



Review



Cite this article: Rannala B, Yang Z. 2025
Reading tree leaves: inferring speciation and
extinction processes using phylogenies. *Phil.
Trans. R. Soc. B* **380**: 20230309.
<https://doi.org/10.1098/rstb.2023.0309>

Received: 18 July 2024

Accepted: 14 October 2024

One contribution of 18 to a theme issue ‘‘A
mathematical theory of evolution’’: phylogenetic
models dating back 100 years’.

Subject Areas:

evolution, computational biology, taxonomy and
systematics

Keywords:

Yule process, generalized birth–death process,
identifiability, statistical inference, phylogenetic
trees

Author for correspondence:

Ziheng Yang

e-mail: z.yang@ucl.ac.uk

Reading tree leaves: inferring speciation and extinction processes using phylogenies

Bruce Rannala¹ and Ziheng Yang²

¹Department of Evolution and Ecology, University of California, Davis, CA 95616, USA

²Department of Genetics, Evolution, and Environment, University College London, London WC1E 6BT, UK

id BR, 0000-0002-8355-9955; ZY, 0000-0003-3351-7981

The birth–death process (BDP) is widely used in evolutionary biology as a model for generating phylogenetic trees of species. The generalized birth–death process (GBDP) allows rate variation over time, with speciation and extinction rates to be arbitrary functions of time. Here we review the probability theory underpinning the GBDP as a model of cladogenesis and recent findings concerning its identifiability. The GBDP with arbitrary continuous rate functions has been shown to be non-identifiable from lineage-through-time data: even with species phylogenies of infinite size the parameters cannot be estimated. However, a restricted class of BDPs with piecewise-constant rates has been shown to be identifiable. We review and illustrate these results using simple examples and discuss their implications for biologists interested in inferring the past tempo and mode of evolution using reconstructed phylogenetic trees.

This article is part of the theme issue ‘‘A mathematical theory of evolution’’: phylogenetic models dating back 100 years’.

1. Introduction

Phylogenetic relationships among species are a consequence of complex historical processes, including population fragmentation and divergence, genetic drift and ecological adaptation, ultimately leading to increased genetic isolation and generating new species. These processes naturally lead to hierarchical relationships among extant and extinct species, including shared ancestral species existing at various times in the past. An evolutionary tree representation of relationships among species has been used since the time of Darwin in the nineteenth century [1]. ‘Tree thinking’ has become one of the fundamental unifying principles of evolutionary biology [2].

Relationships among species over time are predominantly represented as a binary tree (phylogeny) with branch lengths proportional to time (see [figure 1](#)). Although genomic datasets frequently contradict this tree view, supporting more nuanced species relationships defined by introgression and horizontal gene transfer which generates horizontal connections between branches and produces networks rather than binary phylogenetic trees [3], binary trees still provide a useful backbone onto which introgression may be superimposed. Phylogenetic trees also provide a record of the tempo and mode of past evolution in different groups of species [4].

Unique geological and environmental events are expected to play a major role in generating phylogenetic relationships among species. Any process-based mechanistic model of speciation and extinction will be highly complex and group-specific, omit important unobserved factors, and be unrealistic. Nonetheless, biologists have a deep interest in discovering the forces that generated biological diversity, such as geological upheavals, climatic catastrophes and extraterrestrial collisions that may be associated

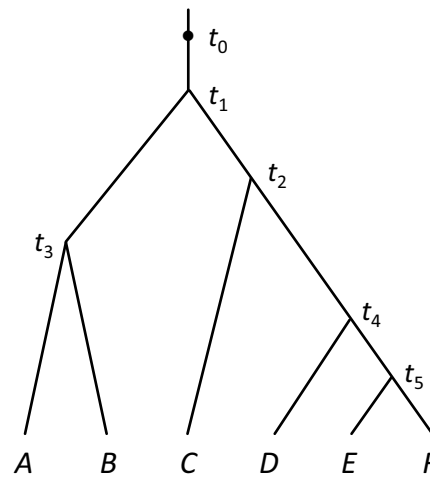


Figure 1. A species phylogeny for six species with internal nodes rank-ordered by age. A tree with labelled tips and rank-ordered internal nodes is called a ‘labelled history’ or ‘ranked rooted tree’. The time of origin, t_0 , is used to formulate the birth–death process (BDP) model of cladogenesis. In models considered in this article, the tree topology is not informative about the parameters in the BDP, and only the speciation times (or equivalently the ‘lineage through time’ or LTT data) constitute the data: $t_1, \dots, t_5$.

with radiative speciations or mass extinctions. Unfortunately, the extreme stochasticity of even ‘neutral’ branching processes with constant speciation and extinction rates can mislead inferences based solely on apparent diversification rates on the phylogenetic tree or based on ‘reading of the tree leaves’. To infer speciation and extinction processes rigorously requires a proper statistical inference framework.

To study speciation and extinction using estimated phylogenies a model of the processes generating the phylogeny is needed. Here we focus on a class of stochastic continuous-time birth–death processes (BDPs) whose study began a century ago with the pure-birth process proposed by Udny Yule [5]. BDPs have provided simple parametric models that allow formal statistical inference procedures to be used to estimate birth (speciation) and death (extinction) rates from phylogenetic trees. BDPs are obviously not an accurate model of the true underlying processes; however, they attempt to capture the most important features with per-species rates of speciation and extinction that may vary through time or across lineages. BDP-based inference methods have been widely used in evolutionary biology [6], but such approaches have also been criticized on several grounds [7–9]. Here we review some recent criticisms and controversies, particularly concerning whether certain BDPs can be distinguished by data. We provide a description of the most widely used BDP model that allows variable rates, the generalized birth–death process (GBDP) [10], and consider the statistical properties of inference methods under the model.

Two distinct approaches have been taken to extend BDPs to account for variable rates of speciation and extinction: variation of rates through time (with rates shared across all contemporary species) [11] and variation of rates across lineages [12,13]. An important difference is that in the first (rate variation through time), all tree topologies with rank-ordered internal nodes and labelled tips (so-called labelled histories; see §4 below) have equal probabilities. As a consequence, divergence times on the phylogeny or the so-called ‘lineages through time’ (LTT) summary [11,14] provide a sufficient statistic for estimating speciation and extinction rates. This is not the case in the second approach (with rate variation across lineages), in which tree topologies are also informative about the rates. In this article, we consider only models of rate variation through time. However, many of our observations are of relevance for inferences under the second class of models as well.

2. Phylogenetic trees as data

The data used to infer speciation and extinction rates are usually phylogenetic trees of species. Here we assume that fossils are excluded and all species tips on the tree are contemporaneous; this is typical of most analyses (we do not consider applications of BDPs as models of virus transmission with non-contemporaneous tips or ancient DNA samples). Such ‘time trees’ are generated by analysing sequence data from a sample of contemporary species using phylogenetic inference methods that assume a molecular clock or a ‘relaxed clock’ [15,16]. The branch lengths of the phylogeny are in units of the expected number of substitutions but can be converted to absolute geological time by use of fossil calibrations. An example of a phylogenetic tree is shown in figure 1. Such trees provide the history of surviving species; extinct species are effectively removed from the tree if they are not direct ancestors. Nee *et al.* [11] referred to this tree of relationships for only the contemporary species as the ‘reconstructed tree’.

Comparative studies using BDPs typically do not distinguish between gene trees and species trees although processes such as incomplete lineage sorting could lead to gene tree versus species tree conflicts and biased species divergence time estimates [17,18]. Also, treating inferred phylogenetic trees as observed data to infer speciation and extinction rates ignores phylogenetic uncertainty (in particular uncertainty in branch lengths). One solution to this problem is to integrate over the trees while estimating rates through Bayesian phylogenetic inference with a BDP used as a prior [19,20]. Another approximation uses trees sampled from a separate Bayesian phylogenetic analysis through Markov chain Monte Carlo (MCMC) and averages inferences over the MCMC sample [21]. Although potentially important, we ignore species tree versus gene tree conflicts and phylogenetic

uncertainty in our considerations, focusing instead on the ‘best-case’ scenario by using a known phylogenetic (species) tree to infer rates.

