# The Fossilised Birth-Death Model is Identifiable

Kate Truman[1,2], Timothy G Vaughan[3,4], Alex Gavryushkin[1,2] and Alexandra "Sasha" Gavryushkina[1,2,*]

[1] *Biological Data Science Laboratory, School of Mathematics and Statistics, University of Canterbury, Christchurch, New Zealand*
[2] *Biomathematics Research Centre, University of Canterbury, Christchurch, New Zealand*
[3] *Department of Biosystems Science and Engineering, ETH Zurich, Basel 4056, Switzerland*
[4] *Swiss Institute of Bioinformatics, Lausanne 1015, Switzerland*
∗ *sasha.gavryushkina@canterbury.ac.nz*

August 29, 2024

**Abstract**

Time-dependent birth-death sampling models have been used in numerous studies for inferring past evolutionary dynamics in different areas, e.g. speciation and extinction rates in macroevolutionary studies, or effective reproductive number in epidemiological studies. These models are branching processes where lineages can bifurcate, die, or be sampled with time-dependent birth, death, and sampling rates, generating phylogenetic trees. It has been shown that in some subclasses of such models, different sets of rates can result in the same distributions of reconstructed phylogenetic trees, and therefore the rates become unidentifiable from the trees regardless of their size. Here we show that widely used time-dependent fossilised birth-death (FBD) models are identifiable. This subclass of models makes more realistic assumptions about the fossilisation process and certain infectious disease transmission processes than the unidentifiable birth-death sampling models. Namely, FBD models assume that sampled lineages stay in the process rather than being immediately removed upon sampling. Identifiability of the time-dependent FBD model justifies using statistical methods that implement this model to infer the underlying temporal diversification or epidemiological dynamics from phylogenetic trees or directly from molecular or other comparative data. We further show that the time-dependent fossilised-birth-death model with an extra parameter, the removal after sampling probability, is unidentifiable. This implies that in scenarios where we do not know how sampling affects lineages we are unable to infer this extra parameter together with birth, death, and sampling rates solely from trees.

**Keywords**

Birth-death sampling models | Phylogenetics | Fossils | Speciation | Extinction

# 1 Introduction

Birth-death models have been widely used to model different processes that are described by phylogenetic trees, from evolutionary histories of species (Raup et al., 1973; Nee et al., 1995; Nee, 2006; Heath et al., 2014; Silvestro et al., 2014; Morlon et al., 2011) to infectious disease transmission histories (Stadler et al., 2012, 2013). A birth-death process describes how lineages bifurcate and die with corresponding birth and death rates, and results in a phylogenetic tree (Feller, 1939; Kendall, 1948). Given molecular or other data of sampled lineages one can reconstruct a phylogenetic tree, and given the tree infer the rates, or infer trees and rates in a joint analysis (Yang and Rannala, 1997; Höhna et al., 2016; Bouckaert et al., 2019). The birth and death rates or their transformations are then interpreted as parameters of the underlying processes, e.g. speciation and extinction rates in macroevolutionary studies or effective reproductive number in epidemiological applications.

An important aspect of modelling the process that produces phylogenetic trees is the fact that such trees are reconstructed from a sample of extant species and do not represent all lineages that are involved in the evolutionary process. Extinct or unsampled lineages remain unobserved. To model present sampling, additionally to the birth and death rates we have a sampling fraction parameter (Stadler, 2009). Furthermore, we consider reconstructed trees that represent the observed part of the evolutionary history (Nee et al., 1995), that is, a phylogeny that relates sampled extant species only.

Temporal changes in the rates of species diversification or infectious disease transmission have been accounted for by time-dependent birth and death rates (Kendall, 1948; Stadler et al., 2013). The probability density functions for reconstructed trees, in this case, were derived for the simple case of piecewise constant rates (Stadler, 2011) or more general case with arbitrary time-dependent rates (Morlon et al., 2011), which enabled efficient inference of the speciation or transmission temporal dynamics (Stadler et al., 2013; Louca and Pennell, 2020a). Importantly, such inference relies on statistical identifiability of the time-dependent birth-death models, which is one of the conditions that guarantee that the estimated rates approach their true values as the tree size increases.

