single tree* T, we perform numerical optimization to find parameter values that maximize the likelihood equation given in Eq. 14. In the case of the general Coxian PH model this amounts to finding the best values of p_i's and λ_i's as in Definition 2, for PH_{Dec} and PH_{Inc} it means finding the best values of x, y, and z.

Alternatively, given the branch lengths of a *tree set* $\{T_1, \ldots, T_M\}$, we apply maximum likelihood estimation to maximize the product

$$\mathcal{L}(\{T_1, \ldots, T_M\} \mid \boldsymbol{\alpha}, \mathbf{Q}) = \mathcal{L}(T_1 \mid \boldsymbol{\alpha}, \mathbf{Q}) \times \cdots \times \mathcal{L}(T_M \mid \boldsymbol{\alpha}, \mathbf{Q}), \tag{16}$$

where we assume trees are independent and apply Eq. 14 to compute the likelihood of observing the individual trees $T_1, \ldots, T_M$.

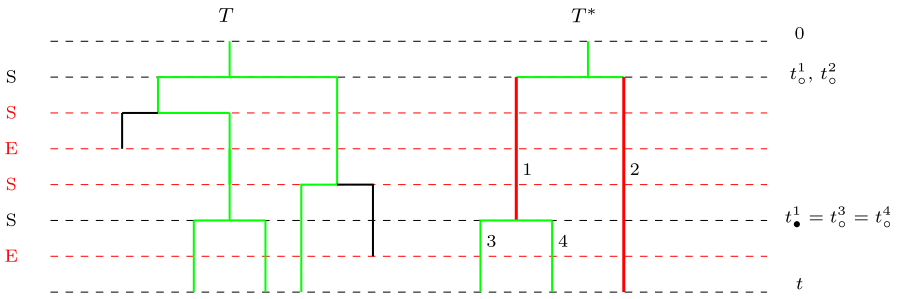

Fig. 3 Original tree T and the reconstructed tree T^*, where S (in black) denotes observed speciation events and E (in red) denotes unobserved extinction events. Speciation events S (in red) followed by extinction events E (in red) on the original tree T, are not observed on the branches (highlighted in red) of the reconstructed tree T^*. The internal branch 1 on T^* was born at time $t^1_\circ$ and the next *observed* speciation event on that branch (on T^*) was at time $t^1_\bullet$. The external branch 2, 3, and 4 on T^* was born at time $t^2_\circ$, $t^3_\circ$, and $t^4_\circ$, respectively, and no speciation events are observed on that branch (on T^*) (Color figure online)

To optimize parameters for the exponential and Weibull distribution, we derive an equivalent expression to Eq. 14 for both distributions. The likelihood expression for the exponential distribution is given by

$$\mathcal{L}(T \mid \lambda) = \prod_{i=1}^{k} \lambda \exp^{-\lambda b_i} \times \prod_{j=1}^{\ell} \exp^{-\lambda \tilde{b}_j}, \tag{17}$$

and for the Weibull distribution

$$\mathcal{L}(T \mid \psi, \phi) = \prod_{i=1}^{k} \frac{\psi}{\phi} \left(\frac{b_i}{\phi}\right)^{\psi-1} \exp^{-(b_i/\phi)^{\psi}} \times \prod_{j=1}^{\ell} \exp^{-\left(\tilde{b}_j/\phi\right)^{\psi}}, \tag{18}$$

where ψ and ϕ are scale and shape parameters, respectively.

Then, we apply maximum likelihood estimation to search for $\lambda > 0$ that maximizes Eq. 16. Similarly, we search for $\psi > 0$ and $\phi > 0$ parameters that maximize Eq. 16.

Finally, we consider a birth-and-death process (BDP) with constant birth rate λ and constant death rate μ. The likelihood expression for the reconstructed tree under such BDP is given in Eq. 20 of Nee et al. (1994b), it is a conditional probability conditioning on the survival of both original branches descending from the root.

Note that the likelihood for the reconstructed tree under any process that includes extinction events needs to consider the possibility that speciation events that end with extinction may occur on internal or external branches and so are not observed on the reconstructed tree (see Fig. 3).

Below, we present our alternative likelihood formula for the reconstructed tree under a BDP. This formula provides new physical interpretations given by Eqs. 19–23, in the context of the dynamics of the process driving the evolution of the phylogenetic tree in time.

Assume that t is the age of the tree with 0 is the time at the start of the root branch and let x_i be the elapsed time from the end of the internal branch i until the end of tree T. That is, if internal branch i is born at time $t_\circ^i$ and gives birth at time $t_\bullet^i$ to another branch, then $x_i = t - t_\bullet^i$ and its length is $b_i = t_\bullet^i - t_\circ^i$. For the external branch j descending from the internal branch i, we have its branch length given by $\tilde{b}_j = t - t_\bullet^i = x_i$.

Then, the likelihood of observing a reconstructed species tree T^* is given by

$$\mathcal{L}(T^* \mid \lambda, \mu) = (\ell - 1)! \prod_{i=1}^{k} G_{x_i,t}(b_i)\lambda \prod_{j=1}^{\ell} D_t^{(1)}(\tilde{b}_j), \tag{19}$$

where $G_{x_i,t}(b_i)$ is the probability of observing reconstructed internal branch i, and $D_t^{(1)}(\tilde{b}_j)$ is the probability of observing reconstructed external branch j, where $G_{x,t}(z)$ is the solution of

$$G_{x,t}(z) = e^{-(\lambda+\mu)z} + \int_{u=0}^{z} e^{-(\lambda+\mu)(z-u)}\lambda \left(2G_{x,t}(u)E(u+x)\right) \mathrm{d}u, \tag{20}$$

$$\frac{\mathrm{d}G_{x,t}(z)}{\mathrm{d}z} = -(\lambda+\mu)G_{x,t}(z) + 2\lambda G_{x,t}(z)E(z+x), \tag{21}$$

and $D_t^{(1)}(z)$ is the solution of

$$D_t^{(1)}(z) = e^{-(\lambda+\mu)z} + \int_{u=0}^{z} e^{-(\lambda+\mu)u}\lambda \left(2D_t^{(1)}(z-u)E(z-u)\right) \mathrm{d}u, \tag{22}$$

$$\frac{\mathrm{d}D_t^{(1)}(z)}{\mathrm{d}z} = -(\lambda+\mu)D_t^{(1)}(z) + 2\lambda E(z)D_t^{(1)}(z), \tag{23}$$

where by Kendall (1948)

$$E(z) = \frac{\mu - \mu e^{(\mu-\lambda)z}}{\lambda - \mu e^{(\mu-\lambda)z}} \tag{24}$$

is the probability that a branch born at time zero becomes extinct by time z. Solving the above equations gives

$$G_{x,t}(z) = \left(\frac{\lambda - \mu e^{(\mu-\lambda)x}}{\lambda - \mu e^{(\mu-\lambda)(z+x)}}\right)^2 e^{(\mu-\lambda)z}, \tag{25}$$

$$D_t^{(1)}(z) = \left(\frac{(\lambda-\mu)e^{\mu z}}{\lambda - \mu e^{(\mu-\lambda)z}}\right)^2 e^{-(\lambda+\mu)z}. \tag{26}$$

The derivation of the differential equations for $D_t^{(1)}(z)$ and $G_{x,t}(z)$ along with their solutions and some intuition are shown in "Appendices 6.4 and 6.5."

Next, we apply our likelihood expression in Eq. 19 to the reconstructed tree T^* in Fig. 3 (ignoring the age of the root) to see that

$$\mathcal{L}_{\text{Nee}}(T^* \mid \lambda, \mu) = \frac{\mathcal{L}(T^* \mid \lambda, \mu)}{(1 - E(x_2))^2}, \tag{27}$$

where $\mathcal{L}_{\text{Nee}}(T^* \mid \lambda, \mu)$ is the likelihood expression given in Eq. 20 in Nee et al. (1994b), and x_2 is the elapsed time from the starting time of the two original branches descending from the root until the end of the tree T^*, as defined in Nee et al. (1994b). This relationship is as expected, since $\mathcal{L}_{\text{Nee}}(T^* \mid \lambda, \mu)$ is a conditional probability of observing the tree T^* given that both original branches have survived until the end of the tree.

3 Simulations

3.1 Comparing Treeset β to the Standard β

To compare β values estimated from individual trees to those estimated for a set of trees, we performed the following simulation. We simulated sets of 1000 trees using *TreeSimGM* package (Hagen and Stadler 2018), where each set of trees had the same number of extant tips $n \in \{10, 20, 30, \ldots, 200\}$ and their times to speciation were drawn from PH distribution with rate matrix

$$\mathbf{Q} = \begin{bmatrix} -2 & 1 & 0 & 0 \\ 0 & -1.1 & 1 & 0 \\ 0 & 0 & -1.01 & 1 \\ 0 & 0 & 0 & -0.001 \end{bmatrix}, \quad \boldsymbol{q} = \begin{bmatrix} 1 \\ 0.1 \\ 0.01 \\ 0.001 \end{bmatrix}. \tag{28}$$

We note that the structure of the exit rate vector $\boldsymbol{q}$ implies that the probability of getting absorbed from later states is less likely than from earlier states. We then repeated the above procedure for sets of trees evolving under the YH model. The YH case is interesting because it is representative of a wider class of models that are known to have $E(\beta) = 0$ (Aldous 2001).

For each set of trees, we computed individual estimates of β for each tree as well as a β estimate for the entire tree set. We also computed 95% confidence intervals for the estimated β values, denoted $\hat{\beta}$, from each tree set. In order to get the lower and upper bound for the confidence intervals, we performed a numerical search over 500 equidistant points between $\hat{\beta} - 5 \times SE\left(\hat{\beta}\right)$ and $\hat{\beta}$ to find the point that corresponds to the lower bound and 500 equidistant points between $\hat{\beta}$ and $\hat{\beta} + 5 \times SE\left(\hat{\beta}\right)$ to find the point that corresponds to the upper bound. The lower and upper bounds were chosen such that their likelihood is equal to the likelihood of the MLE minus a half of the chi-square value with 1 degree of freedom; this gives a 95% confidence interval

(Pawitan 2001). The standard error for $\hat{\beta}$, $SE\left(\hat{\beta}\right)$, was evaluated using

$$SE\left(\hat{\beta}\right) = \frac{1}{\sqrt{I\left(\hat{\beta}\right)}}, \tag{29}$$

where $I\left(\hat{\beta}\right)$ is the Fisher information of $\hat{\beta}$.

The results are summarized in Fig. 4. For both of the generating processes, the distribution of β values is right-skewed (Fig. 4a, c) and the median value for individual trees is higher than the value estimated using the entire tree set particularly for trees with fewer tips (Fig. 4b, d). For the trees generated under the YH process, when estimating the value of β for trees with fewer extant tips we obtained $\beta \approx 0$ when applying the method based on treesets, but median $\beta > 0$ for estimates based on individual trees (Fig. 4c, d). We conclude that the method based on treesets is more accurate for the Yule process, as evidenced by the 95% confidence interval in Fig. 4d. The β values estimated from different sets of trees concentrate around $\beta = 0$ in agreement with the theoretical value for trees evolving under the YH model. We think that the upwards bias in estimation of β arises because, for trees with fewer tips, it not unlikely to get a tree that is maximally balanced (or close to it) and in this case the maximum likelihood procedure for fitting β prefers to make β as large as possible.

3.2 Coxian-PH Models can Generate a Range of Tree Shapes

In Hagen et al. (2015), the authors found that using a Weibull distribution for age dependent speciation had an effect on tree balance (as measured by the β statistic), whereas using a Weibull distribution for extinction had an effect on diversification (as measured by the γ statistic). To test if using PH distributions gives similar results, we simulated trees using the two examples PH_{Dec} and PH_{Inc}. We did not see obvious changes in the β and γ statistics under different parameter values using PH_{Inc}, so we only report results for PH_{Dec}. The simulation procedure was a follows:

- As an example, we set $z = 10$ and mean waiting time to both speciation and extinction $\mathbb{E}_{\text{PH}}(X) = 2$. The choice of $\mathbb{E}_{\text{PH}}(X)$ scales the branch lengths of generated phylogenies, but results will be invariant to this choice of the mean since we only consider tree balance and relative branch lengths. Likewise, the z parameter is chosen arbitrarily as long as it is larger than or equal to 2 in order to preserve a decreasing rate as described in PH_{Dec}.
- We then selected 4 pairs of parameters $0 < x \leq 1$ and $0 < y < 1$ to give a wide range of coefficients of variation (CV). We found choices of x and y where $\text{CV} = \frac{\sigma}{\mu} \in \{30.08, 13.50, 5.56, 1.49\}$. These 4 pairs of x and y are as follows: $(x, y) \in \{(0.1, 0.93), (0.17, 0.88), (0.3, 0.78), (0.68, 0.45)\}$. We also note that fixing either x or y parameters gives less flexibility in choosing (y, z) or (x, z) pairs corresponding to a wide range of CV.
- Using the *TreeSimGM* package (Hagen and Stadler 2018) in *R*, we generated 300 trees with 100 extant tips in which times to speciation followed a PH distribution