3. The Yule process

The Yule process [5] is a continuous-time pure-birth process in which birth (speciation) occurs at the per-lineage rate λ and there are no extinctions. The process begins with at least one lineage and the number of lineages is strictly increasing. For calculating probabilities on phylogenetic trees the only transition probability needed is $p_1(t)$, the probability that a lineage existing at time 0 has exactly one descendant time t later. If the process is initiated with $n_0 = 1$ lineages at time $t = 0$ the finite difference equation is

$$p_1(t + \Delta t) = p_1(t)[1 - \lambda\Delta t] + o(\Delta t), \quad (3.1)$$

and the limiting ordinary differential equation (ODE) is

$$\begin{aligned} \lim_{\Delta t \rightarrow 0} \frac{p_1(t + \Delta t) - p_1(t)}{\Delta t} &= \lim_{\Delta t \rightarrow 0} \left(-p_1(t)\lambda + \frac{o(\Delta t)}{\Delta t} \right) \\ \frac{dp_1(t)}{dt} &= -p_1(t)\lambda. \end{aligned} \quad (3.2)$$

Solving (3.2) with initial condition $p_1(0) = 1$ gives

$$p_1(t) = e^{-\lambda t}. \quad (3.3)$$

The general formula for $p_n(t)$ can be obtained by iteratively solving the successive ODEs. For example,

$$p_2(t + \Delta t) = p_1(t)\lambda\Delta t + p_2(t)(1 - 2\lambda\Delta t) + o(\Delta t), \quad (3.4)$$

which gives the ODE

$$\frac{dp_2(t)}{dt} = -2p_2(t)\lambda + p_1(t)\lambda = -2p_2(t)\lambda + e^{-\lambda t}\lambda. \quad (3.5)$$

With the initial condition $p_2(0) = 0$, this has the solution

$$p_2(t) = e^{-2\lambda t}(e^{\lambda t} - 1). \quad (3.6)$$

In general, $p_n(t)$ is given as

$$p_n(t) = e^{-\lambda t}(1 - e^{-\lambda t})^{n-1}. \quad (3.7)$$

4. Probability of the reconstructed tree

Edwards [22] considered the calculation of the probability of a phylogenetic tree under a Yule process. The same approach applies under the GBDP considered later in this article with an obvious substitution of probability terms. We therefore review the results for the Yule process in some detail. Edwards refers to a phylogenetic tree with labelled tips and rank-ordered internal nodes as a ‘labelled history’ (figure 1). This is the fundamental tree structure under both the coalescent model and the birth–death model. Contribution [23] of this issue further defines the concept of a labelled history, extending it to multifurcating trees. Rank-ordering the heights (times) of speciation events on a tree with a stem at time t_0 , we have $\mathbf{t} = \{t_1, t_2, \dots, t_{n-1}\}$, where t_1 is the height of the root node (figure 1). A speciation event occurs on any of the $i-1$ branches of the tree that exist during time interval $(t_i + \Delta t, t_i)$ with probability $(i-1)\lambda\Delta t_i$ and the resulting lineage leaves exactly one descendant at time present with probability $p_1(t_i)$. The joint probability of having, after time t_0 , n tips and node heights $\mathbf{t}$ is then

$$f(\mathbf{t}, n | t_0) = p_1(t_0) \prod_{i=1}^{n-1} i\lambda p_1(t_i) = (n-1)!\lambda^{n-1} \prod_{i=0}^{n-1} p_1(t_i) = (n-1)!\lambda^{n-1} e^{-\lambda \sum_{i=0}^{n-1} t_i}. \quad (4.1)$$

Integrating over the branch lengths for a fixed stem age gives the marginal probability of n descendants after time t_0 ,

$$f(n | t_0) = \int_{t_1=t_0}^{t_2} \dots \int_{t_{n-1}=0}^{t_{n-2}} (n-1)!\lambda^{n-1} e^{-\lambda \sum_{i=0}^{n-1} t_i} d\mathbf{t} = e^{-n\lambda t_0} (e^{\lambda t_0} - 1)^{n-1}. \quad (4.2)$$

Some authors [24] use the probability density function (PDF) of speciation times conditional on having n tips and the stem age t_0 (i.e. time that BDP was initiated with a single lineage):

$$f(\mathbf{t} | t_0, n) = \frac{(n-1)!\lambda^{n-1} e^{-\lambda \sum_{i=0}^{n-1} t_i}}{p_n(t_0)} = \frac{(n-1)!\lambda^{n-1} e^{-\lambda \sum_{i=0}^{n-1} t_i}}{e^{-n\lambda t_0} (e^{\lambda t_0} - 1)^{n-1}}. \quad (4.3)$$

However, the stem age is not a property of observed phylogenetic trees so it must be dealt with, for example, by integrating it out using a Bayesian prior. Other authors [19] follow [22] and instead condition on the number of tips n and the age of the root (which is observed in the phylogenetic tree):

$$\Pr(n|t_1) = \sum_{i=1}^{n-1} p_i(t_1)p_{n-i}(t_1) = (n-1)e^{-2\lambda t_1}(1 - e^{-\lambda t_1})^{n-2}, \quad (4.4)$$

and thus

$$f(\mathbf{t}|t_1, n) = \frac{(n-1)! \lambda^{n-1} e^{-\lambda \sum_{i=1}^{n-1} t_i}}{\Pr(n|t_1)} = \frac{(n-2)! \lambda^{n-1} e^{-\lambda \sum_{i=1}^{n-1} t_i}}{e^{-2\lambda t_1} (1 - e^{-\lambda t_1})^{n-2}}. \quad (4.5)$$

There may be situations where a stem age is a biologically meaningful concept (and a prior may be available). For example, in branching models of cell fate in an organism, the stem age is the age of the zygote. However, in phylogenetic systematics the meaning of a stem age is typically obscure. Conditioning on stem age, [equation \(4.3\)](#) can be rearranged to give

$$f(\mathbf{t}|t_0, n) = \frac{p_1(t_0)}{p_n(t_0)} \prod_{i=1}^{n-1} i \lambda p_i(t_i) = \frac{1}{(1 - e^{-\lambda t_0})^{n-1}} \prod_{i=1}^{n-1} i \lambda p_i(t_i) = (n-1)! \prod_{i=1}^{n-1} \frac{\lambda p_i(t_i)}{1 - e^{-\lambda t_0}} = (n-1)! \prod_{i=1}^{n-1} \frac{\lambda e^{-\lambda t_i}}{1 - e^{-\lambda t_0}}. \quad (4.6)$$

This is the joint density of the order statistics for $n-1$ independent and identically distributed (i.i.d.) random variables [22] with kernel density:

$$f(x) = \frac{\lambda e^{-\lambda x}}{1 - e^{-\lambda t_0}}, \quad \text{for } 0 \leq x \leq t_0. \quad (4.7)$$

In the case of the Yule process, the kernel density is a simple truncated exponential. This order-statistic form holds for the BDP as well [24] (see also [25]). This representation suggests an efficient method for simulating from a BDP conditional on n and t_0 [24], as used by [26] to simulate large trees under the BDP. In this procedure, $n-1$ i.i.d. variables are simulated from the kernel (eq. 9 of [24]) and rank ordered to obtain a sample from the joint density.

5. The generalized birth–death process

The Yule process described above has obvious limitations. Extinction is not allowed in the model, and the speciation rate is assumed constant. Feller proposed a linear BDP to incorporate extinction, with constant rates of speciation and extinction (λ, μ). Kendall [10] developed a ‘generalized’ birth–death process (GBDP) model in which the birth rate $\lambda(t)$ and death rate $\mu(t)$ are deterministic continuous functions of time. Remarkably, this generalized process still allows exact analytical solutions. Explicit solutions are available for $p_i(t)$, the probability that i descendants exist for a single lineage after t units of time. Of particular interest for calculating the PDF of phylogenetic trees using Edwards’s [22] method are the probabilities $p_1(t)$ and $p_0(t)$ that exactly one descendant, or zero descendants, respectively, exist after time t . Note that the stochastic process runs forward in time but t measures time in the past with the present time being $t = 0$. For example,

$$p_0(t) = 1 - \frac{1}{1 + \int_0^t e^{\int_0^y [\mu(y) - \lambda(y)] dy} \mu(y) dy} \quad (5.1)$$

is the probability that all descendants of a single lineage have died out by time t . [Equation \(5.1\)](#) assumes that all extant species descended from a particular ancestral node are sampled (i.e. complete sampling). This is the *lineage extinction probability* $E(t)$ in [27] and $\Phi(t)$ in [28] (in the special case of complete sampling). We will instead use $p_0(t) \equiv \Phi(t) \equiv E(t)$, to be consistent with Kendall’s original notation.