Kubo and Iwasa (1995) initially showed that time-dependent birth-death models are unidentifiable for reconstructed trees, that is, multiple pairs of birth and death rates have the same probability of producing a reconstructed tree. Despite the age of Kubo and Iwasa's result, and the continuing popularity of these models, widespread concern (Pagel, 2020) about this unidentifiability issue was only recently prompted by the work of Louca and Pennell (2020b). The major implication of the unidentifiability is that we cannot estimate the two rate functions without placing additional constraints (Morlon et al., 2022) on

one of them. Crucially, it has been shown that the number of congruent pairs of birth and death rates, — the rates that produce the same distributions of reconstructed trees, in the absence of such constraints is infinite (Kubo and Iwasa, 1995; Louca and Pennell, 2020b). This rules out manual parameter selection from the class of equally likely parameter pairs as a mitigating strategy.

The unidentifiability of the time-dependent birth-death models have been at least partially alleviated by Legried and Terhorst's findings that piecewise constant, and subsequently, piecewise polynomial birth-death models, are identifiable for reconstructed trees (Legried and Terhorst, 2022, 2023). However, no such result for other classes of models such as those involving exponential growth or decay has been obtained.

In many applications, samples are obtained at different points in time, for example, when fossils are included in the analysis or when heterochronous sampling of pathogen genomic material is employed. Modelling such sampling through time is very important to avoid bias in the rate estimates (Stadler, 2010). For this reason, an additional parameter, sampling rate, was added to the birth-death process, introducing a birth-death sampling process (Morlon et al., 2011; Stadler, 2010; Stadler et al., 2013; Gavryushkina et al., 2014; MacPherson et al., 2021). Early applications of this process to inference assumed that lineages are removed from the process immediately after sampling, which might be the case for some infectious diseases (e.g. HIV), where the diagnosis typically prevents a patient from transmitting the disease further. Louca et al. (2021) showed that this time-dependent birth-death sampling model is again unidentifiable when we consider reconstructed trees.

Removing lineages immediately after sampling is a strong and unrealistic assumption (Andréoletti and Morlon, 2023) for many other processes, e.g. fossilisation or transmission process for highly transmittable diseases where isolation is not performed on the diagnosis, such as influenza or coronavirus. A birth-death sampling model where sampling does not imply removing the lineage from the process is called the fossilised birth-death (FBD) model (Stadler, 2010; Heath et al., 2014). This model is widely used to estimate fossilisation rates and date species phylogenies (Stadler et al., 2018; Wright et al., 2022), however, the identifiability of the time-dependent rates for the FBD model has not been established (Andréoletti and Morlon, 2023).

In this paper, we show that the time-dependent FBD model is identifiable for arbitrary rate functions with strictly positive sampling rate. Although this does not imply that it is practically possible to estimate arbitrary rates from finite phylogenies, it provides a strong theoretical basis for using the FBD model to estimate speciation, extinction, and sampling parameters from phylogenies, and eliminates the requirement that the rates are restricted to be piecewise-polynomial. We demonstrate an important relationship between the distributions of complete and reconstructed trees under the time-dependent FBD model. We also show that the time-dependent birth-death sampling model (MacPherson et al., 2021), in which the removal probability is another unknown time-dependent parameter, is unidentifiable. That is, if the status of a sample after sampling is unknown and likely changes with time, we are again unable to estimate the

birth and death rates.

Further, we use normalised deterministic lineages through time (nLTT) curves to demonstrate how the number of lineages in the reconstructed trees is affected by allowing sampled ancestors, that is, allowing species sampled in the past to be ancestors of younger species, which is the case when sampling does not imply removal. Finally, we discuss the implications of our results for data analysis.

# 2 Materials & Methods

## 2.1 Preliminaries

Here we consider a time-dependent fossilised birth-death (FBD) process (Stadler, 2010; MacPherson et al., 2021; Heath et al., 2014; Nee et al., 1994). The time in this process is usually set backwards, and the process runs from the time of origin $t_0 > 0$ until the present (time zero). It starts with a single lineage, and all lineages arising in the process independently bifurcate, die, and/or are sampled through time with time-dependent Poisson rates $\lambda(t)$, $\mu(t)$, and $\psi(t)$. These rates are lineage-independent, i.e. all lineages at the same time point will have the same birth, death, and sampling rates. At the present time, extant lineages are additionally sampled uniformly at random with probability $\rho_0$, called the *sampling fraction*. This time-dependent FBD process is a special case of the time-dependent birth-death sampling process (MacPherson et al., 2021) in which there is an extra parameter: the *removal probability*, which is a probability of a lineage being removed from the process immediately after sampling. In the FBD process this probability is set to zero, reflecting that sampling does not influence whether lineages remain in the process.