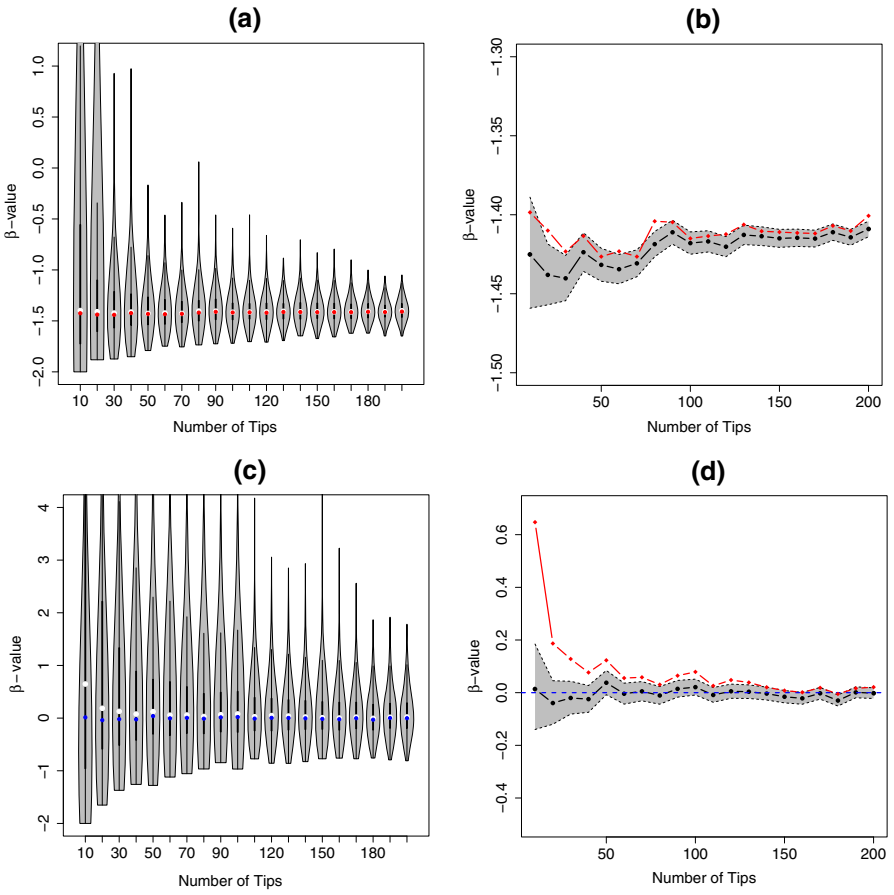

Fig. 4 Estimates of β for individual trees with $n \in \{10, 20, 30, \ldots, 200\}$ tips (**a**–**d**). Estimates of β from treesets are indicated by red dots (**a**) and blue dots (**c**). Trees are simulated according to either Coxian PH distribution for times to speciation events (**a**) or the YH process (**c**). The area of 95% confidence interval of β values from treesets following Coxian PH distribution and the YH process are plotted in (**b**), (**d**), respectively. The black lines represent the treeset β values, and the gray area represents the confidence interval for each treeset β value. The red lines represent the median β values from individual trees. The blue-dashed line represents the theoretical β value for the YH trees ($\beta = 0$) (Color figure online)

with parameters x, y and z, while times to extinction followed an exponential distribution with rate $\lambda = 0.25$. The main goal in choosing trees of size 100 was to have trees that were large enough for β to be accurately estimated for individual trees, but small enough to have reasonable running time. We repeated this procedure for both symmetric and asymmetric speciation modes. Then we repeated everything again but using an exponential distribution for the times to speciation (with $\lambda = 20$) and the PH distributions described above for the times to extinction.

– We measured the effect of different parameter choices above on tree balance using the β statistic. We computed the β statistic both for individual trees, using the *apTreeshape* package (Bortolussi et al. 2006), and for sets of trees based on our

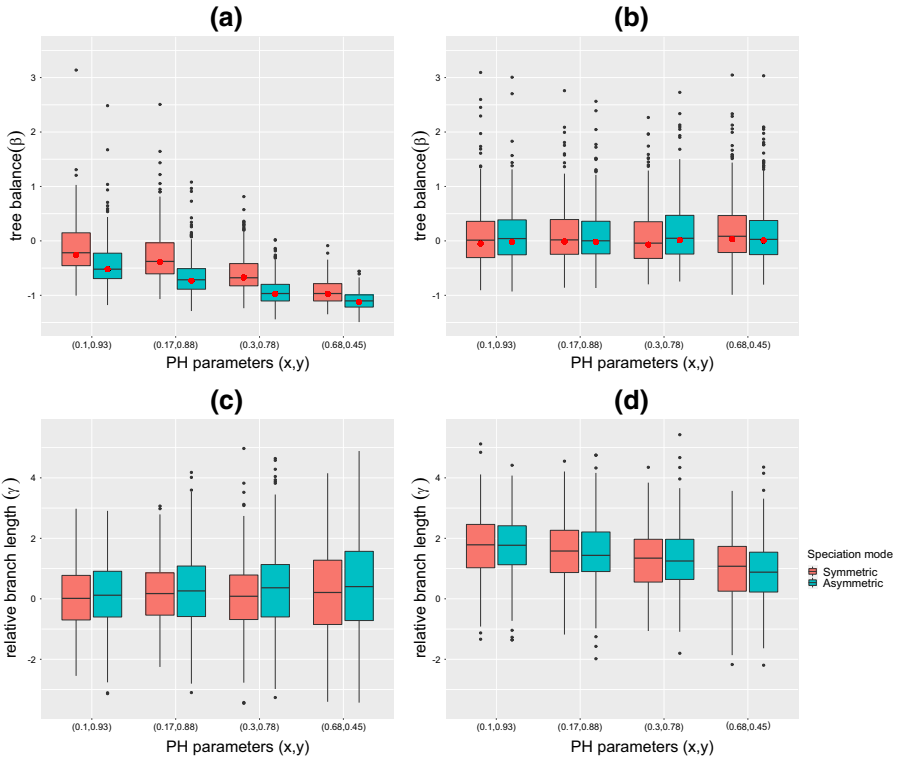

Fig. 5 Effect of speciation and extinction processes on tree balance as measured by the β statistic, and on relative branch lengths as measured by the γ statistic. For each pair of parameters (x, y) in $\mathrm{PH}_{\mathrm{Dec}}$ used to generate either times to speciation (in **a**, **c**) or times to extinction (in **b**, **d**), we simulated 300 trees with 100 extant tips. In **a**, **c** the parameters of the times to speciation are $(x, y) \in \{(0.1, 0.93), (0.17, 0.88), (0.3, 0.78), (0.68, 0.45)\}$ and mean speciation time is $\mathbb{E}_{\mathrm{PH}}(X) = 2$, while times to extinction are drawn from exponential distribution with rate $\lambda = 0.25$. In **b**, **d** times to speciation are drawn from exponential distribution with rate $\lambda = 20$, while the parameters of the times to extinctions are $(x, y) \in \{(0.1, 0.93), (0.17, 0.88), (0.3, 0.78), (0.68, 0.45)\}$ and mean extinction time is $\mathbb{E}_{\mathrm{PH}}(X) = 2$. The red dots show the β statistic for sets of trees (Color figure online)

new approach. We also measured the effect on relative branch lengths as measured by the γ statistic (Pybus and Harvey 2000), which we computed using the *APE* package (Paradis et al. 2004).

The results are presented in Fig. 5. Tree balance is affected by varying the parameters for times to speciation (Fig. 5a), in particular, there are choices of model parameters that match the tree-shape statistics of empirical phylogenies ($\beta = -1$). Tree balance is not significantly affected by the parameters for times to extinction (Fig. 5b). In contrast to the behavior of β, relative branch lengths, as measured by the γ statistic are not affected by the parameters for times to speciation (Fig. 5c), while they are affected by the parameters for times to extinction (Fig. 5d). We did not observe a significant difference in our results between the symmetric and asymmetric speciation modes. These results are congruent with what was found in Hagen et al. (2015).

Table 1 KS tests for hypothesis testing that both fitted and simulated log branch lengths in Fig. 6, 7 and 8 come from the same distribution

Extinction rate λ	KS Statistic	p value
0	0.025	0.101
0.1	0.034	0.007
0.4	0.115	$\ll 0.001$

3.3 Fitting Coxian-PH Distributions to Branch Length Data

In this section, we test if the maximum likelihood approach outlined in Sect. 2.4 is able to fit the speciation process well in cases where: (a) there is no extinction, and (b) the generating model includes extinction. As an example to illustrate the bias introduced by not considering the extinction process in the likelihood function in Eq. 16, we simulated trees using the PH_{Dec} distribution with known parameter values, for the speciation process and an exponential distribution for the extinction process with rate $\lambda \in \{0, 0.1, 0.4\}$, and then fitted the parameters of the PH_{Dec} distribution to the generated branch length data. In total, we generated 50 trees with 50 extant tips each, using *TreeSimGM* package (Hagen and Stadler 2018), which produced 4900 branches.

Using Eq. 14–16, we found the parameters x, y, and z that maximized the likelihood of observing the given set of branch lengths. The optimization was carried out using the built-in *R* function, *optim*, with the "*L-BFGS-B*" method (Byrd et al. 1995) and multiple starting points for x, y, z, followed by local optimization using the "Nelder-Mead" method (Nelder and Mead 1965).

To compare the fitted distribution to the generating distribution we plotted the density of the fitted distribution and the known distribution used to simulate the data. Additionally, using the fitted parameters x, y and z, we generated trees with the same number of tips as in the simulated data, and compared their distribution of branch lengths with that of the simulated trees. Note that we cannot simply compare the branch length histogram from trees generated under the known distribution with its fitted frequency density plot since the generated trees are truncated at some time t (the tree's age). Therefore, to compare distributions of branch lengths we used the two sample Kolmogorov-Smirnov (KS) test of the null hypothesis that both simulated and fitted log branch lengths come from the same distribution (using the built-in *ks.test* function in *R*). The results of this analysis are shown in Figs. 6, 7, 8 and Table 1.

In the scenario without extinction (Fig. 6) the fitting process was able to recover the parameters since the generated trees do not assume extinction, the KS statistic found no significant difference in the log branch lengths produced by the true generating model and the fitted PH_{Dec} model (Table 1). In the scenarios that included extinction, the fitting process was not able to correctly recover the true generating model (Figs. 7, 8). The bias in estimating the speciation process becomes more apparent as we increase the extinction rate (Fig. 8).

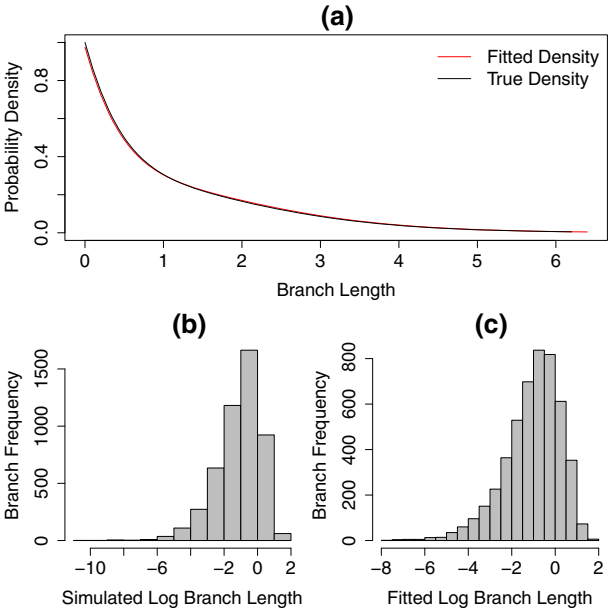

Fig. 6 True density function (black line in **a**) and the fitted density function (red line in **a**) from PH_{Dec} distribution. The values of x-axis on panel **a** show the branch lengths from both fitted and simulated trees. The histograms of the simulated and fitted branch length distributions, shown in log scale, are displayed in (**b**), (**c**), respectively. Data were simulated with no extinction (Color figure online)

4 Empirical Data

In this section, we apply the techniques developed in Sect. 2.4 to two large empirical phylogenies (Zheng and Wiens 2016; Zanne et al. 2014). In order to view these phylogenies and to extract clades of interest, we used Dendroscope 3 software (Huson and Scornavacca 2012). For each dataset, we compared nine models. These included models where the speciation process followed a PH distribution: the general Coxian distribution (Definition 2) with 3, 4, 5, and 6 non-absorbing states, and the two examples PH_{Dec} and PH_{Inc} developed in Sect. 2.2, one model where the speciation process follows an exponential distribution, one where it follows a Weibull distribution, and one where we fit to the constant rate birth–death model (crBD) using the likelihood formula of observing a tree conditioned on survival in Eq. 20 in Nee et al. (1994b) or using the likelihood in Eq. 27. We note that our likelihood formula as in Eq. 14 does not consider permutation on the tips of tree, so it differs from the likelihood from the crBD model by $(N-1)!$ where N denotes the number of tips on tree.

Our general approach for model comparison was to use the Akaike Information Criterion (AIC) (Akaike 1998) which is essentially the log likelihood penalized according to the number of parameters used in the model. We followed the approach suggested in Anderson and Burnham (2004) which is that models with an AIC difference (ΔAIC) of less than two are essentially as good as the best model, and models with ΔAIC less than 6 should not be discounted.

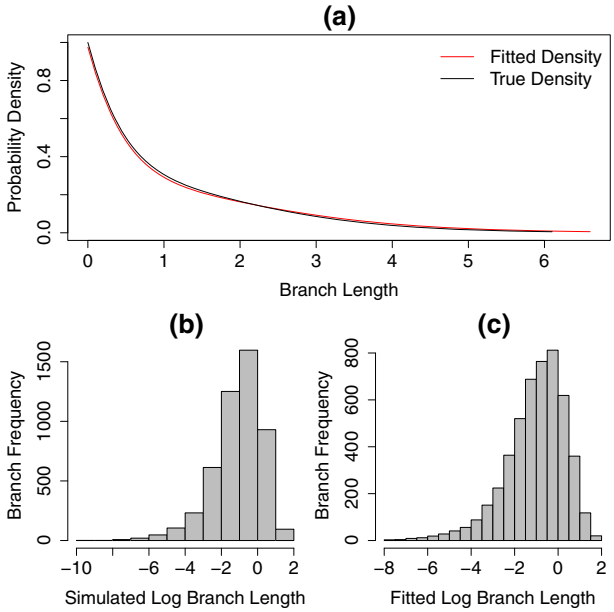

Fig. 7 True density function (black line in **a**) and the fitted density function (red line in **a**) from $\mathrm{PH}_{\mathrm{Dec}}$ distribution. The values of x-axis on panel **a** show the branch lengths from both fitted and simulated trees. The histograms of the simulated and fitted branch length distributions, shown in log scale, are displayed in (**b**), (**c**), respectively. Here, extinction events follow an exponential distribution with rate $\lambda = 0.1$ (Color figure online)

In addition to assessing relative goodness-of-fit via the AIC, and bearing in mind that all of our models are likely to be wrong given that they ignore extinction, we also assessed absolute goodness-of-fit using the KS statistic to compare fitted branch length densities to empirical branch length densities.

Lastly we show the hazard rate function for speciation from the best-fitting model for each clade. We were interested to see how different these would be to the constant hazard rate assumed by most macroevolutionary models or the monotonically decreasing hazard rate given by a Weibull distribution.

4.1 Squamate Phylogeny

We fit the models under consideration to the branch lengths from the squamate phylogeny in Zheng and Wiens (2016). We also examined three major clades of the tree separately, namely the *gekkota* clade (1318 branches), the *iguania* clade (1936 branches), and the *anguimorpha* clade (200 branches), to see if there are any notable differences.