Nee *et al.* [11] added random sampling to the linear BDP. If a fraction ρ of lineages are randomly sampled at present, the equivalent probability is determined by the finite difference equation

$$p_0(t + \Delta t) = p_0(t)[1 - \lambda(t + \Delta t)\Delta t - \mu(t + \Delta t)\Delta t] + \mu(t + \Delta t)\Delta t + p_0(t)^2 \lambda(t + \Delta t)\Delta t + o(\Delta t). \quad (5.2)$$

The three terms represent the probabilities that (i) no events occurred during the interval $(t + \Delta t, t)$, with probability $[1 - \lambda(t + \Delta t)\Delta t - \mu(t + \Delta t)\Delta t]$, and the lineage subsequently went extinct, with probability $p_0(t)$; (ii) a death occurred, with probability $\mu(t + \Delta t)\Delta t$, so that the lineage went extinct during $(t + \Delta t, t)$; and (iii) a birth occurred, with probability $\lambda(t + \Delta t)\Delta t$, and both lineages subsequently went extinct, with probability $p_0(t)^2$. Taking limits, we obtain the ODE

$$\begin{aligned} \lim_{\Delta t \rightarrow 0} \frac{p_0(t + \Delta t) - p_0(t)}{\Delta t} &= \lim_{\Delta t \rightarrow 0} \left(-p_0(t)[\lambda(t + \Delta t) + \mu(t + \Delta t)] + \mu(t + \Delta t) + p_0(t)^2 \lambda(t + \Delta t) + \frac{o(\Delta t)}{\Delta t} \right), \\ \frac{dp_0(t)}{dt} &= -p_0(t)[\lambda(t) + \mu(t)] + \mu(t) + p_0(t)^2 \lambda(t). \end{aligned} \quad (5.3)$$

Solving the equation with the initial condition $p_0(0) = 1 - \rho$ gives

$$p_0(t) = 1 - \frac{e^{\int_0^t \lambda(s) - \mu(s) ds}}{\frac{1}{\rho} + \int_0^t \lambda(s) e^{\int_0^s \lambda(y) - \mu(y) dy} ds}. \quad (5.4)$$

For constant λ and μ [equation \(5.4\)](#) simplifies to

$$p_0(t) = \frac{\rho \mu e^{(\lambda - \mu)t} + \lambda(1 - \rho) - \mu}{\rho \lambda e^{(\lambda - \mu)t} + \lambda(1 - \rho) - \mu}. \quad (5.5)$$

In the case of constant λ and μ with $\rho = 1$, both [equation 5.1](#) and [equation 5.4](#) simplify to

$$p_0(t) = \frac{\mu(e^{(\lambda-\mu)t} - 1)}{\lambda e^{(\lambda-\mu)t} - \mu}. \quad (5.6)$$

In calculating the PDF for labelled histories we also need the probability that a lineage has a single descendant after time t , the solution of which (with complete sampling) from Kendall [10] is

$$p_1(t) = [1 - p_0(t)](1 - \eta_t), \quad (5.7)$$

where

$$\eta_t = 1 - \frac{1}{e^{-\int_0^t [\mu(s) - \lambda(s)] ds} \left[1 + \int_0^t e^{\int_0^s [\mu(y) - \lambda(y)] dy} \mu(s) ds \right]}. \quad (5.8)$$

To obtain the PDF with sampling, a finite difference equation is again formulated

$$p_1(t + \Delta t) = p_1(t)[1 - \lambda(t + \Delta t)\Delta t - \mu(t + \Delta t)\Delta t] + 2p_1(t)p_0(t)\lambda(t + \Delta t)\Delta t + o(\Delta t). \quad (5.9)$$

The two terms are the probabilities (i) that no events occur during the interval $(t + \Delta t, t)$, with probability $1 - \lambda(t + \Delta t)\Delta t - \mu(t + \Delta t)\Delta t$, and exactly one lineage survives after time t , $p_1(t)$; and (ii) that a birth occurs during $(t + \Delta t, t)$, with probability $\lambda(t + \Delta t)\Delta t$, and exactly one of the lineages survives to present, with probability $2p_1(t)p_0(t)$. We have

$$\begin{aligned} \lim_{\Delta t \rightarrow 0} \frac{p_1(t + \Delta t) - p_1(t)}{\Delta t} &= \lim_{\Delta t \rightarrow 0} \left[-p_1(t)[\lambda(t + \Delta t) + \mu(t + \Delta t)] + 2p_1(t)p_0(t)\lambda(t + \Delta t) + \frac{o(\Delta t)}{\Delta t} \right], \\ \frac{dp_1(t)}{dt} &= -p_1(t)[\lambda(t) + \mu(t)] + 2p_1(t)p_0(t)\lambda(t). \end{aligned} \quad (5.10)$$

This ODE is solved using the initial condition $p_1(0) = \rho$ to account for sampling, which gives

$$p_1(t) = \frac{\rho \exp \{2 \int_0^t \lambda(s) p_0(s) ds\}}{\exp \{ \int_0^t [\lambda(s) + \mu(s)] ds \}}. \quad (5.11)$$

If λ and μ are constant this simplifies to

$$p_1(t) = \frac{e^{(\mu - \lambda)t} (\lambda - \mu)^2 \rho}{[(\mu - [1 - \rho]\lambda)e^{(\mu - \lambda)t} - \lambda\rho]^2}. \quad (5.12)$$

In the case that $\rho = 1$ this simplifies to

$$p_1(t) = \frac{(\lambda - \mu)^2 e^{(\lambda - \mu)t}}{[\lambda e^{(\lambda - \mu)t} - \mu]^2}, \quad (5.13)$$

which matches the result obtained by solving eq. 5.7 of [10].

6. Tree likelihoods under the generalized birth–death process and the generalized Yule process

Kubo & Iwasa [14] considered a GBDP in which $\lambda(t)$ varies through time and μ is constant, and another GBDP in which $\mu(t)$ varies through time while λ is constant. They noted that both GBDPs produce the same likelihood on LTT data as a generalized Yule process (GYP) with a time-variable birth rate (and with $\mu = 0$). The fact that multiple GBDPs produce an identical PDF for a phylogeny means that the GBDPs are not statistically identifiable by the phylogeny (see definition in §7 below). For example, the model of variable birth rate $\lambda(t)$ with a constant death rate μ and the model of variable birth rate $\lambda(t)$ with zero death rate $\mu = 0$ are two distinct models with different parameter values that are not identifiable using LTT data.

Louca & Pennell [8] considered GBDPs with both variable birth rate $\lambda(t)$ and variable death rate $\mu(t)$, and found that they produce the same likelihood on species divergence times as a GYP with a time-variable birth rate. Indeed there is a whole class of GBDPs with different time-dependent birth and death rates that have the same likelihood (are ‘congruent’ in the terminology of [8]) and are thus non-identifiable. Note that to demonstrate non-identifiability it is sufficient to show that any GBDP can be represented as a GYP. Suppose that the GBDP starts at time t_0 with a single lineage. The PDF of divergence times (i.e. the likelihood) under the GBDP is

$$f(\mathbf{t}, n | t_0) = p_1(t_0) \prod_{i=1}^{n-1} [\lambda(t_i) p_1(t_i)], \quad (6.1)$$

where $p_1(t)$ is given in [equation \(5.7\)](#). This is the same as the likelihood under the GYP with a time-dependent birth rate

$$\lambda_p(t) = \lambda(t)[1 - p_0(t)]. \quad (6.2)$$

Here $\lambda_p(t)$ is called the ‘pulled’ speciation rate in [8] and is a function of $\lambda(t)$, $\mu(t)$ and ρ in the original GBDP. Define

$$\Lambda_p(t) = \int_0^t \lambda_p(s) ds. \quad (6.3)$$

Then the equivalent GYP likelihood is

$$f_p(\mathbf{t}, n | t_0) = e^{-\Lambda_p(t_0)} \prod_{i=1}^{n-1} \lambda_p(t_i) e^{-\Lambda_p(t_i)} = e^{-\int_0^{t_0} \lambda(s)[1-p_0(s)] ds} \prod_{i=1}^{n-1} \lambda(t_i)[1-p_0(t_i)] e^{-\int_0^{t_i} \lambda(s)[1-p_0(s)] ds}. \quad (6.4)$$

7. Identifiability of birth–death processes

If the probability distributions of the data X are identical under model m with parameters θ and under model m' with parameters θ' , with

$$f(X|m, \theta) = f(X|m', \theta') \quad (7.1)$$

for essentially all possible data X , the models are non-identifiable by data X . Both the terms non-identifiability and unidentifiability are used in the literature and should be considered synonyms. Note that the object of identifiability is the model or parameters, the object of statistical inference. One may use the term *within-model non-identifiability* if $m = m'$ and $\theta \neq \theta'$, or *cross-model non-identifiability* if $m \neq m'$. In the former case, two sets of parameter values in the same model are non-identifiable, whereas in the latter two distinct models are non-identifiable. Non-identifiability reflects an essential indeterminacy in model or problem formulation [29, p. 54] and the definition concerns infinite data or all data (except for isolated data points that collectively have zero probability of occurrence, as implied in the term ‘essentially’ in equation (7.1)).