All lineages created by such a process together with the marks for the sampling events form a phylogenetic tree, which we call a *complete* tree $\mathcal{T}$ (Fig. 1). To write down the probability density of complete trees under the FBD process we need to introduce the following notation, that we keep consistent with MacPherson et al. (MacPherson et al., 2021). We assume that $\rho_0 = 1$ when considering complete trees. Let $N_0$ denote the number of present day taxa, each of which is sampled at present. The tree has $n$ death events at times $(y_1, y_2, \dots y_n)$, $N_0 + n - 1$ branching events at times $(x_1, x_2, \dots x_{N_0+n-1})$, and $m$ sampling events at times $(z_1, z_2, \dots z_m)$. The first branching event at $x_1$ is called the root of the tree. A tree $\mathcal{T}$ can be split into its discrete and continuous components and can be written as a pair $(T, \bar{t})$ where $T$ is the ranked ordered tree topology (Gavryushkin and Drummond, 2016; Steel, 2016; Gavryushkin et al., 2018), and $\bar{t} = (t_1, \dots, t_{N_0+2n+m-1})$ is the vector of times of branching, death, and sampling-through-time events with $t_0 > t_1 > \dots > t_{N_0+2n+m-1} > 0$.

In what follows, we assume that $\lambda(t) \geq 0$, $\mu(t) \geq 0$, and $\psi(t) \geq 0$ for all $t \in [0, t_0]$ and are Lebesgue-integrable on $[0, t_0]$. Following the well described derivation (see, e.g. Stadler, 2010; Morlon et al., 2011; MacPherson et al., 2021), we obtain the probability density function for the complete trees under the

time-dependent FBD process:

$$f^{compl.}\left(\mathcal{T}|\lambda(t),\mu(t),\psi(t),t_0\right)=\Psi(t_0)\prod_{i=1}^{N_0+n-1}\lambda\left(x_i\right)\Psi(x_i)\prod_{j=1}^{n}\frac{\mu\left(y_j\right)}{\Psi(y_j)}\prod_{k=1}^{m}\psi(z_k) \tag{1}$$

where $\Psi(t)=e^{-\int_0^t \lambda(u)+\mu(u)+\psi(u)\,du}$. Each node in the tree contributes to the probability density function as the rate of the event (birth, death, or sampling) at the corresponding time, and each edge contributes the exponential term, $\Psi\left(t\right)$, reflecting that no events took place along them. SI Figure S1 illustrates the parts of a tree that correspond to the exponential of the integral terms.

Note that it is common in the literature to use the term "likelihood" for the probability density function even if it is not considered as a function of parameters. In this paper, we prefer to use the term "probability density function".

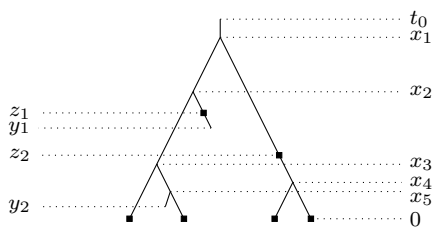
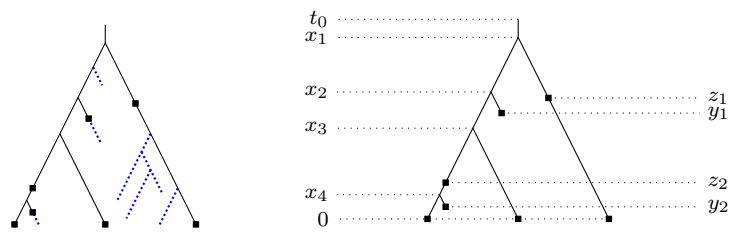



Figure 1: An example of a complete tree generated by a time-dependent Fossilised Birth-Death (FBD) process. The stem of the tree is denoted $t_0$, and the present occurs at time 0. Birth events occur at times $x_i$, death events at times $y_j$, and samples at times $z_k$.

In a complete tree, all speciation, extinction, and sampling events (through time and at present) are observed. However, in practice we are not able to observe (or infer) the complete tree, and therefore parameter estimation relies on the *reconstructed* tree. The reconstructed tree is a tree induced by the sampled nodes and is shown in Figure 2b given the complete tree in Figure 2a.