The model comparison results are summarized in Table 2. The general Coxian model is strongly preferred for the overall tree and for all the clades being studied. In particular, the general Coxian model with three non-absorbing states fits best, but the model with four non-absorbing states is essentially indistinguishable. Additionally,

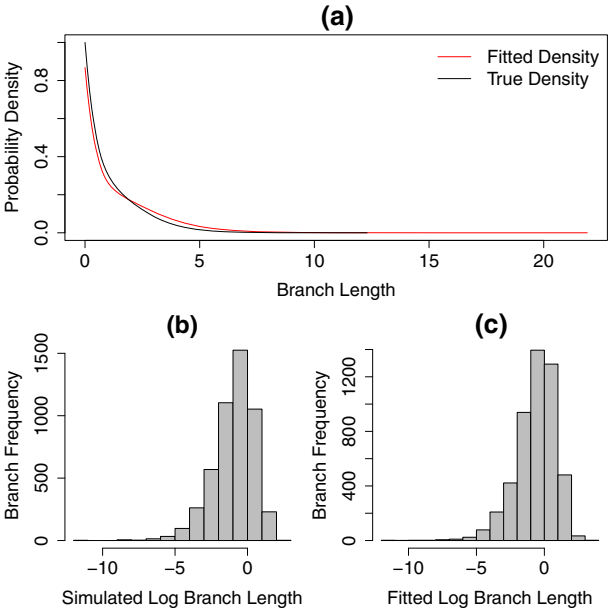

Fig. 8 True density function (black line in **a**) and the fitted density function (red line in **a**) from PH_{Dec} distribution. The values of x-axis on panel **a** show the branch lengths from both fitted and simulated trees. The histograms of the simulated and fitted branch length distributions, shown in log scale, are displayed in (**b**), (**c**), respectively. Here, extinction events follow an exponential distribution with rate $\lambda = 0.4$ (Color figure online)

fitting to the PH_{Inc} example model is significantly worse than other distributions. Moreover, fitting to the crBD model returns zero extinction rate for all the cases and returns the same parameter values for speciation process, comparable to the model that follows exponential speciation rate without extinction.

The absolute goodness-of-fit of different models is assessed in Fig. 9. Visually both general Coxian PH distribution with three and four non-absorbing states give fairly similar densities. These two appear to fit better compared to the other distributions (in agreement with the AIC results in Table 2). Both of these distributions seem to capture the tail behavior fairly well, but do a poorer job of matching the density for shorter branch lengths. The lack of fit to the reconstructed squamate tree and to most clades is supported by the KS tests which show a significant difference between the empirical branch lengths and branch lengths of 10 simulated trees from each best-fitting distribution (Table 3). We use the phytools package (Revell 2012) to simulate trees under the crBD model. Given that earlier results (Hagen et al. (2015) and Fig. 5d) show that the extinction process affects relative branch lengths, we hypothesize that this result could be due to ignoring extinction events in the models. Interestingly, all the distributions, except for PH_{Inc}, show a good fit between the empirical branch lengths of the *anguimorpha* clade and branch lengths of 10 simulated trees from each of these best-fitting distributions (Table 3). We note that this result could be due to the clade having a relatively small number of extant tips (101 tips); therefore, there is a lack of

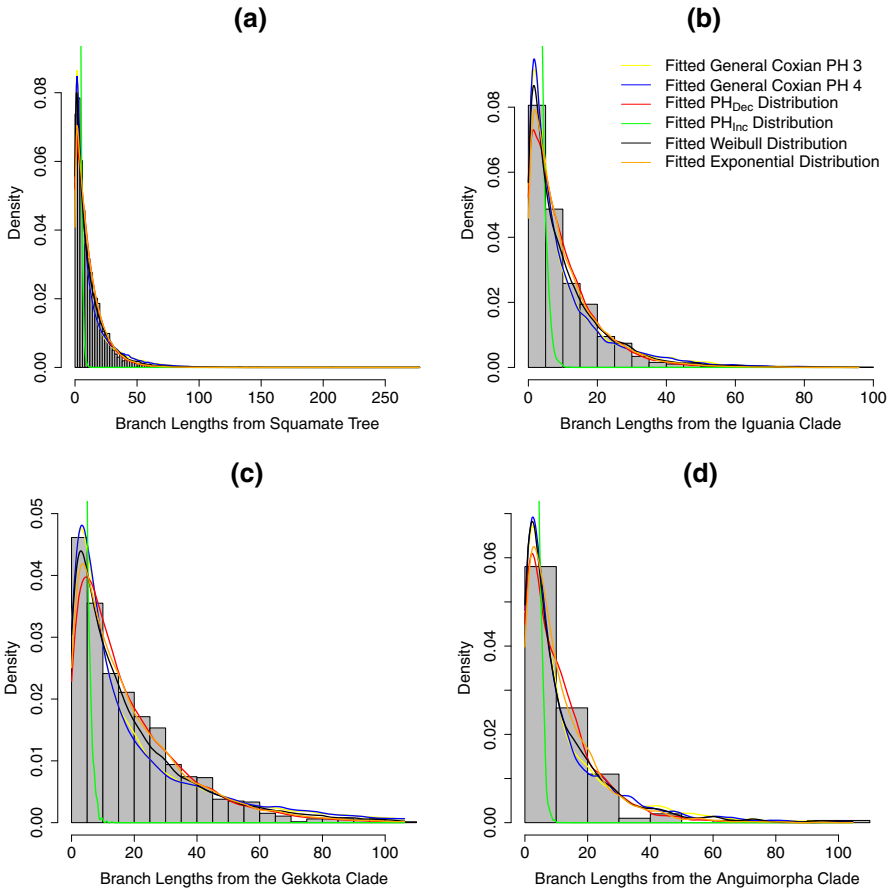

Fig. 9 Histograms of empirical branch length density from the whole squamate tree (**a**), the *iguania* clade (**b**), the *gekkota* clade (**c**), and the *anguimorpha* clade (**d**) with the fitted branch length densities from the six distributions mentioned above. The yellow and blue lines are the fitted densities using the general Coxian PH distribution defined in Definition 2 with 3 and 4 non-absorbing states, respectively, the red line is the fitted density using the Coxian PH_{Dec} example, the green line is the fitted density using the Coxian PH_{Inc} example, and black and orange lines are the fitted density using Weibull and exponential distributions, respectively. The fitted densities for the general Coxian PH distribution with 5 and 6 non-absorbing states are not included because in most cases the distribution with 3 and 4 non-absorbing states fit better, while the fitted density from the crBD model is not included because it is identical to the fitted density using exponential distribution (see Table 2) (Color figure online)

power to distinguish between models. Alternatively, extinction may occur at a lower rate in this clade compared to the other clades.

The hazard rate functions for speciation from the best-fitting general Coxian PH distribution with four non-absorbing states for each the overall squamate phylogeny and the three major clades are shown in Fig. 10. For the overall tree and for each clade, the instantaneous rate of speciation seems to show a slight decline (almost constant) as species get older.

Table 2 Model selection is based on the likelihood of observing the branch lengths given the specified model for times to speciation and no extinction (as per Sect. 2.4)

Model	# branches	# parameters	LogL	AIC	ΔAIC
(a)					
General Coxian PH 3	8336	5	−16727.93	33465.86	0
General Coxian PH 4		7	−16727.02	33468.04	2.18
General Coxian PH 5		9	−16726.87	33471.74	5.88
General Coxian PH 6		11	−16727.22	33476.44	10.58
PH_{Dec}		3	−17793.08	35592.15	2126.29
PH_{Inc}		3	−68410.67	136827.34	103361.50
Exponential distribution		1	−17322.14	34646.29	1180.43
Weibull distribution		2	−17024.64	34053.27	587.41
Constant rate birth–death		2	−17322.14	34648.29	1182.43
(b)					
General Coxian PH 3	1318	5	−2958.58	5927.16	0
General Coxian PH 4		7	−2958.08	5930.15	2.99
General Coxian PH 5		9	−2958.48	5934.96	7.80
General Coxian PH 6		11	−2958.51	5939.02	11.86
PH_{Dec}		3	−3092.96	6191.91	264.75
PH_{Inc}		3	−18783.21	37572.42	31645.25
Exponential distribution		1	−3048.34	6098.69	171.53
Weibull distribution		2	−3006.61	6017.21	90.05
Constant rate birth–death		2	−3048.34	6100.69	173.53
(c)					
General Coxian PH 3	1936	5	−3775.66	7561.33	0
General Coxian PH 4		7	−3774.20	7562.39	1.07
General Coxian PH 5		9	−3773.78	7565.55	4.23
General Coxian PH 6		11	−3773.75	7569.49	8.17
PH_{Dec}		3	−3963.97	7933.93	372.61
PH_{Inc}		3	−12852.02	25710.05	18148.72
Exponential distribution		1	−3860.30	7722.61	161.28
Weibull distribution		2	−3827.23	7658.46	97.14
Constant rate birth–death		2	−3860.30	7724.61	163.28
(d)					
General Coxian PH 3	200	5	−398.14	806.28	0
General Coxian PH 4		7	−398.11	810.22	3.94
General Coxian PH 5		9	−398.12	814.25	7.96
General Coxian PH 6		11	−398.08	818.17	11.88
PH_{Dec}		3	−417.02	840.05	33.76
PH_{Inc}		3	−1607.49	3220.98	2414.69

Table 2 continued

Model	# branches	# parameters	LogL	AIC	ΔAIC
Exponential distribution		1	−410.18	822.37	16.08
Weibull distribution		2	−402.24	808.47	2.18
Constant rate birth–death		2	−410.18	824.37	18.08

The constant rate birth–death (crBD) model is the only model that includes extinction in this comparison. The numbers (3, 4, 5, 6) in the row labels for the general Coxian PH indicate the number of non-absorbing states. For the crBD model from Nee et al. (1994b), we adjust the log likelihood by subtracting $\log((\ell-1)!)$ where ℓ is the number of tips on tree. We select the model that has the lowest AIC value as the base model and compute $\Delta\text{AIC} = \text{AIC}_{\text{other model}} - \text{AIC}_{\text{best model}}$. We use branch lengths from (a) the whole reconstructed squamate tree; and from different clades from the tree, namely (b) the *gekkota* clade, (c) the *iguania* clade, and (d) the *anguimorpha* clade

Table 3 KS tests for hypothesis testing that empirical branch length data of the reconstructed squamate tree and its following clades come from these fitted distributions

Distribution	Squamate	Iguania	Gekkota	Anguimorpha
	KS statistic	KS statistic	KS statistic	KS statistic
Constant rate birth–death	0.028	0.038	0.037	0.071
General Coxian PH 3	0.063	0.059	0.079	0.085
General Coxian PH 4	0.059	0.063	0.075	0.091
General Coxian PH 5	0.059	0.062	0.076	0.085
General Coxian PH 6	0.057	0.059	0.072	0.075
PH_{Dec}	0.062	0.059	0.061	0.069
PH_{Inc}	0.576	0.550	0.716	0.572
Weibull	0.069	0.064	0.071	0.081
Exponential	0.030	0.033	0.041	0.046

For the reconstructed squamate tree, the *iguania*, and the *gekkota* clades, the resulting p values from these KS statistics are all significant ($p < 0.05$), indicating that branch lengths drawn from these fitted distributions are significantly different than the empirical branch lengths. However, in the case of the *anguimorpha* clade, the p value are not significant ($p > 0.05$) in most distributions, except for PH_{Inc}. This indicates that branch lengths drawn from these fitted distributions are not statistically different compared to the empirical branch lengths

4.2 Angiosperm Phylogeny

To see how each model performs on an even larger tree, we also fit branch lengths from four different clades of the angiosperm phylogeny of (Zanne et al. 2014). The four different clades we use are: the *monocotyledoneae* clade (14,118 branches), the *magnoliidae* clade (2092 branches), the *superrosidae* clade (11,323 branches), and the *superasteridae* clade (20,016 branches).

The model comparison results are summarized in Table 4. The general Coxian model are very strongly preferred over all the other models for all of the individual clades. Additionally, fitting to the model that follows PH_{Inc} example is significantly worse than other distributions. Moreover, unlike the results in Table 2, the general Coxian model with four non-absorbing states fit best in this case. Interestingly, fitting

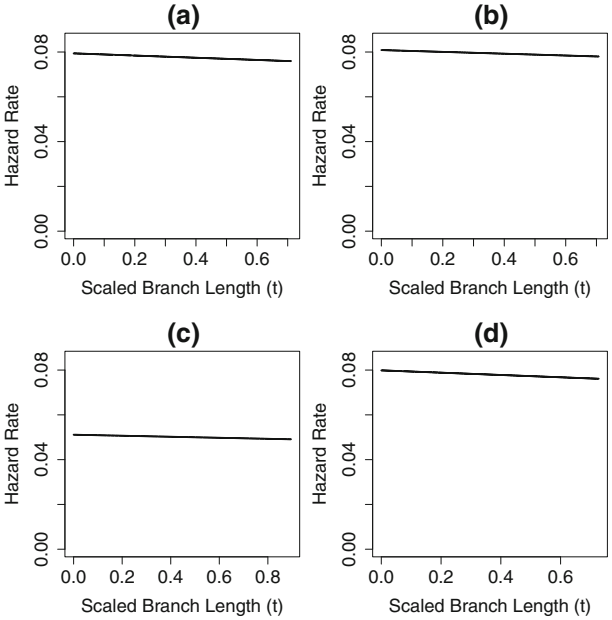

Fig. 10 Hazard rate functions for speciation show the change in the instantaneous probability of speciation as species age, as determined using the best-fitting general Coxian PH distribution with four non-absorbing states for the whole squamate tree (**a**), the *iguania* clade (**b**), the *gekkota* clade (**c**), and the *anguimorpha* clade (**d**). For each tree, branches are scaled by dividing each branch length leading to speciation event with height of the tree

to the crBD model to this set of empirical data returns non-zero extinction rate for all of the individual clades and it fits better compared to the model following an exponential speciation rate without extinction. The absolute goodness-of-fit of different models is assessed in Fig. 11. Visually, both general Coxian PH distributions with three and four non-absorbing states give fairly similar densities. These two appear to fit better compared to the other distributions (in agreement with the AIC results in Table 4). Both of these distributions seem to capture the tail behavior fairly well, but do a poorer job of matching the density for shorter branch lengths. The lack of fit is supported by the KS tests which show a significant difference between the empirical branch lengths and branch lengths of 10 simulated trees from each best-fitting distribution (Table 5). Again, we hypothesize that this result could be due to ignoring extinction events in the model. Here, as with the squamate data, we observe that the density of the fitted distribution of PH_{Inc}, which imposes increasing speciation rates as species age, does not follow the shape of the empirical histograms for any of the clades (Fig. 11).