Some authors [30] define *practical non-identifiability* as ‘the case when distinct parameter combinations cannot be told apart from the limited number of observations available in practice’. Thus, practical non-identifiability simply means low information content in practical datasets. We avoid such terminology. Also, identifiability depends on the type of data; a model that is non-identifiable with data X may be identifiable with extended data X' . For example, models of between-species introgression are known to cause non-identifiability issues [31–33]. Introgression between sister lineages is non-identifiable using data of one sequence per species per locus but is identifiable when multiple samples are available per species [34,35]. Some models of gene flow are non-identifiable using gene tree topologies, but are identifiable using gene trees with branch lengths or sequence data [36].

Non-identifiability comes in many stripes (table 1). In some cases θ and θ' under the same model represent isolated points in the parameter space. One of the best known such cases is the non-identifiability of the label-switching type, as occurs in Bayesian clustering. Let data $X = \{x_i\}$ be a sample from a mixture of two normal distributions, $\mathcal{N}(\mu_1, \sigma_1^2)$ and $\mathcal{N}(\mu_2, \sigma_2^2)$, with the mixing proportions p_1 and $p_2 = 1 - p_1$. Let $\theta = (p_1, \mu_1, \sigma_1^2, \mu_2, \sigma_2^2)$ be the parameter vector. Then $\theta' = (p_2, \mu_2, \sigma_2^2, \mu_1, \sigma_1^2)$ will have exactly the same likelihood, with $f(X|\theta) = f(X|\theta')$ whatever data X are. Thus θ and θ' are non-identifiable. In effect, the labels ‘group 1’ and ‘group 2’ are switched between θ and θ' . This non-identifiability is known as a *label-switching* problem. Models with label-switching non-identifiability can still be used for inference. For example, a *relabelling algorithm* can be used to post-process the MCMC sample to fix the label-switching issue [38,39]. Concerning the linear BDP, several authors have observed that the likelihood of the times under a BDP conditioned on a fixed number of lineages is invariant when the birth rate λ and the death rate μ are swapped [24,25]. This non-identifiability due to the symmetry in λ and μ is thus of the label-switching type, even though the two models are very different biologically. Similar non-identifiability (both within-model and cross-model) of the label-switching type occurs in models of bidirectional introgression in analysis of multilocus genomic data under the multispecies coalescent model [40].

Another common type of non-identifiability occurs when the likelihood function depends on the parameters θ only through a function $h(\theta)$; in other words, the likelihood $L(\theta)$ can be written as $L(h(\theta))$, with $h(\theta)$ typically of a lower dimension than θ . Then θ and θ' are non-identifiable if $h(\theta) = h(\theta')$. Consider, for example, the estimation of divergence times $\mathbf{t} = (t_1, \dots, t_5)$ of figure 1 in a molecular clock dating analysis using a sequence alignment. There are six parameters in the model, $\theta = (\mu, t_1, \dots, t_5)$, where μ is the mutation rate. However, the likelihood for the sequence data depends on $h(\theta) = (t_1\mu, t_2\mu, t_3\mu, t_4\mu, t_5\mu)$, which has five dimensions. If we multiply μ by a factor c and divide all times by the same factor, $h(\theta)$ remains unchanged and all parameters generated through this transform are non-identifiable. There is a one-dimensional ridge on the likelihood surface along which the likelihood is constant.

Non-identifiability of certain parameters may occur when some other parameters take particular values. This is known as *local non-identifiability*. For example, in the clustering problem, if $p_2 = 0$, then μ_2 and σ_2^2 are non-identifiable. In the piecewise-constant model of variable speciation rate (§7b), if a time period has zero duration, the rate for that time period will be non-identifiable. Priors may be used to increase the separation between change points.

(a) Non-identifiability of a generalized birth–death process

Here we illustrate the non-identifiability of the GBDP as discussed in [8,14] using a simple example. We show that the constant-rate BDP with complete sampling $\rho = 1$ produces the same likelihood for LTT data as a pure-birth process with variable birth rate. As a result, the two models are non-identifiable.

Examining equations (6.1) and (6.4) we see that the two models are not identifiable if the following equality holds:

Table 1. Summary of different types of non-identifiability discussed in this article. BDP, birth–death process; GBDP, generalized birth–death process.

condition	identifiability	type	reference
<i>inference under BDPs</i>			
GBDP with continuous rate functions $\lambda(t)$, $\mu(t)$	non-identifiable	cross-model	[8]
GBDP with piecewise-constant rate functions $\lambda(t)$, $\mu(t)$	identifiable	cross-model	[27]
GBDP with piecewise-polynomial rate functions $\lambda(t)$, $\mu(t)$	identifiable	cross-model	[37]
BDP with constant rates λ , μ	$\theta = (\lambda, \mu)$ and $\theta' = (\mu, \lambda)$ are non-identifiable	within-model, label-switching type	[24,25]
<i>other examples discussed in this article</i>			
Bayesian clustering	parameters for different clusters are non-identifiable	within-model, label-switching type	[38,39]
molecular clock dating without fossil calibrations	times and rate are non-identifiable	within-model, dimension reduction	[15]

$$\lambda p_1(t) = \lambda_p(t) e^{-\Lambda_p(t)}. \quad (7.2)$$

On one hand, substituting equation (5.13) into the left hand side of equation (7.2) gives

$$\lambda p_1(t) = \frac{\lambda(\lambda - \mu)^2 e^{(\lambda - \mu)t}}{[\lambda e^{(\lambda - \mu)t} - \mu]^2}. \quad (7.3)$$

On the other hand, from the constant rate BDP we have

$$\Lambda_p(t) = \int_0^t \lambda[1 - p_0(s)] ds = (\lambda - \mu)t + \log\left(\frac{\lambda - e^{(\mu - \lambda)t}\mu}{\lambda - \mu}\right), \quad (7.4)$$

and by substituting equation (5.6) into equation (6.2) we obtain

$$\lambda_p(t) = \lambda[1 - p_0(t)] = \lambda \left(1 - \frac{\mu(e^{(\lambda - \mu)t} - 1)}{\lambda e^{(\lambda - \mu)t} - \mu}\right). \quad (7.5)$$

Substituting equations (7.4) and (7.5) into the right hand side of equation (7.2) gives

$$\lambda_p(t) e^{-\Lambda_p(t)} = \frac{\lambda(\lambda - \mu)^2 e^{(\lambda - \mu)t}}{[\mu - \lambda e^{(\lambda - \mu)t}]^2}. \quad (7.6)$$

Equation (7.3) and equation (7.5) are equal for any t, λ, μ . Thus, the likelihood under the constant-rate BDP (equation 6.1) and the likelihood under the pure-birth process with variable rate $\lambda_p(t)$ of equation (7.5) (equation 6.4) are equal. Note that both models are special cases of the GBDP, so that those two GBDP models are non-identifiable. Figure 2 shows the relationship between $\lambda_p(t)$ of the GYP (equation 7.5) and the constant rates λ and μ in the BDP.