We now introduce a notation for reconstructed trees which is similar to that used in the complete tree case. The sampling at present probability $\rho_0$ can now take any value between zero and one. The reconstructed tree has $M_0$ sampled extant species, $M_0+n-1$ branching events, $n$ sampling events that result in termination of lineages in the tree (tips), and $m$ remaining sampling events that lie within lineages of reconstructed tree and are called *sampled ancestors*. We denote the times of the branching events as $(x_1,x_2,\ldots x_{M_0+n-1})$, the times of the terminal sampling events as $(y_1,y_2,\ldots y_n)$, and the times of the remaining sampling events as $(z_1,z_2,\ldots z_m)$ (Fig. 2b).

To derive the probability density function for reconstructed trees, we need to define function $E(t)$, which is the probability that a lineage alive at time $t$ does not leave sampled descendants. It was shown (MacPherson et al., 2021) that




(a) Complete tree (b) Reconstructed tree

Figure 2: An example of a complete tree (a) and the corresponding reconstructed tree (b) obtained from the complete tree by trimming unobserved lineages (blue, dotted). The unobserved lineages are the lineages that do not terminate with samples (solid squares). In the resulting reconstructed tree, there are three types of nodes: branching nodes at times ($x_i$), tip samples ($y_i$ and present time), and sampled ancestors ($z_i$).

this function is the solution to the following differential equation:

$$\frac{\mathrm{d}E}{\mathrm{d}t} = -\left(\lambda + \mu + \psi\right) E + \lambda E^2 + \mu \tag{2}$$

with initial condition $E(0) = 1 - \rho_0$. The following proposition establishes the conditions on the rate functions which guarantee that the solution to this equation exists and is unique.

**Proposition 2.1.** Equation (2) has a unique extended solution on $[0, t_0]$ if the following conditions are satisfied:

- $\lambda \geq 0$, $\mu \geq 0$, and $\psi \geq 0$ are Lebesgue-integrable on $[0, t_0]$,
- $\int_0^t \lambda(s) e^{-\int_0^s \lambda(u)-\mu(u)-\psi(u)du} ds < \infty$ for $t \in [0, t_0]$, and
- either $\rho_0 > 0$ or $\psi(0) > 0$ and $\lambda$, $\mu$, and $\psi$ are continuous at zero.

*Proof.* See SI Section S2 □

In what follows, whenever reconstructed trees are considered, we assume that $\lambda$, $\mu$, and $\psi$ satisfy these conditions.

The probability density function for reconstructed trees (MacPherson et al., 2021) is:

$$f^{\text{rec.}}\left(\mathcal{T} | \lambda(t), \mu(t), \psi(t), \rho_0, t_0\right) = \rho_0^{M_0} \Phi(t_0) \prod_{i=1}^{M_0+n-1} \lambda\left(x_i\right) \Phi(x_i) \prod_{j=1}^{n} \frac{\psi\left(y_j\right) E(y_j)}{\Phi(y_j)} \prod_{k=1}^{m} \psi\left(z_k\right) \tag{3}$$

where $\Phi(t) = e^{\int_0^t 2\lambda(u)E(u)-\lambda(u)-\mu(u)-\psi(u)\, du}$. Note that $1 - E\left(t_0\right)$ is the probability that at least one lineage was sampled either in the past or the present (Morlon

et al., 2011). Then the probability density function for reconstructed trees, conditioning on the event of sampling at least one lineage ($S$), is:

$$f^{\text{rec.}}\left(\mathcal{T}|\lambda(t),\mu(t),\psi(t),\rho_0,t_0,S\right)=\frac{f^{\text{rec.}}\left(\mathcal{T}|\lambda(t),\mu(t),\psi(t),\rho_0,t_0\right)}{1-E\left(t_0\right)} \tag{4}$$

Consider a parametric statistical model, which is a class of distributions $P_\theta$ across a set of parameters $\Theta$: $\mathcal{P}_\Theta=\{P_\theta:\theta\in\Theta\}$.

**Definition 2.1** (Identifiability)**.** A statistical model $\mathcal{P}_\Theta=\{P_\theta:\theta\in\Theta\}$ is identifiable if $\theta\mapsto P_\theta$ is injective; that is, for all $\theta_1,\theta_2\in\Theta$, we have $P_{\theta_1}=P_{\theta_2}\Rightarrow\theta_1=\theta_2$. Two probability density functions $P_{\theta_1},P_{\theta_2}$ are considered equal if the set of points where they differ has Lebesgue measure zero.