The hazard rate functions for speciation from the best-fitting general Coxian PH distribution with four non-absorbing states for the four major clades of the angiosperm phylogeny are shown in Fig. 12. The instantaneous rate of speciation declines in each case and the rate of decline appears to be different in major clades of the angiosperm tree.

Table 4 Model selection is based on the likelihood of observing the branch lengths given the specified model for times to speciation and no extinction (as per Sect. 2.4)

Model	# branches	# parameters	LogL	AIC	ΔAIC
(a)					
General Coxian PH 3	14118	5	−18498.30	37006.59	112.74
General Coxian PH 4		7	−18439.92	36893.85	0
General Coxian PH 5		9	−18439.56	36897.13	3.28
General Coxian PH 6		11	−18439.46	36900.92	7.07
PH_{Dec}		3	−18788.38	37582.76	688.91
PH_{Inc}		3	−45591.17	91188.34	54294.49
Exponential distribution		1	−22149.68	44301.37	7407.52
Weibull distribution		2	−18633.08	37270.16	376.31
Constant rate birth–death		2	−19979.50	39962.99	3069.14
(b)					
General Coxian PH 3	2092	5	−3369.34	6748.68	40.07
General Coxian PH 4		7	−3347.30	6708.60	0
General Coxian PH 5		9	−3346.96	6711.93	3.33
General Coxian PH 6		11	−3346.95	6715.89	7.29
PH_{Dec}		3	−3476.51	6959.01	250.41
PH_{Inc}		3	−8926.00	17857.99	11149.39
Exponential distribution		1	−3633.44	7268.87	560.27
Weibull distribution		2	−3395.17	6794.34	85.73
Constant rate birth–death		2	−3493.25	6990.51	281.90
(c)					
General Coxian PH 3	20016	5	−29808.68	59627.35	117.45
General Coxian PH 4		7	−29747.95	59509.90	0
General Coxian PH 5		9	−29747.57	59513.15	3.25
General Coxian PH 6		11	−29747.56	59517.11	7.21
PH_{Dec}		3	−30533.85	61073.71	1563.81
PH_{Inc}		3	−59551.17	119108.33	59598.44
Exponential distribution		1	−33668.54	67339.07	7829.17
Weibull distribution		2	−30064.19	60132.39	622.49
Constant rate birth–death		2	−31765.43	63534.87	4024.97
(d)					
General Coxian PH 3	11323	5	−29977.30	59964.60	0
General Coxian PH 4		7	−29977.32	59968.64	4.04
General Coxian PH 5		9	−29977.33	59972.66	8.06
General Coxian PH 6		11	−29977.35	59976.71	12.11
PH_{Dec}		3	−30717.66	61441.32	1476.72
PH_{Inc}		3	−72613.88	145233.76	85269.16
Exponential distribution		1	−33136.25	66274.49	6309.89

Table 4 continued

Model	# branches	# parameters	LogL	AIC	ΔAIC
Weibull distribution		2	−30183.84	60371.68	407.08
Constant rate birth–death		2	−31791.90	63587.81	3623.21

The constant rate birth–death (crBD) model is the only model that includes extinction in this comparison. The numbers (3, 4, 5, 6) in the row labels for the general Coxian PH indicate the number of non-absorbing states. For the crBD model from Nee et al. (1994b), we adjust the log likelihood by subtracting $\log((\ell-1)!)$ where ℓ is the number of tips on tree. We select the model that has the lowest AIC value as the base model and compute $\Delta\text{AIC} = \text{AIC}_{\text{othermodel}} - \text{AIC}_{\text{bestmodel}}$. We use branch lengths from different clades of the angiosperm phylogeny, namely (a) the *monocotyledoneae* clade, (b) *magnoliidae* clade, (c) *superasteridae* clade and (d) the *superrosidae* clade

5 Discussion and Conclusion

Our macroevolutionary model for phylogenetic trees where times to speciation or extinction events are drawn from a Coxian PH distribution can produce phylogenetic trees with a range of tree shapes. The model provides a good fit to empirical data compared to exponential and Weibull distributions. The idea of applying PH distributions is motivated by the following two properties. First, it is well known that PH distributions are dense in the field of all positive-valued distributions (Asmussen et al. 1996), and thus, they are very flexible when fitting to empirical distributions. In particular, it implies that waiting times to either speciation or extinction events that follow any positive real-value distributions, such as exponential and Weibull, are well approximated using PH distribution with some given structure. Second, evolution of species trees or a species tree can be modeled as a forward-in-time process which follows an acyclic PH distribution. It is also known in the literature that any acyclic PH distribution can be represented as a Coxian PH distribution (Cumani 1982; Asmussen et al. 1996). Using a Coxian distribution is particularly useful here because its structure allows for the process to reach the absorbing state from any of the non-absorbing states, as described in Definition 2. This implies, using a general Coxian PH distribution, we can create an example where either speciation or extinction rates decrease or increase over time, by only changing parameter values inside the rate matrix $\mathbf{Q}$, such as ones in PH_{Dec} and PH_{inc}. However, we recommend using the general Coxian PH distribution when used to fit to empirical data.

We have demonstrated that trees generated under our model can have a range of different levels of tree balance as measured by the β statistic (Fig. 5). Thus, it is possible to fit parameters of our model to empirical tree shapes. The ability to get tree shapes that vary from the uniform distribution on ranked tree shapes (URT) in our model is expected based on the work of Lambert and Stadler (2013). A model with Coxian PH distributed times to speciation and exponentially distributed times to extinction is in class 4 of the scheme given in Lambert and Stadler (2013), in which the speciation process depends on a non-heritable trait (in this case species age).

In our simulations, we found that tree balance is mainly controlled by the speciation process and is largely invariant to the extinction process. In contrast to the behavior of β, the relative branch lengths, as measured by the γ statistic, are to a large extent

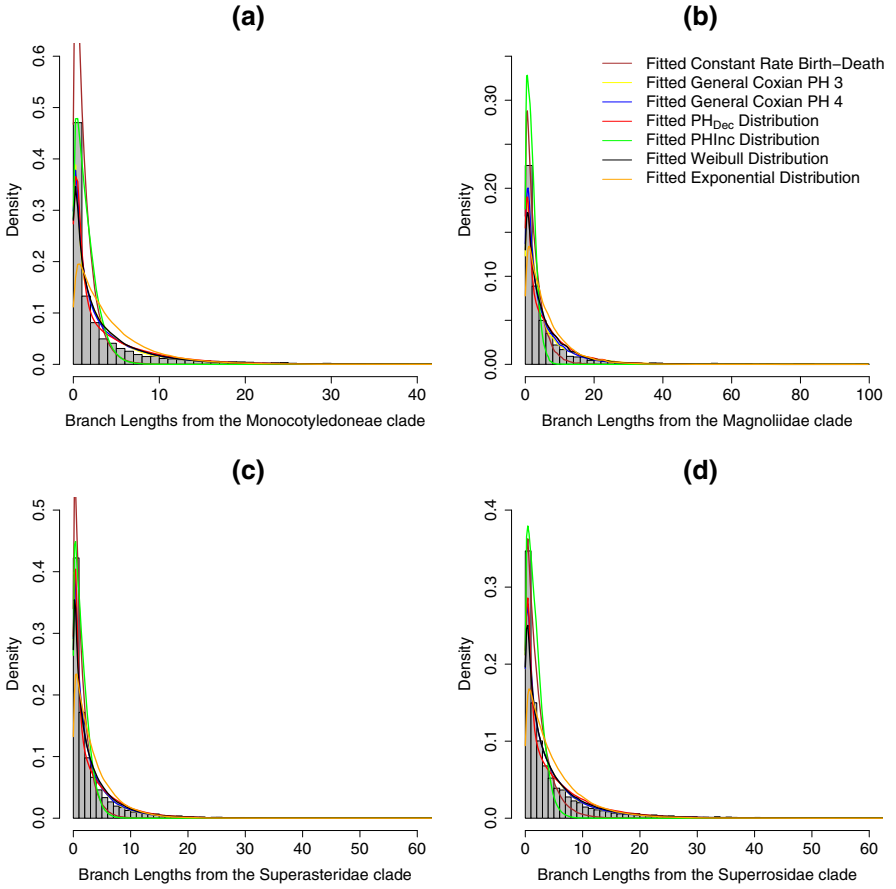

Fig. 11 Histograms of empirical branch length density from the *monocotyledoneae* clade (**a**), the *magnoliidae* (**b**), the *superasteridae* clade (**c**) and the *superrosidae* clade (**d**) with the fitted branch length densities from the six distributions mentioned earlier. The yellow and blue lines are the fitted densities using the general Coxian PH distribution defined in Definition 2 with 3 and 4 non-absorbing states, respectively, the red line is the fitted density using the Coxian PH_{Dec} example, the green line is the fitted density using the Coxian PH_{Inc} example, and black and orange lines are the fitted density using Weibull and exponential distributions, respectively. The fitted densities for the general Coxian PH distribution with 5 and 6 non-absorbing states are not included because in most cases having the distribution with less number of non-absorbing states (e.g., four non-absorbing states) fit better (see Table 4) (Color figure online)

controlled by the extinction process, but relatively invariant to the speciation process. Interestingly, unlike the β statistic where we found model parameters that gave values around -1, we did not find any model parameters that led to negative values of γ. We also found that using symmetric or asymmetric speciation modes did not have much effect on tree balance. These findings agree with the results in Hagen et al. (2015) in which speciation and extinction processes were modeled using Weibull distribution.

We proposed a method of computing the β statistic based on sets of trees. We have demonstrated that computing the β statistic based on individual trees can be upwardly biased, particularly for trees with smaller numbers of taxa. For trees generated by a

Table 5 KS tests for hypothesis testing that empirical branch length data of the following clades from the reconstructed angiosperm come from these fitted distributions

Distribution	Monocotyledoneae	Magnoliidae	Superasteridae	Superrosidae
	KS statistic	KS statistic	KS statistic	KS statistic
Constant rate birth–death	0.232	0.169	0.179	0.178
General Coxian PH 3	0.080	0.057	0.048	0.044
General Coxian PH 4	0.082	0.039	0.042	0.043
General Coxian PH 5	0.078	0.047	0.042	0.044
General Coxian PH 6	0.076	0.047	0.043	0.040
PH_{Dec}	0.073	0.072	0.045	0.055
PH_{Inc}	0.217	0.290	0.188	0.264
Weibull	0.100	0.080	0.066	0.061
Exponential	0.261	0.173	0.183	0.178

The resulting p values from these KS statistics are all significant ($p < 0.05$), indicating that branch lengths drawn from these fitted distributions are significantly different than the empirical branch lengths

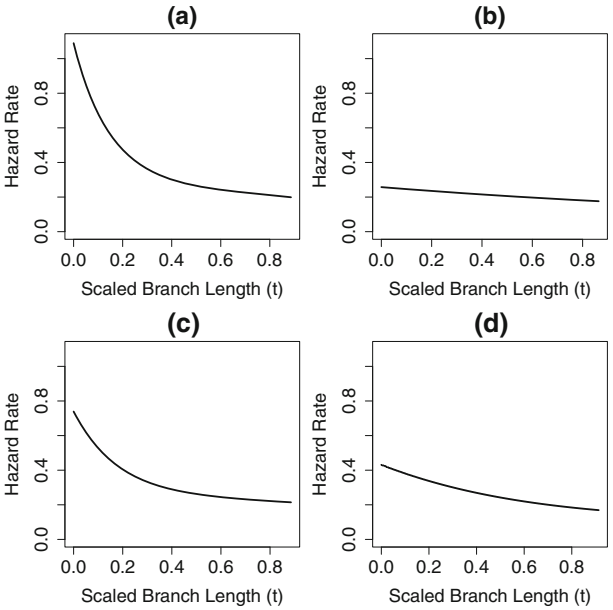

Fig. 12 Hazard rate functions for speciation show the change in the instantaneous probability of speciation as species age, as determined using the best-fitting general Coxian PH distribution with four non-absorbing states for the *monocotyledoneae* clade (**a**), the *magnoliidae* clade (**b**), the *superasteridae* clade (**c**), and *superrosidae* clade (**d**)

YH process, computing the β statistic based on sets of trees gives a more accurate result (Fig. 4). This approach of computing a β value for a set of trees is useful in the context of simulated tree data, but beyond simulation studies, there may be other contexts where it is useful to estimate β for a set of trees. For example, when studying

bio-geographic patterns researchers may have multiple species trees for the same set of geographic regions. It would also be possible to compute a single β value for a set of gene trees.

We derived a likelihood expression for the probability of observing any reconstructed tree (Eqs. 14–16) that has evolved with PH distributed times to speciation (and no extinction); we applied it to both simulated and empirical data by applying the maximum likelihood method. We note that fitting parameters based on branch lengths taken from trees that include extinction, produces some bias in estimation of the speciation process (Fig. 7). The bias becomes more apparent with increasing rates of extinction (Fig. 8). In future work, we aim to generalize Eq. 14 to include extinction. Such an extension can potentially be done in a similar manner as the derivation for the likelihood under a BDP process as described in Eq. 19. Once we derive a generalized likelihood function, we will compare its performance with likelihood functions that consider both speciation and extinction events, such as in Rabosky (2006).

In Sect. 2.4, we have also given a different approach for deriving the likelihood expression of observing a tree evolving under a constant rate birth–death process. This expression in Eq. 19 provides new physical interpretations in the context of the process driving the evolution of phylogenetic tree, and it also has a nice relationship with the formula in Nee et al. (1994b) as described in Eq. 27. In terms of fitting the model to empirical data, we note that the likelihood must be conditioned on the survival of the original two branches descending from the root of the tree as seen in Eq. 27. This agrees with what Stadler (2013a) stated in her paper.

Finally, we have fitted the parameters of our model to the empirical data consisting of branch lengths from various clades in the squamate and angiosperm reconstructed phylogenies (Zheng and Wiens 2016; Zanne et al. 2014). In both cases, we found that the extra flexibility permitted by the Coxian PH distribution was favored by the AIC over the simpler Weibull and Exponential models. Interestingly, in both cases, the model using the Coxian PH distribution without extinction process still fits better than the constant rate birth–death model from Nee et al. (1994b) that includes extinction. Moreover, in one example, fitting using the Coxian PH distribution with three non-absorbing states is preferable, but fitting using the distribution with four non-absorbing states is mostly preferred. Meanwhile, fitting to the same distribution with more than four non-absorbing states was always less favorable in the examples we looked at while also adding more computational time.