(b) Identifiability of birth–death processes with piecewise-constant rates

The GBDP with arbitrary speciation and extinction rate functions appears to be too general to allow identifiability. Legried & Terhorst [27] considered a restricted class of GBDPs with rate functions $\lambda(t)$ and $\mu(t)$ to be piecewise-constant over K pieces (time periods). Such models are also known as *multiple-change-point models* in the statistics literature [41], with the boundaries between pieces of constant rate known as change points. Legried & Terhorst [27] showed that these models are identifiable as long as the number of tree tips is not too small relative to K . They further argue that with a large K these models can approximate any GBDP to an arbitrary degree, although this does not mean that real datasets contain information to recover arbitrary rate functions with any reliability (see below). Note that the result of §7a, that the BDP with constant rates λ and μ and the pure-birth process with variable rate $\lambda_p(t)$ are non-identifiable, is not in contradiction with Legried & Terhorst [27]. The variable-rate pure-birth process of equation (7.5) is not within the set of models considered by those authors, and there does not exist a piecewise-constant rate function for the pure-birth process that gives the same likelihood as the constant-rate BDP.

In simple cases (with certain regularity conditions [42] assumed), identifiability of parameters is indicated by the uniqueness of the maximum likelihood estimates and the non-singularity of Fisher's information matrix, evaluated at the true parameter value [42,43]. Here we use this idea to illustrate that the simplest member of this class of cladogenesis models, a Yule process with two birth rates on different intervals, is identifiable. For simplicity, we assume that the time of the rate change is specified.

Consider a pure-birth process with complete sampling that begins at time t_0 in the past with rate λ_0 and switches to rate λ_1 at time T . With $t > T$ assumed, equation (5.11) simplifies to

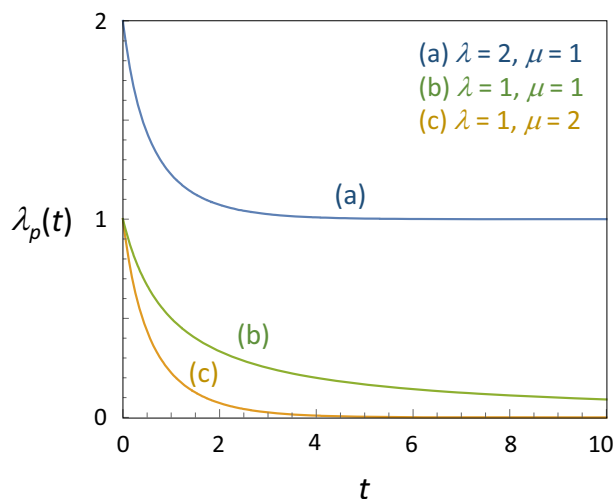


Figure 2. The ‘pulled speciation rate’, $\lambda_p(t)$, for the equivalent generalized Yule process of [8], which is non-identifiable with the birth–death process (BDP) model with constant birth and death rates (λ, μ) for three sets of parameter values. The BDP model with constant birth rate λ and constant death rate μ is non-identifiable, using lineage through time data, with the Yule process with variable birth rate $\lambda_p(t)$ given in equation (7.5).

$$p_1(t) = e^{-\int_0^t \lambda(s) ds} = e^{-[\lambda_0(t-T) + \lambda_1 T]} = e^{-\lambda_0(t-T)} e^{-\lambda_1 T}, \quad t > T. \quad (7.7)$$

This is also the probability that the lineage does not split over either time interval (t, T) or $(T, 0)$.

Now consider the reconstructed tree of n lineages where one lineage exists at time t_0 , I lineages were born during interval (t_0, T) and J were born during $(T, 0)$ (figure 3). Let $\mathbf{t}^0$ and $\mathbf{t}^1$ be vectors of times of speciation events occurring in the epochs (t_0, T) and $(T, 0)$, respectively. For simplicity, we assume that t_0 and T are known, and there are only two parameters to estimate in the model: λ_0, λ_1 . The likelihood for the observed times, according to equation (4.6), is

$$\begin{aligned} L(\lambda_0, \lambda_1) &= f(\mathbf{t} | t_0, n) \\ &= \prod_{i=1}^I i \lambda_0 e^{-\lambda_0(t_i^0 - T) - \lambda_1 T} \prod_{j=1}^J (I + j) \lambda_1 e^{-\lambda_1 t_j^1}, \\ &= (I + J)! \lambda_0^I e^{T I (\lambda_0 - \lambda_1)} e^{-\lambda_0 (\sum_{i=1}^I t_i^0)} \lambda_1^J e^{-\lambda_1 (\sum_{j=1}^J t_j^1)}, \end{aligned} \quad (7.8)$$

with the log-likelihood:

$$\log L = I \log \lambda_0 + J \log \lambda_1 + T I (\lambda_0 - \lambda_1) - \lambda_0 \sum_{i=1}^I t_i^0 - \lambda_1 \sum_{j=1}^J t_j^1. \quad (7.9)$$

The partial derivatives are

$$\frac{\partial \log L}{\partial \lambda_0} = \frac{I}{\lambda_0} + IT - \sum_{i=1}^I t_i^0, \quad \frac{\partial \log L}{\partial \lambda_1} = \frac{J}{\lambda_1} - IT - \sum_{j=1}^J t_j^1, \quad (7.10)$$

while the second derivatives are

$$\frac{\partial^2 \log L}{\partial \lambda_0^2} = -\frac{I}{\lambda_0^2}, \quad \frac{\partial^2 \log L}{\partial \lambda_1^2} = -\frac{J}{\lambda_1^2}. \quad (7.11)$$

The maximum likelihood estimates are thus

$$\hat{\lambda}_0 = \frac{I}{(\sum_{i=1}^I t_i^0) - IT}, \quad \hat{\lambda}_1 = \frac{J}{(\sum_{j=1}^J t_j^1) + IT}. \quad (7.12)$$

The information matrix is

$$F = - \begin{pmatrix} \frac{\partial^2 \log L}{\partial \lambda_0^2} & \frac{\partial^2 \log L}{\partial \lambda_0 \partial \lambda_1} \\ \frac{\partial^2 \log L}{\partial \lambda_1 \partial \lambda_0} & \frac{\partial^2 \log L}{\partial \lambda_1^2} \end{pmatrix} = \begin{pmatrix} \frac{I}{\lambda_0^2} & 0 \\ 0 & \frac{J}{\lambda_1^2} \end{pmatrix}. \quad (7.13)$$

The true parameter values are positive real numbers. If $I > 0, J > 0$, the information matrix will be non-singular and the parameters are identifiable. Note that here we have conditioned on the numbers of speciation times in each epoch: I, J . In the constant-rate Yule process described in §3, the number of times $(n - 1)$ is given by the number of tips n . Similarly, in the two-rate process $n = I + J + 1$.

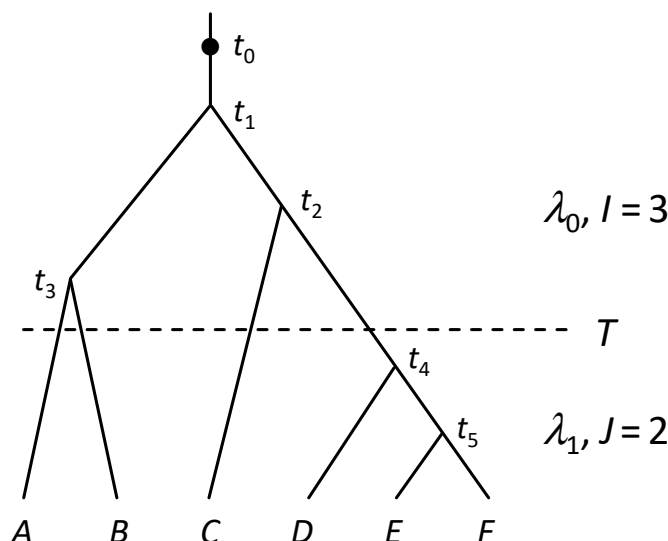


Figure 3. A piecewise-constant rate Yule model with two time periods of different rates. During the time period $(0, T)$, the birth rate is λ_1 and $J = 2$ speciation events are observed, whereas in the time period (T, t_0) , the birth rate is λ_0 and $I = 3$ speciation events are observed.