The parameters $\theta$ of the FBD model are $(\lambda,\mu,\psi,\rho_0,t_0)$. We consider two models for which we show identifiability. The complete tree model is described by probability densities $f_\theta^{\text{compl.}}$ defined by equation (1) with restrictions on the parameters formulated in Theorem 3.1 that define set $\Theta_{\text{compl.}}$. The reconstructed tree model is described by probability densities $f_\theta^{\text{rec.}}$ defined by equation (3) and has a slightly different set of allowable (with restrictions from theorem 3.2) parameters $\Theta_{\text{rec.}}$.

### nLTT code implementation

In unidentifiable models (birth-death or birth-death sampling with removal at sampling cases) there are sets of birth, death, and sampling rates for which the probability density functions and therefore reconstructed tree distributions are identical (Louca and Pennell, 2020b; Louca et al., 2021). Such parameters form congruence classes, where any pair of parameters are congruent if they produce the same distributions of reconstructed trees.

To demonstrate congruent scenarios Louca and Pennell (2020b) and Louca et al. (2021) used deterministic lineage through time (dLTT) curves. This curve is the deterministic limit of the number of observed lineages at times $t$ generated in the stochastic birth-death sampling processes (Kubo and Iwasa, 1995). The stochastic process describes the number of lineages present, given diversification and sampling events which are realisations of a birth-death sampling process. Deterministic calculations assume that the timing of events is solely determined by the birth, death, and sampling rates with no random component. dLTT curve also corresponds to the expected number of lineages in the reconstructed trees generated by the stochastic birth-death sampling process (SI S4.2). Therefore, pairs of rates that produce different dLTT curves also produce different distributions of reconstructed trees.

Here we follow Louca et al. (2021) who considered a normalised lineage through time (nLTT) curve, which is a dLTT curve normalised by its area under the curve (Janzen et al., 2015). Obviously, different nLTT curves imply different dLTT curves. To plot nLTT curves Louca et al. (2021) used the R package *castor* (Louca and Doebeli, 2018). They randomly selected birth,

death, and sampling rates, which were generated under either an exponential or an Ornstein-Uhlenbeck stochastic process. Here we either randomly select exponential rates as in Louca et al. (2021) or manually select linear rates to avoid technical difficulties with obtaining congruent scenarios with positive rates. Given the selected rates $\theta_1 = (\lambda_1, \mu_1, \psi_1)$, we use the castor function **congruent_hbds_model** and calculations based on the pulled diversification rate to obtain congruent rates $\theta_2 = (\lambda_2, \mu_2, \psi_2)$. Then we calculate deterministic values using the **simulate_deterministic_hbds** function by generating a tree with between 100,000 and 200,000 tips, given the rates ($\theta_1$ or $\theta_2$) and sampling fraction $\rho_0 = 0$ to achieve a good approximation of the nLTTs. Rather than utilising the removal probability $r$, the package functions use the retention probability $\kappa = 1 - r$, which we specify accordingly.

We use a fork of *castor v.1.7.11* suitable for calculating nLTT curves for trees with sampled ancestors. This is available at https://github.com/bioDS/castor. The R code from Louca et al. (Louca et al., 2021) was used as a basis for the simulations. The deterministic parameter calculation code for FBD trees is available at https://github.com/bioDS/FBD-dLTTs.

# 3 Results

## 3.1 Identifiability of the time-dependent FBD model

Here we show the identifiability of the time-dependent FBD model parameters from complete trees.

**Theorem 3.1.** Suppose $t_0$ is fixed and $\lambda(t) \geq 0$, $\mu(t) \geq 0$, and $\psi(t) \geq 0$ are Lebesgue-integrable on $[0, t_0]$. Then the time-dependent FBD complete tree model (the probability density function for this model is defined by equation 1) is identifiable.