In the squamate phylogeny (Zheng and Wiens 2016), all the clades we examined (*iguania*, *gekkota*, *anguimorpha*) showed rates of speciation that declined slightly as species got older (Fig. 10). The whole squamate phylogeny also showed slight declining rates of speciation (almost constant rate). On the other hand, two clades (*monocotyledoneae*, *superasteridae*) from angiosperm phylogeny (Zanne et al. 2014) considered in this study showed apparent declining rates of speciation as species got older (Fig. 12a, c), while the other clades in the phylogeny (*magnoliidae*, *superrosidae*) only showed rates of speciation that decreased slightly (Fig. 12b, d). We caution against reading too much into these results as the model does not include extinction or account for incomplete sampling.

In summary, we have demonstrated that our macroevolutionary model with Coxian PH distribution, provides a better fit to empirical phylogenies, when compared to

models with other distributions, including exponential and Weibull (Tables 2, 4). We conclude that it is necessary to use distributions with sufficient complexity, such as Coxian PH distributions, to provide a better fit to empirical phylogenies.

Acknowledgements We would like to thank the Australian Research Council for funding this research through Discovery Project DP180100352. We also would like to thank Oskar Hagen from ETH Zürich for the insight in solving an issue with generating trees using the *TreeSimGM* package.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions

Data Availability The datasets and all the relevant code, including functions for fitting empirical data to a phase-type model and for computing treeset β values, are available in the DRYAD repository, https://doi.org/10.5061/dryad.w9ghx3fpk

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

6 Appendix

6.1 Equivalence of Formulas for $q_n(i, \beta)$

There are two different formulas for computing the probability of observing i tips on the left given n extant tips on a tree, $q_n(i, \beta)$. The first expression includes a product of gamma functions with a normalizing constant, $a_n(\beta)$, as seen in Eq. 4 from Aldous (1996), while the second expression includes a product of beta functions with a normalizing constant, $\hat{a_n}(\beta)$, as seen in the *maxlik.betasplit* command from *apTreeShape* package (https://github.com/bcm-uga/apTreeshape/blob/master/R/maxlik.betasplit.R). Here, we show that both expressions are equivalent by showing that both normalizing constants are related.

Recall from Aldous (1996), we have

$$q_n(i, \beta) = \frac{1}{a_n(\beta)} \frac{\Gamma(\beta + i + 1)\Gamma(\beta + n - i + 1)}{\Gamma(i + 1)\Gamma(n - i + 1)}, 1 \leq i \leq n - 1, \tag{30}$$

where $a_n(\beta)$ is a normalizing constant and $\Gamma(x)$ is the gamma function.

Recall from the *maxlik.betasplit* command, we have

$$\hat{q}_n(i, \beta) = \frac{1}{\hat{a_n}(\beta)} \frac{B(\beta + i + 1, \beta + n - i + 1)}{B(i + 1, n - i + 1)}, 1 \leq i \leq n - 1, \tag{31}$$

where $\hat{a_n}(\beta)$ is a normalizing constant and $B(x, y)$ is beta function.

Proof Using the relation between gamma and beta functions where $B(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$, we can write Eq. 31 as,

$$\hat{q}_n(i, \beta) = \frac{1}{\hat{a_n}(\beta)} \frac{\frac{\Gamma(\beta+i+1)\Gamma(\beta+n-i+1)}{\Gamma(2\beta+n+2)}}{\frac{\Gamma(i+1)\Gamma(n-i+1)}{\Gamma(n+2)}} \tag{32}$$

$$= \frac{\Gamma(n+2)}{\hat{a_n}(\beta)\Gamma(2\beta+n+2)} \frac{\Gamma(\beta+i+1)\Gamma(\beta+n-i+1)}{\Gamma(i+1)\Gamma(n-i+1)} \tag{33}$$

$$= \frac{\Gamma(n+2)}{\hat{a_n}(\beta)\Gamma(2\beta+n+2)} a_n(\beta) q_n(i, \beta). \tag{34}$$

Hence, $\hat{q}_n(i, \beta) = q_n(i, \beta)$ if and only if $\frac{1}{\hat{a_n}(\beta)} = \frac{\Gamma(n+2)}{a_n(\beta)\Gamma(2\beta+n+2)}$. That is, $\hat{a}_n(\beta) = \frac{a_n(\beta)\Gamma(2\beta+n+2)}{\Gamma(n+2)}$. □

6.2 Equivalent Formula of $q_n(i, \beta)$ for Large n and i

Here, we show the work to approximate Eqs. 30 and 31 for large n and i, where n is the number of extant tips on the tree and i is the number of left tips on the tree. We use this approximation due to computational limitation of evaluating gamma function for large number. The formula also appears in the *maxlik.betasplit* from the *TreeSimGM* package (Hagen and Stadler 2018).

Lemma 1 *Given large n and i, Eqs.* 30 *and* 31 *can be approximated using the following formula,*

$$\hat{q}_n(i, \beta) = \frac{1}{\hat{a}_n(\beta)} \left(\frac{i}{n}\right)^{\beta} \left(1 - \frac{i}{n}\right)^{\beta}, \tag{35}$$

where $\hat{a}_n(\beta)$ *is the normalizing constant for* $\hat{q}_n(i, \beta)$.

Proof Recall the Stirling's approximation for gamma function is given by

$$\Gamma(z) \approx \sqrt{\frac{2\pi}{z}} \left(\frac{z}{e}\right)^{z}. \tag{36}$$

Then, we claim that

Lemma 2

$$\frac{\Gamma(x+\beta+1)}{\Gamma(x+\alpha+1)} \approx x^{\beta-\alpha} \text{ for large } x. \tag{37}$$

Proof By Stirling's approximation with $z = x + \beta$ and $z = x + \alpha$, we have

$$\frac{\Gamma(x+\beta+1)}{\Gamma(x+\alpha+1)} = \frac{(x+\beta)\Gamma(x+\beta)}{(x+\alpha)\Gamma(x+\alpha)}, \text{since } x+\beta \text{and } x+\alpha \in Z \tag{38}$$

$$\approx \frac{(x+\beta)\sqrt{\frac{2\pi}{x+\beta}}\left(\frac{x+\beta}{e}\right)^{x+\beta}}{(x+\alpha)\sqrt{\frac{2\pi}{x+\alpha}}\left(\frac{x+\alpha}{e}\right)^{x+\alpha}} \tag{39}$$

$$= \frac{\sqrt{2\pi(x+\beta)}\left(\frac{x+\beta}{e}\right)^{x+\beta}}{\sqrt{2\pi(x+\alpha)}\left(\frac{x+\alpha}{e}\right)^{x+\alpha}} \tag{40}$$

$$= \left(\frac{x+\beta}{x+\alpha}\right)^{\frac{1}{2}} \frac{(x+\beta)^{x+\beta}}{(x+\alpha)^{x+\alpha}} \frac{1}{e^{\beta-\alpha}} \tag{41}$$

$$= \frac{(x+\beta)^{x+\beta+1/2}}{(x+\alpha)^{x+\alpha+1/2}} \frac{1}{e^{\beta-\alpha}} \tag{42}$$

$$= \frac{(x+\alpha+\beta-\alpha)^{x+\alpha+1/2}}{(x+\alpha)^{x+\alpha+1/2}} \frac{(x+\beta)^{\beta-\alpha}}{e^{\beta-\alpha}} \tag{43}$$

$$= \left(1+\frac{\beta-\alpha}{x+\alpha}\right)^{x+\alpha+1/2} \left(\frac{x+\beta}{x}\right)^{\beta-\alpha} \frac{x^{\beta-\alpha}}{e^{\beta-\alpha}} \tag{44}$$

$$= \left(1+\frac{\beta-\alpha}{x+\alpha}\right)^{x+\alpha+1/2} \left(1+\frac{\beta}{x}\right)^{\beta-\alpha} \left(\frac{x}{e}\right)^{\beta-\alpha}. \tag{45}$$

We observe here that $\left(1+\frac{\beta-\alpha}{x+\alpha}\right)^{x+\alpha+1/2} \to e^{\beta-\alpha}$ as $x \to \infty$ and $\left(1+\frac{\beta}{x}\right)^{\beta-\alpha} \to 1$ as $x \to \infty$. Therefore,

$$\frac{\Gamma(x+\beta+1)}{\Gamma(x+\alpha+1)} \approx e^{\beta-\alpha}\left(\frac{x}{e}\right)^{\beta-\alpha} \tag{46}$$

$$= x^{\beta-\alpha}. \tag{47}$$

□

Recall that $q_n(i,\beta) = \frac{1}{a_n(\beta)} \frac{\Gamma(\beta+i+1)\Gamma(\beta+n-i+1)}{\Gamma(i+1)\Gamma(n-i+1)}$. Then, we apply Eq. 47 for large n and i,

$$q_n(i,\beta) = \frac{1}{a_n(\beta)} \frac{\Gamma(\beta+i+1)}{\Gamma(i+1)} \frac{\Gamma(\beta+n-i+1)}{\Gamma(n-i+1)} \tag{48}$$

$$\approx \frac{1}{a_n(\beta)} i^{\beta}(n-i)^{\beta} \tag{49}$$

$$= \frac{n^{2\beta}}{a_n(\beta)} \left(\frac{i}{n}\right)^{\beta} \left(1-\frac{i}{n}\right)^{\beta}. \tag{50}$$

That is, $q_n(i,\beta) = \hat{q}_n(i,\beta)$ if and only if $\hat{a}_n(\beta) = \frac{a_n(\beta)}{n^{2\beta}}$. □

To verify the result, we conduct a simulation for $n = 500$ and $\beta = -1$ (see Fig. 13).

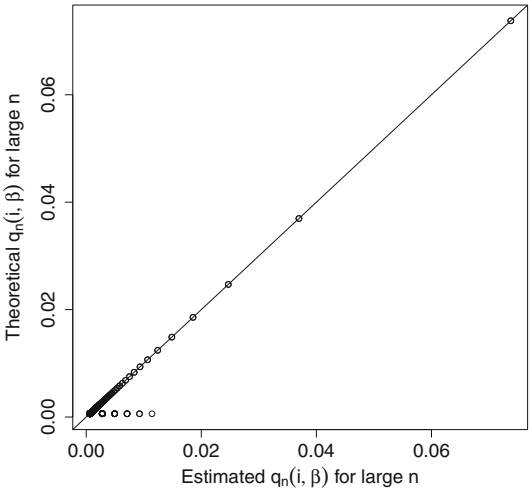

Fig. 13 Comparison of the probability $q_n(i, \beta)$ defined in Eqs. 35 and 30 for $n = 500$ and $\beta = -1$. The x-axis represents the probability defined in Eq. 35 while the y-axis represents the probability defined in Eq. 30

6.3 Expression of First and Second Moments from Coxian PH Distribution

In this section, we derive the expressions for first and second moments from a Coxian PH distribution, then we also derive those expressions for the two examples of a Coxian PH distribution used on this paper. The structure of the rate matrix $\mathbf{Q}$ follows canonical form 3 described in Okamura and Dohi (2016).

Consider a Coxian PH distribution with four non-absorbing states defined by its rate matrix given as follows

$$\mathbf{Q} = \begin{bmatrix} -\lambda_1 & p_1\lambda_1 & & \\ & -\lambda_2 & p_2\lambda_2 & \\ & & -\lambda_3 & p_3\lambda_3 \\ & & & -\lambda_4 \end{bmatrix}, \tag{51}$$

where $0 < p_1, p_2, p_3 \leq 1$. Furthermore, we have the condition that $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \lambda_4$ based on the result in Cumani (1982) and Dehon and Latouche (1982) for acyclic PH distributions.

In order to derive the expression of first and second moments of a Coxian PH distribution, we compute the inverse matrix in Eq. 51 using the identity matrix of the same size and performing elementary row operations to derive $\left(\mathbf{I}|(\mathbf{Q})^{-\mathbf{1}}\right)$ from $(\mathbf{Q}|\mathbf{I})$.

$$\left(\begin{array}{cccc|cccc} -\lambda_1 & p_1\lambda_1 & 0 & 0 & 1 & 0 & 0 & 0 \\ -0 & -\lambda_2 & p_2\lambda_2 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & -\lambda_3 & p_3\lambda_3 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -\lambda_4 & 0 & 0 & 0 & 1 \end{array}\right) \begin{array}{l} \\ \xrightarrow{-\frac{1}{\lambda_1}r_1} \\ \xrightarrow{-\frac{1}{\lambda_2}r_2} \\ \\ \end{array} \left(\begin{array}{cccc|cccc} 1 & -p_1 & 0 & 0 & -\frac{1}{\lambda_1} & 0 & 0 & 0 \\ 0 & 1 & -p_2 & 0 & 0 & -\frac{1}{\lambda_2} & 0 & 0 \\ 0 & 0 & -\lambda_3 & p_3\lambda_3 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -\lambda_4 & 0 & 0 & 0 & 1 \end{array}\right)$$

$$
\xrightarrow[\substack{-\frac{1}{\lambda_3}r_3}]{r_1+p_1r_2}
\left(\begin{array}{cccc|cccc}
1 & 0 & -p_1p_2 & 0 & -\frac{1}{\lambda_1} & -\frac{p_1}{\lambda_2} & 0 & 0 \\
0 & 1 & -p_2 & 0 & 0 & -\frac{1}{\lambda_2} & 0 & 0 \\
0 & 0 & 1 & -p_3 & 0 & 0 & -\frac{1}{\lambda_3} & 0 \\
0 & 0 & 0 & -\lambda_4 & 0 & 0 & 0 & 1
\end{array}\right)
$$

$$
\xrightarrow[r_2+p_2r_3]{r_1+p_1p_2r_3}
\left(\begin{array}{cccc|cccc}
1 & 0 & 0 & -p_1p_2p_3 & -\frac{1}{\lambda_1} & -\frac{p_1}{\lambda_2} & -\frac{p_1p_2}{\lambda_3} & 0 \\
0 & 1 & 0 & -p_2p_3 & 0 & -\frac{1}{\lambda_2} & -\frac{p_2}{\lambda_3} & 0 \\
0 & 0 & 1 & -p_3 & 0 & 0 & -\frac{1}{\lambda_3} & 0 \\
0 & 0 & 0 & -\lambda_4 & 0 & 0 & 0 & 1
\end{array}\right)
$$

$$
\xrightarrow{-\frac{1}{\lambda_4}r_4}
\left(\begin{array}{cccc|cccc}
1 & 0 & 0 & -p_1p_2p_3 & -\frac{1}{\lambda_1} & -\frac{p_1}{\lambda_2} & -\frac{p_1p_2}{\lambda_3} & 0 \\
0 & 1 & 0 & -p_2p_3 & 0 & -\frac{1}{\lambda_2} & -\frac{p_2}{\lambda_3} & 0 \\
0 & 0 & 1 & -p_3 & 0 & 0 & -\frac{1}{\lambda_3} & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & -\frac{1}{\lambda_4}
\end{array}\right)
$$

$$
\xrightarrow[\substack{r_3+p_3r_4}]{\substack{r_1+p_1p_2p_3r_4 \\ r_2+p_2p_3r_4}}
\left(\begin{array}{cccc|cccc}
1 & 0 & 0 & 0 & -\frac{1}{\lambda_1} & -\frac{p_1}{\lambda_2} & -\frac{p_1p_2}{\lambda_3} & -\frac{p_1p_2p_3}{\lambda_4} \\
0 & 1 & 0 & 0 & 0 & -\frac{1}{\lambda_2} & -\frac{p_2}{\lambda_3} & -\frac{p_2p_3}{\lambda_4} \\
0 & 0 & 1 & 0 & 0 & 0 & -\frac{1}{\lambda_3} & -\frac{p_3}{\lambda_4} \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & -\frac{1}{\lambda_4}
\end{array}\right).
$$