8. The limits of statistical information

The fact that the GBDP with piecewise-constant speciation and extinction rates is identifiable [27] may be encouraging for biologists interested in inferring speciation and extinction rates from reconstructed phylogenies. However, as noted by [27] identifiability does not ensure that information is available in real datasets to produce parameter estimates that are reliable and precise enough to draw useful biological conclusions. Simulations and other exploratory analyses may be used to determine how informative typical datasets are. We note that when the speciation rate and extinction rate vary over time, there is a general difficulty of small sample sizes. The case is similar to a coin-tossing experiment in which a sequence of coins with different probabilities for heads are flipped, each only once or a very few times, with the aim of estimating the probabilities for all coins. Furthermore, most studies applying the GBDP to reconstructed phylogenies (the LTT data) have assumed that the true divergence times are known, whereas in practice estimated divergence times involve considerable errors and uncertainties owing to the challenges of summarizing the fossil evidence, estimating branch lengths on molecular phylogenies, and the confounding effects of rates and times in analysis of molecular data (see [15] for a review).

Here we conduct a simulation experiment under a simple scenario with only one rate change to explore the information content in LTT data. We simulate divergence times with an extension of the two-rate Yule process model described in §7b above, and ask whether the change in speciation rate at time T can be recovered using a piecewise-constant Yule process model. We extend the model slightly by assuming that at the end of the first epoch (at time T), mass extinction occurs with the probability of any lineage surviving to be p while speciation rate is much higher after the mass extinction than before, with $\lambda_1 \gg \lambda_0$. This mimics a mass-extinction scenario such as occurs in the fossil record at the Cretaceous–Palaeogene extinction event (K–Pg boundary), where extinction opens up new niches previously occupied by other species, increasing the speciation rate. However, many lineages existing before the extinction event will leave no descendants, so that it may not be possible to estimate the speciation rate for those lineages (λ_0) reliably.

For simplicity, we assume that the time of origin (t_0), the survival probability at extinction (p), the time of the mass extinction (T) and the numbers of speciation events in the two time epochs (I, J) are all known, so that there are only two parameters in the model: λ_0 and λ_1 . If a lineage arises in the first epoch at time $t > T$ the probability that it leaves exactly one descendant at present is given by averaging over the number of lineages (m) right before the mass extinction (with probability $p_m(t)$ of equation (3.7)), with only one of them surviving the mass extinction and leaving one descendant in the sample:

$$p_1^*(t | \lambda_0, \lambda_1, T, p) = \sum_{m=1}^{\infty} e^{-m\lambda_0(t-T)} (e^{\lambda_0(t-T)} - 1)^{m-1} \cdot m p (1-p)^{m-1} \cdot e^{-\lambda_1 T} = \frac{p e^{\lambda_0(t+T) - \lambda_1 T}}{[e^{\lambda_0 T} (p-1) - e^{\lambda_0 t} p]^2}. \quad (8.1)$$

Note that

$$\lim_{p \rightarrow 1} p_1^*(t | \lambda_0 = \lambda, \lambda_1 = \lambda, T, p) = e^{-\lambda t}. \quad (8.2)$$

The likelihood is then

$$L(\lambda_0, \lambda_1) = \prod_{i=1}^I \lambda_0 i p_1^*(t_i^0 | \lambda_0, \lambda_1, T, p) \prod_{j=1}^J \lambda_1 (I+j) e^{-\lambda_1 t_j^1}. \quad (8.3)$$

Note that the limit of equation (8.3) as $p \rightarrow 1$ is equation (7.8) as expected. The log-likelihood is

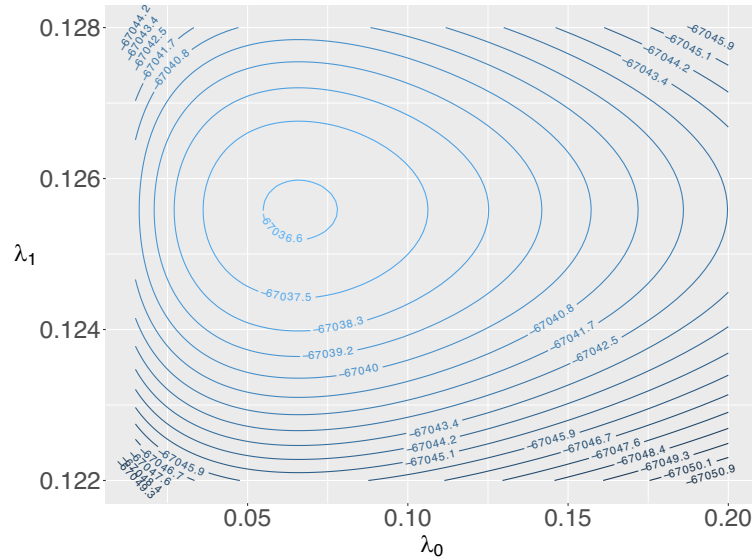


Figure 4. Log-likelihood surface for a realization of the two-rate epoch Yule process with a mass extinction event between epochs (figure 3). Time is given in millions of years and the true rates are $\lambda_0 = 0.025$, $\lambda_1 = 0.125$. LTT data are simulated assuming a mass extinction event at $T = 66$ Ma, with the survival probability at the extinction to be $p = 0.25$, and the process originated at $t_0 = 170$ Ma. Note the very different scales for the λ_0 and λ_1 axes.

$$\begin{aligned} \log L &= I \log \lambda_0 + \sum_{i=1}^I \log[p_1^*(t_i^0 | \lambda_0, \lambda_1, T, p)] + J \log \lambda_1 - \lambda_1 \sum_{j=1}^J t_j^1, \\ &= I \log \lambda_0 + \sum_{i=1}^I \log \left[\frac{p e^{\lambda_0(t_i^0 + T) - \lambda_1 T}}{[e^{\lambda_0 T}(p-1) - e^{\lambda_0 t_i^0} p]^2} \right] + J \log \lambda_1 - \lambda_1 \sum_{j=1}^J t_j^1. \end{aligned} \quad (8.4)$$

We simulated a dataset using $\lambda_0 = 0.025$, $\lambda_1 = 0.125$, $t_0 = 170$, $T = 66$ and $p = 0.25$. In this case, the process starts at 170 Ma with a speciation rate of 1 speciation per 40 Myr, and a mass extinction occurs at 66 Ma that caused 75% of species to become extinct (similar to the K-Pg extinction). Following this event, the speciation rate increases fivefold to 1.25 speciations per 10 Myr. The simulated data had $I = 3$ and $J = 21\,798$, so the number of species in the reconstructed tree is 21 801, larger than many empirical datasets of this type.

A contour plot of the log-likelihood surface is shown in figure 4. Here we used the correct model with the log-likelihood given by equation (8.4), with $p = 0.25$ and $T = 66$ set to their true values. Although λ_1 is relatively precisely estimated there is a great deal of uncertainty about the speciation rate λ_0 before the mass extinction and it is difficult to infer a change in speciation rate after the extinction event from this dataset. Thus, even for very large datasets, it can be very challenging to identify even a large change in speciation rate between epochs. The two rates are identifiable in this case, but the information content in typical datasets may be very limited. Note that we have chosen this example simply to illustrate the potential effects of mass extinctions on rate inference; obviously a much more extensive simulation study would be needed to reach any general conclusions about the power of inference that is expected in real data analyses.

May *et al.* [44] developed a Bayesian inference model with both temporal shifts of speciation and extinction rates and periodic mass extinction events. Although they were specifically interested in detecting mass extinction events and estimating their time of occurrence (rather than rates before and after such events) they did find from simulations that the power to detect mass extinction events depends on how early they occur in relation to the phylogenetic tree. Ancient events near the root are more difficult to detect: ‘a mass-extinction event that occurs too close to the root of the study tree will be difficult to detect because too few lineages will have participated in that event’ [44, p.955]. This is similar to the situation concerning the ancestral rate (λ_0) in our example.

9. Discussion

Recent analyses suggest that the GBDP with arbitrary time-dependent speciation and extinction rates is not identifiable using LTT data, while the model becomes identifiable if the rate function is constrained to be piecewise-constant. We review those theoretical results, and discuss their implications to inference using empirical phylogenies. We demonstrate that in at least some cases even large phylogenies with thousands of species may not contain enough information to make useful inferences of past speciation and extinction rates.