*Proof.* The key ingredient of the proof of the theorem is the following lemma:

**Lemma 3.1.** Assume $f^{\text{compl.}}_{\theta_1} = f^{\text{compl.}}_{\theta_2}$ almost everywhere, then the following equalities hold:

$$\lambda_1(t)\Psi_1(t) = \lambda_2(t)\Psi_2(t) \text{ for almost all } t \in [0, t_0] \tag{5}$$

$$\lambda_1(t)\mu_1(t) = \lambda_2(t)\mu_2(t) \text{ for almost all } t \in [0, t_0] \tag{6}$$

$$\psi_1(t) = \psi_2(t) \text{ for almost all } t \in [0, t_0] \tag{7}$$

$$\Psi_1(t_0) = \Psi_2(t_0) \tag{8}$$

*Proof.* Equalities (5)–(7) are obtained by considering trees that differ by one event as follows. To show (5), we note that the densities should be equal for almost all trees with $s$ edges and $s_e$ extant edges, and also for almost all trees with $s + 1$ edges and $s_e + 1$ extant edges. Given a pair of trees in which one tree is obtained from the other by attaching an additional edge at time $t$ (Figs. 3a and 3b), their densities differ by one term, namely, $\lambda(t)\Psi(t)$, which implies equality (5).

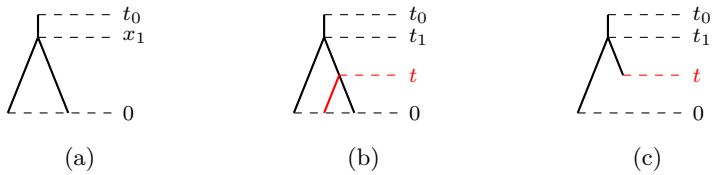



Figure 3: Tree modifications used in the proof of Lemma 3.1. Subfigure (a) depicts the original tree, while Subfigure (b) shows the addition of an extant edge born at time $t$ and Subfigure (c) shows the contraction of an extant edge such that it dies at time $t$.

By considering pairs of trees in which one edge switches from being extant to going extinct at time $t$ (Figs. 3a and 3c), we obtain $\frac{\mu_1(t)}{\Psi_1(t)} = \frac{\mu_2(t)}{\Psi_2(t)}$, which together with (5) implies (6). Similarly, considering trees with an additional sampling event, we obtain (7). Since all the individual terms except for $\Psi(t_0)$ are equal in the probability density functions of all trees under $\theta_1$ and $\theta_2$, we conclude that the $\Psi(t_0)$ terms should also be equal (8). The detailed proof can be found in SI section S3.2. □

We now finish the proof of the theorem as follows. Denote $\gamma(t) = \frac{\Psi_1(t)}{\Psi_2(t)}$. Equalities (5)-(7) imply that $\gamma$ must satisfy the following differential equation:

$$\frac{d\gamma}{dt} = \lambda_1\gamma^2 - (\lambda_1 + \mu_1)\gamma + \mu_1 \quad \text{for almost all } t \in [0, t_0] \tag{9}$$

Definition of $\gamma$ and equality (8) implies the boundary condition $\gamma(0) = \gamma(t_0) = 1$. $\gamma(t) = 1$ is a solution to this equation, which must be unique (SI section S3.3). Therefore, $\lambda_1(t) = \lambda_2(t)$ and $\mu_1(t) = \mu_2(t)$ for almost all $t \in [0, t_0]$. □

**Theorem 3.2.** Suppose $t_0$ is fixed, parameters $\lambda$, $\mu$, $\psi$ and $\rho_0$ satisfy the conditions of Proposition 2.1, and $\psi(t) > 0$ for almost all $t \in [0, t_0]$. Then the time-dependent FBD reconstructed tree model (the probability density function is defined by equation 3) is identifiable.

*Proof.* We prove this theorem by showing that there exists a one-to-one transformation of the FBD parameters (Eq. (10), Lemma 3.2) such that the distribution of the reconstructed trees produced by the process with original parameters and conditioned on sampling of at least one lineage is the same as the distribution of complete trees produced by the FBD process with transformed parameters (Lemma 3.3).Then we apply Theorem 3.1 to complete the proof.

Consider the following parameter transformation of original rates to *pulled* rates:

$$
\begin{aligned}
\tilde{\lambda} &= \lambda(1-E) \\
\tilde{\mu} &= \frac{\psi E}{1-E} \\
\tilde{\psi} &= \psi \\
\tilde{\rho}_0 &= 1
\end{aligned}
\tag{10}
$$

where $E(t)$ is defined in Equation (2). Note that since $\psi(t) > 0$ we have $E(t) < 1$ for all $t \in (0, t_0]$ (SI Section S2). Then if $\rho_0 = 0$, implying $E(0) = 1$, we set $\tilde{\mu}(0)$ to zero. Thus, all the rates are well-defined for all $t$.