Therefore,

$$
\mathbf{Q}^{-1} = \begin{bmatrix}
-\frac{1}{\lambda_1} & -\frac{p_1}{\lambda_2} & -\frac{p_1p_2}{\lambda_3} & -\frac{p_1p_2p_3}{\lambda_4} \\
0 & -\frac{1}{\lambda_2} & -\frac{p_2}{\lambda_3} & -\frac{p_2p_3}{\lambda_4} \\
0 & 0 & -\frac{1}{\lambda_3} & -\frac{p_3}{\lambda_4} \\
0 & 0 & 0 & -\frac{1}{\lambda_4}
\end{bmatrix}
$$

and $\mathbf{Q}\mathbf{Q}^{-1} = \mathbf{I}$ where $\mathbf{I}$ is the identity matrix.

Furthermore,

$$
\begin{aligned}
\mathbf{Q}^{-2} &= \left(\mathbf{Q}^{-1}\right)^2 \\
&= \begin{bmatrix}
\frac{1}{\lambda_1^2} & \frac{p_1}{\lambda_2}\left(\frac{1}{\lambda_1}+\frac{1}{\lambda_2}\right) & \frac{p_1p_2}{\lambda_3}\left(\frac{1}{\lambda_1}+\frac{1}{\lambda_2}+\frac{1}{\lambda_3}\right) & \frac{p_1p_2p_3}{\lambda_4}\left(\frac{1}{\lambda_1}+\frac{1}{\lambda_2}+\frac{1}{\lambda_3}+\frac{1}{\lambda_4}\right) \\
0 & \frac{1}{\lambda_2^2} & \frac{p_2}{\lambda_3}\left(\frac{1}{\lambda_2}+\frac{1}{\lambda_3}\right) & \frac{p_2p_3}{\lambda_4}\left(\frac{1}{\lambda_2}+\frac{1}{\lambda_3}+\frac{1}{\lambda_4}\right) \\
0 & 0 & \frac{1}{\lambda_3^2} & \frac{p_3}{\lambda_4}\left(\frac{1}{\lambda_3}+\frac{1}{\lambda_4}\right) \\
0 & 0 & 0 & \frac{1}{\lambda_4^2}
\end{bmatrix}.
\end{aligned}
$$

Hence, the expressions for first and second moments from a Coxian PH distribution with the initial probability distribution $\boldsymbol{\alpha} = [1, 0, 0, 0]$ and the rate matrix given by Eq. 51 are as follows

$$
\mathbb{E}_{\mathrm{PH}}(X) = -\boldsymbol{\alpha}\mathbf{Q}^{-1}\mathbf{1}
$$

$$
\begin{aligned}
&= -\begin{bmatrix}1 & 0 & 0 & 0\end{bmatrix}\begin{bmatrix}-\frac{1}{\lambda_1} & -\frac{p_1}{\lambda_2} & -\frac{p_1p_2}{\lambda_3} & -\frac{p_1p_2p_3}{\lambda_4}\\ 0 & -\frac{1}{\lambda_2} & -\frac{p_2}{\lambda_3} & -\frac{p_2p_3}{\lambda_4}\\ 0 & 0 & -\frac{1}{\lambda_3} & -\frac{p_3}{\lambda_4}\\ 0 & 0 & 0 & -\frac{1}{\lambda_4}\end{bmatrix}\begin{bmatrix}1\\1\\1\\1\end{bmatrix}\\
&= \frac{1}{\lambda_1} + \frac{p_1}{\lambda_2} + \frac{p_1p_2}{\lambda_3} + \frac{p_1p_2p_3}{\lambda_4},\\
\mathbb{E}_{\mathrm{PH}}(X) &= \frac{1}{\lambda_1} + p_1\left(\frac{1}{\lambda_2} + p_2\left(\frac{1}{\lambda_3} + \frac{p_3}{\lambda_4}\right)\right). \qquad (52)\\
\mathbb{E}_{\mathrm{PH}}\left(X^2\right) &= 2\boldsymbol{\alpha}\mathbf{Q}^{-2}\mathbf{1}\\
&= 2\begin{bmatrix}1 & 0 & 0 & 0\end{bmatrix}\\
&\quad\begin{bmatrix}\frac{1}{\lambda_1^2} & \frac{p_1}{\lambda_2}\left(\frac{1}{\lambda_1}+\frac{1}{\lambda_2}\right) & \frac{p_1p_2}{\lambda_3}\left(\frac{1}{\lambda_1}+\frac{1}{\lambda_2}+\frac{1}{\lambda_3}\right) & \frac{p_1p_2p_3}{\lambda_4}\left(\frac{1}{\lambda_1}+\frac{1}{\lambda_2}+\frac{1}{\lambda_3}+\frac{1}{\lambda_4}\right)\\ 0 & \frac{1}{\lambda_2^2} & \frac{p_2}{\lambda_3}\left(\frac{1}{\lambda_2}+\frac{1}{\lambda_3}\right) & \frac{p_2p_3}{\lambda_4}\left(\frac{1}{\lambda_2}+\frac{1}{\lambda_3}+\frac{1}{\lambda_4}\right)\\ 0 & 0 & \frac{1}{\lambda_3^2} & \frac{p_3}{\lambda_4}\left(\frac{1}{\lambda_3}+\frac{1}{\lambda_4}\right)\\ 0 & 0 & 0 & \frac{1}{\lambda_4^2}\end{bmatrix}\begin{bmatrix}1\\1\\1\\1\end{bmatrix},\\
\mathbb{E}_{\mathrm{PH}}\left(X^2\right) &= 2\left[\frac{1}{\lambda_1^2} + \frac{p_1}{\lambda_2}\left(\frac{1}{\lambda_1}+\frac{1}{\lambda_2}\right) + \frac{p_1p_2}{\lambda_3}\left(\frac{1}{\lambda_1}+\frac{1}{\lambda_2}+\frac{1}{\lambda_3}\right)\right.\\
&\quad\left. + \frac{p_1p_2p_3}{\lambda_4}\left(\frac{1}{\lambda_1}+\frac{1}{\lambda_2}+\frac{1}{\lambda_3}+\frac{1}{\lambda_4}\right)\right]. \qquad (53)
\end{aligned}
$$

Next, to get the expressions for first and second moment from $\mathrm{PH}_{\mathrm{Dec}}$, we use Eqs. 52 and 53 and the following substitutions,

$$
\begin{aligned}
&\lambda_1 = z, \lambda_2 = 1 + x, \lambda_3 = 1 + x^2, \lambda_4 = x^3,\\
&p_1 = 1 - y, p_2 = 1 - y^2, p_3 = 1 - y^3. \qquad (54)
\end{aligned}
$$

On the other hand, to derive the expressions for both moments from $\mathrm{PH}_{\mathrm{Inc}}$, we use the following substitutions to Eqs. 52 and 53,

$$
\begin{aligned}
&\lambda_1 = 1 + x^3, \lambda_2 = 1 + x^2, \lambda_3 = 1 + x, \lambda_4 = z\\
&p_1 = 1 - y^4, p_2 = 1 - y^3, p_3 = 1 - y^2 \qquad (55)
\end{aligned}
$$

6.4 Deriving and Solving the Differential Equation of $D_t^{(1)}(z)$

In this section, we show the derivation of the differential equation of the probability of observing a reconstructed external branch with length z on a tree with age t, $D_t^{(1)}(z)$, shown in Eq. 23, using physical interpretations. Then, we derive the solution to the differential equation as shown in Eq. 26.

We can write $D_t^{(1)}(z)$ by conditioning on the time of the first event on that external branch with elapsed time z on a tree with age t. That is, (1) the branch has not undergone any observable event yet at time z, which occurs with probability $e^{-(\lambda+\mu)z}$ or (2) the branch has a child at some time $u \leq z$, which occurs with probability $e^{-(\lambda+\mu)u}\lambda$, and so the two branches evolve independently of each other where the child branch

becomes extinct by time z with probability $E(z-u)$ and the initial branch survives at time z with probability $D_t^{(1)}(z-u)$ or vice versa. Thus,

$$D_t^{(1)}(z) = e^{-(\lambda+\mu)z} + \int_{u=0}^{z} e^{-(\lambda+\mu)u}\lambda\left(2D_t^{(1)}(z-u)E(z-u)\right)\mathrm{d}u.$$

Then, by taking derivative with respect to z in the above equation, we have,

$$\frac{\mathrm{d}D_t^{(1)}(z)}{\mathrm{d}z} = \frac{\mathrm{d}}{\mathrm{d}z}\left(e^{-(\lambda+\mu)z}\right) + \frac{\mathrm{d}}{\mathrm{d}z}\left(\int_{u=0}^{z} e^{-(\lambda+\mu)u}\lambda\left(2D_t^{(1)}(z-u)E(z-u)\right)\mathrm{d}u\right).$$

Next, by applying the Leibniz integral rule and noting that $E(z-z) = E(0) = 0$, we have,

$$\begin{aligned}\frac{\mathrm{d}D_t^{(1)}(z)}{\mathrm{d}z} &= -(\lambda+\mu)e^{-(\lambda+\mu)z} + \left(\int_{u=0}^{z}\frac{\partial}{\partial z}\left(e^{-(\lambda+\mu)u}\lambda\left(2D_t^{(1)}(z-u)E(z-u)\right)\right)\mathrm{d}u\right)\\ &= -(\lambda+\mu)e^{-(\lambda+\mu)z} + \int_{u=0}^{z} e^{-(\lambda+\mu)u}2\lambda\\ &\quad\left(\frac{\partial D_t^{(1)}(z-u)}{\partial z}E(z-u) + D_t^{(1)}(z-u)\frac{\partial E(z-u)}{\partial z}\right)\mathrm{d}u.\end{aligned}$$

Next, applying integration by parts and noting:

$$\frac{\partial}{\partial u}(D_t^{(1)}(z-u)E(z-u)) = -\left(\frac{\partial D_t^{(1)}(z-u)}{\partial z}E(z-u) + D_t^{(1)}(z-u)\frac{\partial E(z-u)}{\partial z}\right)$$

we get,

$$\begin{aligned}\frac{\mathrm{d}D_t^{(1)}(z)}{\mathrm{d}z} &= -(\lambda+\mu)e^{-(\lambda+\mu)z}\\ &\quad + \left(-2\lambda e^{-(\lambda+\mu)u}D_t^{(1)}(z-u)E(z-u)|_{u=0}^{z}\right.\\ &\quad \left. - \int_{u=0}^{z} 2\lambda(\lambda+\mu)e^{-(\lambda+\mu)u}D_t^{(1)}(z-u)E(z-u)\mathrm{d}u\right)\\ &= -(\lambda+\mu)e^{-(\lambda+\mu)z} + 2\lambda D_t^{(1)}(z)E(z) - (\lambda+\mu)\\ &\quad \int_{u=0}^{z} 2\lambda e^{-(\lambda+\mu)u}D_t^{(1)}(z-u)E(z-u)\mathrm{d}u\\ &= -(\lambda+\mu)\left(e^{-(\lambda+\mu)z} + \int_{u=0}^{z} e^{-(\lambda+\mu)u}\lambda\left(2D_t^{(1)}(z-u)E(z-u)\right)\mathrm{d}u\right)\\ &\quad +2\lambda D_t^{(1)}(z)E(z)\\ \frac{\mathrm{d}D_t^{(1)}(z)}{\mathrm{d}z} &= -(\lambda+\mu)D_t^{(1)}(z) + 2\lambda E(z)D_t^{(1)}(z), \quad \text{as in Eq. (23)}\end{aligned}$$

Lemma 3

$$D_t^{(1)}(z) = \left(\frac{(\lambda-\mu)e^{\mu z}}{\lambda-\mu e^{(\mu-\lambda)z}}\right)^2 e^{-(\lambda+\mu)z}$$

is the solution to the differential equation,

$$\frac{\mathrm{d}D_t^{(1)}(z)}{\mathrm{d}z} = -(\lambda+\mu)D_t^{(1)}(z) + 2\lambda E(z)D_t^{(1)}(z),$$

where $D_t^{(1)}(0) = 1$.