Several recent papers have responded directly to the non-identifiability result of [8], even suggesting possible solutions. Morlon *et al.* [30] suggested the use of informative priors on rates or other sources of biological information to reduce the space of congruent models. Other authors [45] analysed empirical and simulated datasets to examine whether trends such as increasing rates of net cladogenesis can be identified, assuming that the congruent class of non-identifiable models has been found and relying on programs such as CRABS [46] to enumerate all such models. Extinction rates are particularly difficult to infer and one suggestion was to use a GYP to detect general trends of increasing or decreasing rates, again using the CRABS

program [46] to identify congruent models of speciation and extinction rates. Kopperud *et al.* [45] found that many proposed congruent models were invalid (e.g. with negative extinction rates), so the space of valid models may be much smaller. They pointed out that statistical uncertainty about rates may be as important as unidentifiability of congruence classes.

The findings discussed in §7b that piecewise-constant [27] (and piecewise-polynomial, [37]) time-dependent rate functions are identifiable appear to offer hopes for evolutionary biologists interested in using timetrees to infer variable speciation and extinction rates. Although identifiability with infinite data is reassuring, the information in finite datasets about the rates during particular epochs may be very weak. To account for this, inference methods should be explicit about uncertainty of parameters and numbers of rate epochs. Bayesian methods with diffuse priors may prove more reliable than likelihood in reflecting the true levels of uncertainty in such cases. Another solution is to use simplified models that reduce the numbers of intervals with distinct rates, introduce autocorrelation, or reduce the number of parameters that need to be estimated from the data. Another venue for exploration is GBDP in which samples are collected through time (rather than all at the present time). Fossils add useful information and may make the model identifiable. A recent study [47] claims that the fossilized GBDP is identifiable.

Some readers may be surprised by the poor information content about the ancestral speciation rate before a mass extinction in our example in §8. However, in the simulation only four lineages existed prior to the extinction that are ancestral to all contemporary species, so there is clearly very limited information in this case. Some studies examining the power to infer ancient rate changes have produced more optimistic results. Stadler [48] developed a likelihood inference method for a BDP with piecewise-constant rates with a fixed number of rate shifts, attempting to infer both the number and time points of rate shifts and the rates on each interval. A small simulation with a single rate shift suggested that the method could accurately identify both the rates and the time of the rate shift. The simulations considered only very low extinction rates (up to one-quarter of the speciation rate in one simulation and three orders of magnitude lower extinction rates than the speciation rates in another). For mass extinction events one expects the extinction rate to be potentially orders of magnitude greater than the speciation rate, and a high extinction rate during a mass extinction will greatly reduce the number of ancestral lineages available for inferring rates from a reconstructed tree. Similarly, there is typically more power to detect recent than ancient diversification rate shifts (as found in [48] for mammalian diversification after the K-Pg mass extinction). More studies are needed to examine the performance of inference with extinction rates similar to or higher than speciation rates. The effect of mass extinctions in reducing information about past diversification rates shares many similarities with the related problem of inferring ancestral population sizes from genetic data in the presence of a population bottleneck [49,50]. In both cases, low numbers of lineages surviving the event lead to loss of information about rates prior to the event.

We have focused on the identifiability and inference properties of the GBDP with rate variation through time. Models of among-lineage rate variation, often referred to as *state-dependent extinction models* (SSEs) [13], are another generalization of the GBDP, so identifiability problems are likely to exist for such models as well. The difficulty of estimating parameters under SSE models has already been a subject of debate, although some of the problems have been due to implementations [7]. Existing inference methods invoking different assumptions are noted to produce very different results [51]. In some models, rates are positive random variables drawn from a continuous distribution at rate-change events, while in others there exist a fixed set of rates. Dragomir *et al.* [52] studied a multitype BDP (a particular type of SSE) with rates changing at random times but drawn from a fixed set of rates and proved that parameters under the model are identifiable. Recently, several contradictory studies have been posted that claim to prove either identifiability [53] or non-identifiability of general SSEs [54]. Hopefully, definitive answers will emerge soon. In conclusion, the problem of inferring rates of speciation and extinction using stochastic birth–death models, a legacy of Udny Yule’s [5] model in the 1920s, continues to provide rich challenges and unanswered questions for biologists and mathematicians in the twenty-first century.

Ethics. This work did not require ethical approval from a human subject or animal welfare committee.

Data accessibility. This article has no additional data.

Declaration of AI use. We have not used AI-assisted technologies in creating this article.

Authors’ contributions. B.R.: conceptualization, funding acquisition, investigation, methodology, project administration, software, writing—original draft, writing—review and editing; Z.Y.: conceptualization, funding acquisition, investigation, methodology, project administration, writing—original draft, writing—review and editing.

Both authors gave final approval for publication and agreed to be held accountable for the work performed herein.

Conflict of interest declaration. We declare we have no competing interests.

Funding. This study has been supported by an NIH Grant (GM123306) to B.R., and a Biotechnology and Biological Sciences Research Council grant (BB/Y004132/1) and Templeton Foundation grant to Z.Y.

References

1. Mindell DP. 2013 The tree of life: metaphor, model, and heuristic device. *Syst. Biol.* **62**, 479–489. (doi:10.1093/sysbio/sys115)
2. Baum DA, Smith SD. 2012 *Tree Thinking: an Introduction to Phylogenetic Biology*. Greenwood Village, CO: Roberts Company Publishers.
3. Taylor SA, Larson EL. 2019 Insights from genomes into the evolutionary importance and prevalence of hybridization in nature. *Nat. Ecol. Evol.* **3**, 170–177. (doi:10.1038/s41559-018-0777-y)
4. Ricklefs RE. 2007 Estimating diversification rates from phylogenetic information. *Trends Ecol. Evol.* **22**, 601–610. (doi:10.1016/j.tree.2007.06.013)
5. Yule GU. 1925 II.—A mathematical theory of evolution, based on the conclusions of Dr. J. C. Willis, FRS. *Phil. Trans. R. Soc. Lond. B* **213**, 21–87. (doi:10.1098/rstb.1925.0002)
6. Morlon H. 2014 Phylogenetic approaches for studying diversification. *Ecol. Lett.* **17**, 508–525. (doi:10.1111/ele.12251)