We say that two sets of rates $\left[\lambda_1(t), \mu_1(t), \psi_1(t)\right]$ and $\left[\lambda_2(t), \mu_2(t), \psi_2(t)\right]$ are different if at least one of the rates ($\lambda$, $\mu$, or $\psi$) differ on at least a non-zero measure subset of $[0, t_0]$.

**Lemma 3.2.** Parameter transformation (10) is a one-to-one correspondence.

*Proof.* See SI Section S4.3. □

**Lemma 3.3.** The distribution of reconstructed trees for the time-dependent FBD process, conditioned on sampling of at least one lineage (through-time or at present), is the same as the distribution of the complete trees under the time-dependent FBD process with pulled rates and a sampling probability of one. That is,

$$
f^{\text{rec.}}\left(\mathcal{T} | \lambda, \mu, \psi, \rho_0, t_0, S\right) = f^{\text{compl.}}\left(\mathcal{T} \mid \tilde{\lambda}, \tilde{\mu}, \tilde{\psi}, t_0\right)
$$

*Proof.* The proof follows from substituting the pulled rates into the probability density function for the complete trees (SI Section S4). □

To finish the proof of the theorem suppose $f^{\text{rec.}}_{\theta_1} = f^{\text{rec.}}_{\theta_2}$. Then $1 - E_1(t_0) = 1 - E_2(t_0)$, because $1 - E(t_0)$ is the cumulative probability of all surviving trees. So $\frac{f^{\text{rec.}}_{\theta_1}}{1-E_1(t_0)} = \frac{f^{\text{rec.}}_{\theta_2}}{1-E_2(t_0)}$. Lemma 3.3 implies that $f^{\text{compl.}}_{\tilde{\theta}_1} = f^{\text{compl.}}_{\tilde{\theta}_2}$. Then it follows from Theorem 3.1 that $\tilde{\theta}_1 = \tilde{\theta}_2$, and from Lemma 3.2 that $\theta_1 = \theta_2$. □

Note, that if $\psi$ is equal to zero on some time interval within $[0, t_0]$ then only birth or death events can happen at this time interval. Therefore within this time interval, the process can be described by only one pulled birth rate as shown in (Louca and Pennell, 2020b) and the model becomes unidentifiable. In other words, by choosing identical birth, death and sampling rates whenever $\psi$ is greater than zero and congruent birth and death rates (as defined for unidentifiable birth-death models) whenever $\psi$ is zero we obtain congruent rates for the reconstructed FBD model.

## 3.2 Unidentifiability of the birth-death model with removal at sampling

In this section we address the identifiability problem for a time-dependent birth-death sampling model with removal at sampling (MacPherson et al., 2021). This model is defined by a time-dependent FBD process where sampled lineages are removed from the process after sampling with time-dependent removal probability $r(t)$, where $0 \leq r(t) \leq 1$.

For this model, we can also define pulled rates:

$$\begin{aligned} \tilde{\lambda} &= \lambda(1-E) \\ \tilde{\mu} &= \frac{\psi E}{1-E} + \psi r \\ \tilde{\psi} &= \psi(1-r) \end{aligned} \tag{11}$$

where $r \neq 0$ on some non-zero measure subset of $[0, t_0]$. Similarly to the FBD case, the probability density function of the reconstructed trees with original rates is the same as the probability density function of complete trees under the model with pulled rates (see SI section S5.2).

Now, for any given $\lambda > 0$, $\mu \geq 0$, $\psi > 0$, $1 \geq r \geq 0$, $\rho_0 \in (0, 1]$, and for any alternative rate $\psi^* > 0$ (here we also assume $\lambda$, $\psi$, and $\psi*$ are differentiable almost everywhere on $[0, t_0]$) we define

$$\lambda^* = \frac{\lambda\psi}{\psi^*} \tag{12}$$

$$\mu^* = \lambda^* - \psi^* + \frac{1}{\lambda^*}\frac{\mathrm{d}\lambda^*}{\mathrm{d}t} - \lambda + \mu + \psi - \frac{1}{\lambda}\frac{\mathrm{d}\lambda}{\mathrm{d}t} \tag{13}$$

$$r^* = 1 - \frac{\psi(1-r)}{\psi^*} \tag{14}$$

Equation (14) implies $\tilde{\psi} = \tilde{\psi^*}$. Louca et al. (Louca et al., 2021) showed that (12) and (13) imply:

$$\begin{aligned} \lambda(1-E) &= \lambda^*(1-E^*) \\ \frac{\psi}{1-E} &= \frac{\psi^*}{1-E^*} \end{aligned}$$

Therefore we have $\tilde{\lambda} = \tilde{\lambda^*}$. Noting that $\tilde{\mu} = \frac{\psi}{1-E} - \tilde{\psi}$ we have $\tilde{\mu} = \tilde{\mu^*}$.