Proof Substitute $E(z)$ from Eq. (24) (see also Kendall 1948) to the above differential equation, we have,

$$\frac{\mathrm{d}D_t^{(1)}(z)}{\mathrm{d}z} = -(\lambda+\mu)D_t^{(1)}(z) + 2\lambda\left(\frac{\mu-\mu e^{(\mu-\lambda)z}}{\lambda-\mu e^{(\mu-\lambda)z}}\right)D_t^{(1)}(z). \tag{56}$$

Then,

$$\begin{aligned}
\frac{\mathrm{d}D_t^{(1)}(z)}{\mathrm{d}z} &= -(\lambda+\mu)D_t^{(1)}(z) + 2\lambda\left(\frac{\mu-\mu e^{(\mu-\lambda)z}}{\lambda-\mu e^{(\mu-\lambda)z}}\right)D_t^{(1)}(z)\\
\frac{\mathrm{d}D_t^{(1)}(z)}{\mathrm{d}z} &= -(\lambda+\mu)D_t^{(1)}(z) + g(z)D_t^{(1)}(z), \quad g(z) = 2\lambda\left(\frac{\mu-\mu e^{(\mu-\lambda)z}}{\lambda-\mu e^{(\mu-\lambda)z}}\right)\\
\frac{\mathrm{d}D_t^{(1)}(z)}{\mathrm{d}z} &= \left(g(z)-(\lambda+\mu)\right)D_t^{(1)}(z)\\
\int\frac{dD_t^{(1)}(z)}{D_t^{(1)}(z)} &= \int\left(g(z)-(\lambda+\mu)\right)\mathrm{d}z + C\\
\ln\left(D_t^{(1)}(z)\right) &= -(\lambda+\mu)z + \int g(z)\mathrm{d}z + C.
\end{aligned} \tag{57}$$

Next, we solve $\int g(z)\mathrm{d}z$,

$$\begin{aligned}
\int g(z)\mathrm{d}z &= 2\lambda\int\frac{\mu-\mu e^{(\mu-\lambda)z}}{\lambda-\mu e^{(\mu-\lambda)z}}\mathrm{d}z\\
&= 2\lambda\left(\underbrace{\int\frac{\mu\mathrm{d}z}{\lambda-\mu e^{(\mu-\lambda)z}}}_{I} - \underbrace{\int\frac{\mu e^{(\mu-\lambda)z}\mathrm{d}z}{\lambda-\mu e^{(\mu-\lambda)z}}}_{II}\right).
\end{aligned}$$

We solve the integrals in I and II,

$$\begin{aligned}
I &= \int \frac{\mu \mathrm{d}z}{\lambda - \mu e^{(\mu-\lambda)z}} \\
&= \frac{\mu}{\mu-\lambda} \int \frac{\mathrm{d}v}{v(\lambda - v)}, \quad v = \mu e^{(\mu-\lambda)z}, \quad dv = v(\mu-\lambda)dz \\
&= \frac{\mu}{\mu-\lambda} \left[\int \frac{\mathrm{d}v}{\lambda v} + \int \frac{\mathrm{d}v}{\lambda(\lambda - v)} \right] \\
&= \frac{\mu}{\mu-\lambda} \left[\frac{1}{\lambda} \ln v - \frac{1}{\lambda} \ln(\lambda - v) + c_0 \right] \\
&= \frac{\mu}{\lambda(\mu-\lambda)} \left[\ln \left(\frac{v}{\lambda - v} \right) + c_0 \right] \\
I &= \frac{\mu}{\lambda(\mu-\lambda)} \ln \left(\frac{\mu e^{(\mu-\lambda)z}}{\lambda - \mu e^{(\mu-\lambda)z}} \right) + c_1,
\end{aligned}$$

and

$$\begin{aligned}
II &= \int \frac{\mu e^{(\mu-\lambda)z} \mathrm{d}z}{\lambda - \mu e^{(\mu-\lambda)z}} \\
&= \int \frac{m \mathrm{d}m}{(\mu-\lambda)m(\lambda - m)}, \quad m = \mu e^{(\mu-\lambda)z}, \quad \mathrm{d}m = (\mu-\lambda)m\mathrm{d}z \\
&= \frac{1}{\mu-\lambda} \int \frac{\mathrm{d}m}{\lambda - m} \\
&= \frac{1}{\lambda-\mu} \ln(\lambda - m) + c \\
II &= \frac{1}{\lambda-\mu} \ln \left(\lambda - \mu e^{(\mu-\lambda)z} \right) + c.
\end{aligned}$$

Thus,

$$\begin{aligned}
\int g(z)\mathrm{d}z &= 2\lambda \left(\frac{\mu}{\lambda(\mu-\lambda)} \ln \left(\frac{\mu e^{(\mu-\lambda)z}}{\lambda - \mu e^{(\mu-\lambda)z}} \right) - \frac{1}{\lambda-\mu} \ln \left(\lambda - \mu e^{(\mu-\lambda)z} \right) \right) + c \\
&= 2\lambda \left(\frac{\mu}{\lambda(\mu-\lambda)} \ln \left(\mu e^{(\mu-\lambda)z} \right) + \frac{\lambda-\mu}{\lambda(\mu-\lambda)} \ln \left(\lambda - \mu e^{(\mu-\lambda)z} \right) \right) + c \\
&= 2 \left(\frac{\mu}{\mu-\lambda} \ln \left(\mu e^{(\mu-\lambda)z} \right) - \ln \left(\lambda - \mu e^{(\mu-\lambda)z} \right) \right) + c.
\end{aligned}$$

Substituting back into Eq. (58), we get,

$$\ln\left(D_t^{(1)}(z)\right) = -(\lambda+\mu)z + 2\left(\frac{\mu}{\mu-\lambda}\ln\left(\mu e^{(\mu-\lambda)z}\right) - \ln\left(\lambda-\mu e^{(\mu-\lambda)z}\right)\right) + C$$

$$\ln\left(D_t^{(1)}(z)\right) = -(\lambda+\mu)z + 2\ln\left(\frac{\left(\mu e^{(\mu-\lambda)z}\right)^{\frac{\mu}{\mu-\lambda}}}{\left(\lambda-\mu e^{(\mu-\lambda)z}\right)}\right) + C$$

$$D_t^{(1)}(z) = K\left(\frac{\left(\mu e^{(\mu-\lambda)z}\right)^{\frac{\mu}{\mu-\lambda}}}{\left(\lambda-\mu e^{(\mu-\lambda)z}\right)}\right)^2 e^{-(\lambda+\mu)z}.$$

Since $D_t^{(1)}(0) = 1$, we have

$$K = \left(\frac{\lambda-\mu}{\mu^{\frac{\mu}{\mu-\lambda}}}\right)^2.$$

Therefore,

$$\begin{aligned} D_t^{(1)}(z) &= \left[\frac{\lambda-\mu}{\mu^{\frac{\mu}{\mu-\lambda}}} \times \frac{\left(\mu e^{(\mu-\lambda)z}\right)^{\frac{\mu}{\mu-\lambda}}}{\left(\lambda-\mu e^{(\mu-\lambda)z}\right)}\right]^2 e^{-(\lambda+\mu)z} \\ &= \left[\frac{(\lambda-\mu)e^{\mu z}}{\lambda-\mu e^{(\mu-\lambda)z}}\right]^2 e^{-(\lambda+\mu)z}. \end{aligned}$$

□

6.5 Deriving and Solving the Differential Equation of $G_{x,t}(z)$

Here, we show the derivation of the differential equation of the probability of observing a reconstructed internal branch with length z on a tree with age t, $G_{x,t}(z)$, shown in Eq. 21, using physical interpretations where x is the length of an external branch descending from that internal branch. Then, we derive the solution to the differential equation as shown in Eq. 25.

We can write $G_{x,t}(z)$ by conditioning on the time of the first event on that internal branch with elapsed time z on a tree with age t (the elapsed time since the beginning of the tree starting at time 0). That is, (1) the branch has not undergone any observable event yet at time $z \leq t$, which occurs with probability $e^{-(\lambda+\mu)z}$ or (2) the branch has a child at some time $z-u \leq z$, which occurs with probability $e^{-(\lambda+\mu)(z-u)}\lambda$, and so the two branches evolve independently of each other where the child branch becomes extinct by time t with elapsed time $u+x$ with probability $E(u+x)$ and the initial branch survives until time $u \leq z$ with probability $G_{x,t}(u)$ or vice versa. Thus,

$$G_{x,t}(z) = e^{-(\lambda+\mu)z} + \int_{u=0}^{z} e^{-(\lambda+\mu)(z-u)}\lambda\left(2G_{x,t}(u)E(u+x)\right)\mathrm{d}u.$$

Taking the derivative with respect to z on both sides of the equation above, we have

$$\frac{\mathrm{d}G_{x,t}(z)}{\mathrm{d}z} = \frac{\mathrm{d}}{\mathrm{d}z}\left(e^{-(\lambda+\mu)z}\right) + \frac{\mathrm{d}}{\mathrm{d}z}\left(\int_{u=0}^{z} e^{-(\lambda+\mu)(z-u)}\lambda\left(2G_{x,t}(u)E(u+x)\right)\mathrm{d}u\right).$$

Then, by applying the Leibniz integral rule we have,

$$\begin{aligned}
\frac{\mathrm{d}G_{x,t}(z)}{\mathrm{d}z} &= -(\lambda+\mu)e^{-(\lambda+\mu)z} + 2\lambda G_{x,t}(z)E(z+x) \\
&\quad + \int_{u=0}^{z} \frac{\mathrm{d}}{\mathrm{d}z}\left(e^{-(\lambda+\mu)(z-u)}\lambda(2G_{x,t}(u)E(u+x))\right)\mathrm{d}u \\
&= -(\lambda+\mu)e^{-(\lambda+\mu)z} + 2\lambda G_{x,t}(z)E(z+x) - (\lambda+\mu) \\
&\quad \int_{u=0}^{z} e^{-(\lambda+\mu)(z-u)}\lambda 2G_{x,t}(u)E(u+x)\mathrm{d}u \\
&= -(\lambda+\mu)\left(e^{-(\lambda+\mu)z} + \int_{u=0}^{z} e^{-(\lambda+\mu)(z-u)}\lambda 2G_{x,t}(u)E(u+x)\mathrm{d}u\right) \\
&\quad +2\lambda G_{x,t}(z)E(z+x) \\
&= -(\lambda+\mu)G_{x,t}(z) + 2\lambda G_{x,t}(z)E(z+x), \quad \text{as in Eq. 21.}
\end{aligned}$$

Lemma 4

$$G_{x,t}(z) = \left(\frac{\lambda - \mu e^{(\mu-\lambda)x}}{\lambda - \mu e^{(\mu-\lambda)(z+x)}}\right)^2 e^{(\mu-\lambda)z}$$

is the solution to the differential equation,

$$\frac{\mathrm{d}G_{x,t}(z)}{\mathrm{d}z} = -(\lambda+\mu)G_{x,t}(z) + 2\lambda G_{x,t}(z)E(z+x),$$

where $G_{x,t}(0) = 1$.

Proof We solve the differential equation for $G_{x,t}(z)$ as follows

$$\begin{aligned}
\frac{\mathrm{d}G_{x,t}(z)}{\mathrm{d}z} &= -(\lambda+\mu)G_{x,t}(z) + 2\lambda G_{x,t}(z)E(z+x) \\
\int \frac{dG_{x,t}(z)}{G_{x,t}(z)} &= \int \left(-(\lambda+\mu) + 2\lambda E(z+x)\right)\mathrm{d}z + C \\
\ln\left(G_{x,t}(z)\right) &= -(\lambda+\mu)z + 2\lambda \int E(z+x)\mathrm{d}z + C \\
\ln\left(G_{x,t}(z)\right) &= -(\lambda+\mu)z + 2\lambda \int E(v)\mathrm{d}v + C, \quad v = z+x, \quad dv = dz \\
G_{x,t}(z) &= Ke^{-(\lambda+\mu)z + 2\lambda\int E(v)\mathrm{d}v}. \qquad (58)
\end{aligned}$$

Next, we solve $\int E(v)\mathrm{d}v$.

$$\begin{aligned}\int E(v)\mathrm{d}v &= \int \frac{\mu-\mu e^{-(\lambda-\mu)v}}{\lambda-\mu e^{-(\lambda-\mu)v}}\mathrm{d}v \\ &= \int \frac{\mu-w}{-(\lambda-w)}\frac{1}{(\lambda-\mu)w}\mathrm{d}w, \quad w=\mu e^{-(\lambda-\mu)v}, \quad \mathrm{d}w=-(\lambda-\mu)w\mathrm{d}\mathrm{d}v \\ &= \frac{1}{\lambda-\mu}\int \frac{w-\mu}{(\lambda-w)w}\mathrm{d}w \\ &= \frac{1}{\lambda-\mu}\left(\int \frac{1}{\lambda-w}\mathrm{d}w - \int \frac{\mu}{(\lambda-w)w}\mathrm{d}w\right) \\ &= \frac{1}{\lambda-\mu}(A-B).\end{aligned}$$

Note that

$$\begin{aligned}A &= \int \frac{1}{\lambda-w}\mathrm{d}w \\ &= -\int \frac{\mathrm{d}(\lambda-w)}{\lambda-w} \\ &= -\ln(\lambda-w)+C. \\ B &= \int \frac{\mu}{(\lambda-w)w}\mathrm{d}w \\ &= \mu\int \frac{1}{(\lambda-w)w}\mathrm{d}w \\ &= \mu\left(\int \frac{1}{\lambda(\lambda-w)}\mathrm{d}w + \int \frac{1}{\lambda w}\mathrm{d}w\right) \\ &= \mu\left(-\frac{1}{\lambda}\ln(\lambda-w)+\frac{1}{\lambda}\ln(w)\right) \\ &= \frac{\mu}{\lambda}(\ln(w)-\ln(\lambda-w)) \\ &= \ln\left(\frac{w}{\lambda-w}\right)^{\frac{\mu}{\lambda}}+C.\end{aligned}$$

Thus,

$$\begin{aligned}\int E(v)\mathrm{d}v &= \frac{1}{\lambda-\mu}\left(-\ln(\lambda-w)-\ln\left(\frac{w}{\lambda-w}\right)^{\frac{\mu}{\lambda}}\right) \\ &= \frac{1}{\mu-\lambda}\left(\ln\left((\lambda-w)\frac{w^{\frac{\mu}{\lambda}}}{(\lambda-w)^{\frac{\mu}{\lambda}}}\right)\right) \\ &= \frac{1}{\mu-\lambda}\left(\ln\left(\frac{w^{\frac{\mu}{\lambda}}}{(\lambda-w)^{\frac{\mu}{\lambda}-1}}\right)\right)\end{aligned}$$

$$= \ln\left(\frac{w^{\frac{\mu}{\lambda}}}{(\lambda - w)^{\frac{\mu}{\lambda}-1}}\right)^{\frac{1}{\mu-\lambda}}$$
$$= \ln\left(\frac{(\mu e^{-(\lambda-\mu)v})^{\frac{\mu}{\lambda}}}{(\lambda - \mu e^{-(\lambda-\mu)v})^{\frac{\mu}{\lambda}-1}}\right)^{\frac{1}{\mu-\lambda}}.$$