7. Moore BR, Höhna S, May MR, Rannala B, Huelsenbeck JP. 2016 Critically evaluating the theory and performance of Bayesian analysis of macroevolutionary mixtures. *Proc. Natl Acad. Sci. USA* **113**, 9569–9574. (doi:10.1073/pnas.1518659113)
8. Louca S, Pennell MW. 2020 Extant timetrees are consistent with a myriad of diversification histories. *Nature* **580**, 502–505. (doi:10.1038/s41586-020-2176-1)
9. Helmstetter AJ *et al.* 2022 Pulled diversification rates, lineages-through-time plots, and modern macroevolutionary modeling. *Syst. Biol.* **71**, 758–773. (doi:10.1093/sysbio/syab083)
10. Kendall DG. 1948 On the generalized “birth-and-death” process. *Ann. Math. Stat.* **19**, 1–15. (doi:10.1214/aoms/1177730285)
11. Nee S, May RM, Harvey PH. 1994 The reconstructed evolutionary process. *Phil. Trans. R. Soc. Lond. B* **344**, 305–311. (doi:10.1098/rstb.1994.0068)
12. Sanderson MJ, Donoghue MJ. 1996 Reconstructing shifts in diversification rates on phylogenetic trees. *Trends Ecol. Evol.* **11**, 15–20. (doi:10.1016/0169-5347(96)81059-7)
13. Maddison WP, Midford PE, Otto SP. 2007 Estimating a binary character’s effect on speciation and extinction. *Syst. Biol.* **56**, 701–710. (doi:10.1080/10635150701607033)
14. Kubo T, Iwasa Y. 1995 Inferring the rates of branching and extinction from molecular phylogenies. *Evolution* **49**, 694–704. (doi:10.1111/j.1558-5646.1995.tb02306.x)
15. dos Reis M, Donoghue PCJ, Yang Z. 2016 Bayesian molecular clock dating of species divergences in the genomics era. *Nat. Rev. Genet.* **17**, 71–80. (doi:10.1038/nrg.2015.8)
16. Rannala B, Yang Z *et al.* 2013 Molecular clock dating. In *The Princeton guide to evolution* (ed. JB Losos *et al.*), pp. 67–74. Princeton, NJ: Princeton University Press. (doi:10.1515/9781400848065-010)
17. Angelis K, Dos Reis M. 2015 The impact of ancestral population size and incomplete lineage sorting on Bayesian estimation of species divergence times. *Curr. Zool.* **61**, 874–885. (doi:10.1093/czoolo/61.5.874)
18. Rannala B, Edwards SV, Leaché A, Yang Z. 2020 The multi-species coalescent model and species tree inference. In *Phylogenetics in the genomic era* (eds C Scornavacca, F Delsuc, N Galtier), pp. 3.3:1–3.3:21. See <https://hal.inria.fr/PGE>.
19. Rannala B, Yang Z. 1996 Probability distribution of molecular evolutionary trees: a new method of phylogenetic inference. *J. Mol. Evol.* **43**, 304–311. (doi:10.1007/pl00006090)
20. Huelsenbeck JP, Rannala B, Masly JP. 2000 Accommodating phylogenetic uncertainty in evolutionary studies. *Science* **288**, 2349–2350. (doi:10.1126/science.288.5475.2349)
21. Pagel M, Meade A, Barker D. 2004 Bayesian estimation of ancestral character states on phylogenies. *Syst. Biol.* **53**, 673–684. (doi:10.1080/10635150490522232)
22. Edwards AWF. 1970 Estimation of the branch points of a branching diffusion process. *J. R. Stat. Soc. Ser. B* **32**, 155–164. (doi:10.1111/j.2517-6161.1970.tb00828.x)
23. Dickey EH, Rosenberg NA. 2025 Labelled histories with multifurcation and simultaneity. *Phil. Trans. R. Soc. B* **380**, 20230307. (doi:10.1098/rstb.2023.0307)
24. Rannala B. 1997 Gene genealogy in a population of variable size. *Heredity* **78**, 417–423. (doi:10.1038/sj.hdy.6881060)
25. Stadler T, Steel M. 2019 Swapping birth and death: symmetries and transformations in phylodynamic models. *Syst. Biol.* **68**, 852–858. (doi:10.1093/sysbio/syz039)
26. Rannala B, Huelsenbeck JP, Yang Z, Nielsen R. 1998 Taxon sampling and the accuracy of large phylogenies. *Syst. Biol.* **47**, 702–710. (doi:10.1080/106351598260680)
27. Legried B, Terhorst J. 2022 A class of identifiable phylogenetic birth–death models. *Proc. Natl Acad. Sci. USA* **119**, e2119513119. (doi:10.1073/pnas.2119513119)
28. Morlon H, Parsons TL, Plotkin JB. 2011 Reconciling molecular phylogenies with the fossil record. *Proc. Natl Acad. Sci. USA* **108**, 16327–16332. (doi:10.1073/pnas.1102543108)
29. Stuart A, Ord K, Arnold S. 1999 *Kendall’s advanced theory of statistics*, 6th edn. London, UK: Edward Arnold.
30. Morlon H, Robin S, Hartig F. 2022 Studying speciation and extinction dynamics from phylogenies: addressing identifiability issues. *Trends Ecol. Evol.* **37**, 497–506. (doi:10.1016/j.tree.2022.02.004)
31. Yu Y, Degnan JH, Nakhleh L. 2012 The probability of a gene tree topology within a phylogenetic network with applications to hybridization detection. *PLoS Genet.* **8**, e1002660. (doi:10.1371/journal.pgen.1002660)
32. Pardi F, Scornavacca C. 2015 Reconstructible phylogenetic networks: do not distinguish the indistinguishable. *PLoS Comput. Biol.* **11**, e1004135. (doi:10.1371/journal.pcbi.1004135)
33. Zhu S, Degnan JH. 2016 Displayed trees do not determine distinguishability under the network multispecies coalescent. *Syst. Biol.* **66**, 283–298. (doi:10.1093/sysbio/syy097)
34. Degnan JH. 2018 Modeling hybridization under the network multispecies coalescent. *Syst. Biol.* **67**, 786–799. (doi:10.1093/sysbio/syy040)
35. Jiao X, Flouri T, Yang Z. 2021 Multispecies coalescent and its applications to infer species phylogenies and cross-species gene flow. *Natl Sci. Rev.* **8**, nwab127. (doi:10.1093/nsr/nwab127)
36. Pang XX, Zhang DY. 2024 Detection of ghost introgression requires exploiting topological and branch length information. *Syst. Biol.* **73**, 207–222. (doi:10.1093/sysbio/syad077)
37. Legried B, Terhorst J. 2023 Identifiability and inference of phylogenetic birth–death models. *J. Theor. Biol.* **568**, 111520. (doi:10.1016/j.jtbi.2023.111520)
38. Celeux G, Hurm M, Robert CP. 2000 Computational and inferential difficulties with mixture posterior distributions. *J. Am. Stat. Assoc.* **95**, 957. (doi:10.2307/2669477)
39. Stephens M. 2000 Dealing with label switching in mixture models. *J. R. Stat. Soc. Ser. B* **62**, 795–809. (doi:10.1111/1467-9868.00265)
40. Yang Z, Flouri T. 2022 Estimation of cross-species introgression rates using genomic data despite model unidentifiability. *Mol. Biol. Evol.* **39**, 10. (doi:10.1093/molbev/msac083)
41. Green PJ. 1995 Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* **82**, 71132. (doi:10.2307/2337340)
42. Bowden R. 1973 The theory of parametric identification. *Econometrica* **41**, 1069–1074. (doi:10.2307/1914036)
43. Rothenberg TJ. 1971 Identification in parametric models. *Econometrica* **39**, 577–591. (doi:10.2307/1913267)
44. May MR, Höhna S, Moore BR. 2016 A Bayesian approach for detecting the impact of mass-extinction events on molecular phylogenies when rates of lineage diversification may vary. *Methods Ecol. Evol.* **7**, 947–959. (doi:10.1111/2041-210X.12563)
45. Kopperud BT, Magee AF, Höhna S. 2023 Rapidly changing speciation and extinction rates can be inferred in spite of nonidentifiability. *Proc. Natl Acad. Sci.* **120**, e2208851120. (doi:10.1073/pnas.2208851120)
46. Höhna S, Kopperud BT, Magee AF. 2022 CRABS: congruent rate analyses in birth–death scenarios. *Methods Ecol. Evol.* **13**, 2709–2718. (doi:10.1111/2041-210X.13997)
47. Truman K, Vaughan TG, Gavryushkina A, Gavryushkina AS. 2024 The Fossilised Birth-Death Model is Identifiable. *Syst. Biol.* **2024**, syae058. (doi:10.1093/sysbio/syae058)
48. Stadler T. 2011 Mammalian phylogeny reveals recent diversification rate shifts. *Proc. Natl Acad. Sci. USA* **108**, 6187–6192. (doi:10.1073/pnas.1016876108)
49. Terhorst J, Song YS. 2015 Fundamental limits on the accuracy of demographic inference based on the sample frequency spectrum. *Proc. Natl Acad. Sci. USA* **112**, 7677–7682. (doi:10.1073/pnas.1503717112)
50. Baharian S, Gravel S. 2018 On the decidability of population size histories from finite allele frequency spectra. *Theor. Popul. Biol.* **120**, 42–51. (doi:10.1016/j.tpb.2017.12.008)
51. Martínez-Gómez J, Song MJ, Tribble CM, Kopperud BT, Freyman WA, Höhna S, Specht CD, Rothfels CJ. 2024 Commonly used Bayesian diversification methods lead to biologically meaningful differences in branch-specific rates on empirical phylogenies. *Evol. Lett.* **8**, 189–199. (doi:10.1093/evlett/qrada044)
52. Dragomir D, Allman ES, Rhodes JA. 2023 Parameter identifiability of a multitype pure-birth model of speciation. *J. Comput. Biol.* **30**, 277–292. (doi:10.1089/cmb.2022.0330)
53. O’Meara B, Beaulieu J. 2021 Potential survival of some, but not all, diversification methods. *EcoEvoRxiv*. (doi:10.32942/osf.io/w5nvd)
54. Tarasov S, Uyeda J. 2024 Nonidentifiability of state-dependent diversification models (SSEs) is ubiquitous but not problematic for phylogenetics. *bioRxiv*. (doi:10.1101/2022.07.04.498736)