Substituting this back into Eq. (58), we get,

$$\begin{aligned}
G_{x,t}(z) &= K e^{-(\lambda+\mu)z + 2\lambda\int E(v)\mathrm{d}v} \\
&= K\left(e^{-(\lambda+\mu)z}\left(\frac{(\mu e^{-(\lambda-\mu)v})^{\frac{\mu}{\lambda}}}{(\lambda - \mu e^{-(\lambda-\mu)v})^{\frac{\mu}{\lambda}-1}}\right)^{\frac{2\lambda}{\mu-\lambda}}\right) \\
&= K\left(e^{-(\lambda+\mu)z}\frac{\left(\mu e^{-(\lambda-\mu)v}\right)^{\frac{2\mu}{\mu-\lambda}}}{\left(\lambda - \mu e^{-(\lambda-\mu)v}\right)^2}\right) \\
&= K\left(\frac{\left(\mu e^{-(\lambda-\mu)v}\right)^{\frac{\mu}{\mu-\lambda}}}{\lambda - \mu e^{-(\lambda-\mu)v}}\right)^2 e^{-(\lambda+\mu)z} \\
&= K\left(\frac{\mu^{\frac{\mu}{\mu-\lambda}} e^{\mu v}}{\lambda - \mu e^{-(\lambda-\mu)v}}\right)^2 e^{-(\lambda+\mu)z} \\
&= K\left(\frac{\mu^{\frac{\mu}{\mu-\lambda}} e^{\mu(z+x)}}{\lambda - \mu e^{(\mu-\lambda)(z+x)}}\right)^2 e^{-(\lambda+\mu)z}.
\end{aligned}$$

Given that $G_{x,t}(0) = 1$, we have,

$$K = \left(\frac{\lambda - \mu e^{(\mu-\lambda)x}}{\mu^{\frac{\mu}{\mu-\lambda}} e^{\mu x}}\right)^2.$$

Thus,

$$\begin{aligned}
G_{x,t}(z) &= \left(\frac{\lambda - \mu e^{(\mu-\lambda)x}}{\mu^{\frac{\mu}{\mu-\lambda}} e^{\mu x}}\right)^2 \left(\frac{\mu^{\frac{\mu}{\mu-\lambda}} e^{\mu(z+x)}}{\lambda - \mu e^{(\mu-\lambda)(z+x)}}\right)^2 e^{-(\lambda+\mu)z} \\
&= \frac{\left(\lambda - \mu e^{(\mu-\lambda)x}\right)^2 e^{(\mu-\lambda)z}}{\left(\lambda - \mu e^{(\mu-\lambda)(z+x)}\right)^2} \\
&= \left(\frac{\lambda - \mu e^{(\mu-\lambda)x}}{\lambda - \mu e^{(\mu-\lambda)(z+x)}}\right)^2 e^{(\mu-\lambda)z}. \qquad (59)
\end{aligned}$$

□

6.6 Showing the Relationship with $\mathcal{L}_{\text{Nee}}(T^* \mid \lambda, \mu)$

Here, we show the relationship between the likelihood $\mathcal{L}(T^* \mid \lambda, \mu)$ shown in Eq. 19 with the likelihood $\mathcal{L}_{\text{Nee}}(T^* \mid \lambda, \mu)$ in Nee et al. (1994b). We show this for the case of the reconstructed tree T^* with three tips as pictured in Fig. 3. On this reconstructed tree with age t, ignoring the age of the root, the external branch 2 is one of the two original branches descending from the root, so it has length $\tilde{b}_2 = x_2$. We assume that the other two external branches have length $\tilde{b}_3 = \tilde{b}_4 = x_3$. Thus, the internal branch 1 has length $b_1 = x_2 - x_3$.

Lemma 5

$$\mathcal{L}_{\text{Nee}}(T^* \mid \lambda, \mu) = \frac{\mathcal{L}(T^* \mid \lambda, \mu)}{(1 - E(x_2))^2},$$

where $\mathcal{L}(T^* \mid \lambda, \mu)$ is given in Eq. 19, $\mathcal{L}_{\text{Nee}}(T^* \mid \lambda, \mu)$ is given in Eq. (20) in Nee et al. (1994b)*, x_2 is the elapsed time from the starting time of the two original branches descending from the root of the tree T^* until the end of the tree, and $E(x_2)$ is the extinction probability for one of those branches.*

Proof Under the reconstructed tree T^* with three tips, the likelihood in Eq. 19 is simplified to,

$$\begin{aligned}
\mathcal{L}(T^* \mid \lambda, \mu) &= 2! \times G_{x_1,t}(b_1) \times \lambda \times D_t^{(1)}(\tilde{b}_2) \times D_t^{(1)}(\tilde{b}_3) \times D_t^{(1)}(\tilde{b}_4) \\
&= 2\lambda \left(\frac{\lambda - \mu e^{-(\lambda-\mu)x_3}}{\lambda - \mu e^{-(\lambda-\mu)x_2}} \right)^2 e^{-(\lambda-\mu)(x_2-x_3)} \left(\frac{(\lambda-\mu)e^{\mu x_2}}{\lambda - \mu e^{-(\lambda-\mu)x_2}} \right)^2 e^{-(\lambda+\mu)x_2} \\
&\quad \left(\frac{(\lambda-\mu)e^{\mu x_3}}{\lambda - \mu e^{-(\lambda-\mu)x_3}} \right)^4 e^{-2(\lambda+\mu)x_3} \\
&= 2\lambda \left(\frac{1}{\lambda - \mu e^{-(\lambda-\mu)x_2}} \right)^2 \left(\frac{(\lambda-\mu)e^{-(\lambda-\mu)x_2}}{\lambda - \mu e^{-(\lambda-\mu)x_2}} \right)^2 \\
&\quad \left(\frac{(\lambda-\mu)e^{-(\lambda-\mu)x_3}}{\lambda - \mu e^{-(\lambda-\mu)x_3}} \right)^2 \left(\frac{(\lambda-\mu)^3}{\lambda - \mu e^{-(\lambda-\mu)x_3}} \right) \\
&= 2\lambda \left(\frac{\lambda-\mu}{\lambda - \mu e^{-(\lambda-\mu)x_3}} \right) \left(\frac{(\lambda-\mu)e^{-(\lambda-\mu)x_2}}{\lambda - \mu e^{-(\lambda-\mu)x_2}} \right)^2 \\
&\quad \left(\frac{(\lambda-\mu)e^{-(\lambda-\mu)x_3}}{\lambda - \mu e^{-(\lambda-\mu)x_3}} \right) \left(\frac{\lambda-\mu}{\lambda - \mu e^{-(\lambda-\mu)x_2}} \right)^2 \\
&= \mathcal{L}_{\text{Nee}}(T^* \mid \lambda, \mu)(1 - E(x_2))^2 \\
&\quad \mathcal{L}_{\text{Nee}}(T^* \mid \lambda, \mu) = \frac{\mathcal{L}(T^* \mid \lambda, \mu)}{(1 - E(x_2))^2}.
\end{aligned}$$

□

References

Akaike H (1998) Information theory and an extension of the maximum likelihood principle. In: Parzen E, Tanabe K, Kitagawa G (eds) Selected papers of Hirotugu Akaike. Springer, New York, pp 199–213

Aldous DJ (1996) Probability distributions on cladograms. In: Aldous D, Pemantle R (eds) Random discrete structures. Springer, New York, pp 1–18

Aldous DJ (2001) Stochastic models and descriptive statistics for phylogenetic trees, from Yule to today. Stat Sci 16(1):23–34

Anderson D, Burnham K (2004) Model selection and multi-model inference, vol 63. Springer, New York, p 10

Asmussen S, Nerman O, Olsson M (1996) Fitting phase-type distributions via the EM algorithm. Scand J Stat 23:419–441

Bellman R, Harris TE (1948) On the theory of age-dependent stochastic branching processes. Proc Natl Acad Sci USA 34(12):601

Bortolussi N, Durand E, Blum M, François O (2006) apTreeshape: statistical analysis of phylogenetic tree shape. Bioinformatics 22(3):363–364

Byrd RH, Lu P, Nocedal J, Zhu C (1995) A limited memory algorithm for bound constrained optimization. SIAM J Sci Comput 16(5):1190–1208

Colless DH (1982) Review of phylogenetics: the theory and practice of phylogenetic systematics. Syst Zool 31(1):100–104

Cumani A (1982) On the canonical representation of homogeneous Markov processes modelling failure-time distributions. Microelectron Reliab 22(3):583–602

Dehon M, Latouche G (1982) A geometric interpretation of the relations between the exponential and generalized Erlang distributions. Adv Appl Probab 14(4):885–897

Etienne RS, Haegeman B, Stadler T, Aze T, Pearson PN, Purvis A, Phillimore AB (2012) Diversity-dependence brings molecular phylogenies closer to agreement with the fossil record. Proc R Soc B Biol Sci 279(1732):1300–1309

FitzJohn RG (2012) Diversitree: comparative phylogenetic analyses of diversification in R. Methods Ecol Evol 3(6):1084–1092

Hagen O, Stadler T (2018) TreeSimGM: simulating phylogenetic trees under general Bellman–Harris models with lineage-specific shifts of speciation and extinction in R. Methods Ecol Evol 9(3):754–760

Hagen O, Hartmann K, Steel M, Stadler T (2015) Age-dependent speciation can explain the shape of empirical phylogenies. Syst Biol 64(3):432–440. https://doi.org/10.1093/sysbio/syv001

Hagen O, Andermann T, Quental TB, Antonelli A, Silvestro D (2018) Estimating age-dependent extinction: contrasting evidence from fossils and phylogenies. Syst Biol 67(3):458–474

Harvey PH, Pagel MD (1991) The comparative method in evolutionary biology, vol 239. Oxford University Press, Oxford

Huson DH, Scornavacca C (2012) Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. Syst Biol 61(6):1061–1067

Kendall DG (1948) On the generalized birth-and-death process. Ann Math Stat 19(1):1–15

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

Louca S, Pennell MW (2020) Extant timetrees are consistent with a myriad of diversification histories. Nature 580(7804):502–505

Maddison WP, Midford PE, Otto SP (2007) Estimating a binary character's effect on speciation and extinction. Syst Biol 56(5):701–710

Marshall AH, McClean SI (2004) Using Coxian phase-type distributions to identify patient characteristics for duration of stay in hospital. Health Care Manag Sci 7(4):285–289

Morlon H (2014) Phylogenetic approaches for studying diversification. Ecol Lett 17(4):508–525

Morlon H, Potts MD, Plotkin JB (2010) Inferring the dynamics of diversification: a coalescent approach. PLoS Biol 8(9):e1000493

Morlon H, Parsons TL, Plotkin JB (2011) Reconciling molecular phylogenies with the fossil record. Proc Natl Acad Sci USA 108(39):16327–16332

Nee S, Mooers AO, Harvey PH (1992) Tempo and mode of evolution revealed from molecular phylogenies. Proc Natl Acad Sci USA 89(17):8322–8326

Nee S, Holmes EC, May RM, Harvey PH (1994a) Extinction rates can be estimated from molecular phylogenies. Philos Trans R Soc Lond B 344(1307):77–82
Nee S, May RM, Harvey PH (1994b) The reconstructed evolutionary process. Philos Trans R Soc Lond B Biol Sci 344(1309):305–311
Nelder JA, Mead R (1965) A simplex method for function minimization. Comput J 7(4):308–313
Neuts MF (1975) Probability distributions of phase-type. Liber Amicorum Prof Emeritus H Florin, Department of Mathematics, University of Louvain
Neuts MF (1981) Matrix-geometric solutions in stochastic models: an algorithmic approach. Johns Hopkins University Press, Baltimore
Okamura H, Dohi T (2016) Ph fitting algorithm and its application to reliability engineering. J Oper Res Soc Jpn 59(1):72–109
Paradis E, Claude J, Strimmer K (2004) APE: analyses of phylogenetics and evolution in R language. Bioinformatics 20(2):289–290
Pawitan Y (2001) In all likelihood: statistical modelling and inference using likelihood. Oxford University Press, Oxford
Phillimore AB, Price TD (2008) Density-dependent cladogenesis in birds. PLoS Biol 6(3):e71
Pybus OG, Harvey PH (2000) Testing macro-evolutionary models using incomplete molecular phylogenies. Proc R Soc B 267(1459):2267–2272
Quental TB, Marshall CR (2010) Diversity dynamics: molecular phylogenies need the fossil record. Trends Ecol Evol 25(8):434–441
Rabosky DL (2006) Likelihood methods for detecting temporal shifts in diversification rates. Evolution 60(6):1152–1164
Rabosky DL, Lovette IJ (2008) Density-dependent diversification in north American wood warblers. Proc R Soc B Biol Sci 275(1649):2363–2371
Revell LJ (2012) phytools: an r package for phylogenetic comparative biology (and other things). Methods Ecol Evol 2:217–223
Ricklefs RE (2007) Estimating diversification rates from phylogenetic information. Trends Ecol Evol 22(11):601–610
Sackin M (1972) good and bad phenograms. Syst Biol 21(2):225–226
Simpson GG (1944) Tempo and mode in evolution. Columbia University Press, New York
Stadler T (2013a) How can we improve accuracy of macroevolutionary rate estimates? Syst Biol 62(2):321–329
Stadler T (2013b) Recovering speciation and extinction dynamics based on phylogenies. J Evol Biol 26(6):1203–1219
Stanley SM (1998) Macroevolution: pattern and process. Johns Hopkins University Press, Baltimore
Steel M (2016) Phylogeny: discrete and random processes in evolution. SIAM, Philadelphia
Thummler A, Buchholz P, Telek M (2006) A novel approach for phase-type fitting with the EM algorithm. IEEE Trans Dependable Secure Comput 3(3):245–258
Verbelen R (2013) Phase-type distributions & mixtures of erlangs. Ph.D. thesis, University of Leuven
Yule GU (1925) Ii.—a mathematical theory of evolution, based on the conclusions of dr. jc willis, fr s. Philos Trans R Soc Lond B 213(402–410):21–87
Zanne AE, Tank DC, Cornwell WK, Eastman JM, Smith SA, FitzJohn RG, McGlinn DJ, O'Meara BC, Moles AT, Reich PB et al (2014) Three keys to the radiation of angiosperms into freezing environments. Nature 506(7486):89–92
Zheng Y, Wiens JJ (2016) Combining phylogenomic and supermatrix approaches, and a time-calibrated phylogeny for squamate reptiles (lizards and snakes) based on 52 genes and 4162 species. Mol Phylogenet Evol 94:537–547