Thus for a given model with parameters $\theta$ and an arbitrary $\psi^*$ distinct from $\psi$ we obtained another model defined by distinct parameters $\theta^*$ for which the pulled rates are equal and therefore the probability distributions over reconstructed trees are also equal. Given that we have found distinct parameter sets that produce the same probability distributions, the following theorem holds:

**Theorem 3.3.** Suppose $t_0$ is fixed, parameters $\lambda$, $\mu$, $\psi$ and $\rho_0$ satisfy the conditions of Proposition 2.1, $\psi(t) > 0$ for almost all $t \in [0, t_0]$, $\lambda$ and $\psi$ are differentiable almost everywhere on $[0, t_0]$, $0 \leq r(t) \leq 1$, and $r \neq 0$ on an non-zero

measure subset of $[0, t_0]$. Then the time-dependent birth-death sampling model with removal at sampling is unidentifiable for reconstructed trees.

In the above theorem, we assumed a restricted set of parameters for which the unidentifiability holds. For any wider parameter set that contains this restricted set, the model remains unidentifiable. However, if stricter or non-overlapping conditions apply, the model may become identifiable, e.g. fixing $r$ to zero (FBD model identifiability) or fixing $\psi$ to zero and considering piecewise constant or piecewise polynomial birth and death rates (Legried and Terhorst, 2022, 2023). Furthermore, if we fix the removal probability to a function that is not equal to one then given that the pulled rate transformation 11 is one-to-one (SI section S5), we obtain the following corollary:

**Corollary 3.1.** Suppose $t_0$ is fixed, parameters $\lambda$, $\mu$, $\psi$ and $\rho_0$ satisfy the conditions of Proposition 2.1, $\psi(t) > 0$ for almost all $t \in [0, t_0]$, $0 \leq r(t) \leq 1$, and $r(t)$ is a fixed function such that $r(t) \neq 1$ almost everywhere on $[0, t_0]$. Then the time-dependent birth-death sampling model with fixed removal at sampling probability is identifiable for reconstructed trees.

### 3.3 nLTT curves

Louca et al. (2021) illustrated the congruency of the birth-death model by comparing the nLTT curves produced using different congruent rate functions. We follow this methodology to assess deterministic properties given particular speciation, extinction, and sampling rates, in our case under the FBD process. We use two sets of rate functions, $\theta_1$ and $\theta_2$. Given $\theta_1 = (\lambda_1, \mu_1, \psi_1)$, $\theta_2 = (\lambda_2, \mu_2, \psi_2)$ is chosen such that $\theta_1$ and $\theta_2$ result in the same nLTT curve under the time-dependent birth-death sampling model with a removal probability $r = 1$. We then use $\theta_1$ and $\theta_2$ to produce nLTT curves for scenarios under the process with smaller removal probabilities, specifically $r = 0.5$ and $r = 0$, which allow for sampled ancestors.

Figure 4 illustrates the diversification rates and deterministic curves for a scenario with linear diversification rates for $\theta_1$, while Figure 5 depicts a scenario with exponential forms for $\theta_1$. Note that restricting $\theta_1$ to linear or exponential rates does not place the same restriction on $\theta_2$. We observe that the introduction of sampled ancestors for the same speciation, extinction, and sampling rates affects the nLTT curve, particularly near its peak. As time passes and sampling events occur without removing lineages, we expect the total number of speciation, extinction, and sampling events to increase with the number of lineages, given the same lineage-independent rates.

While using a removal probability of less than one results in different nLTT curves for scenarios that were previously congruent under $r = 1$, the difference in the nLTT may be small for all times considered. An example of this is shown in Figure 6. This small difference between nLTT curves is observed in scenarios where the rate of sampling is low compared to the total tree length. In such cases, there are relatively few time points at which a lineage may be removed

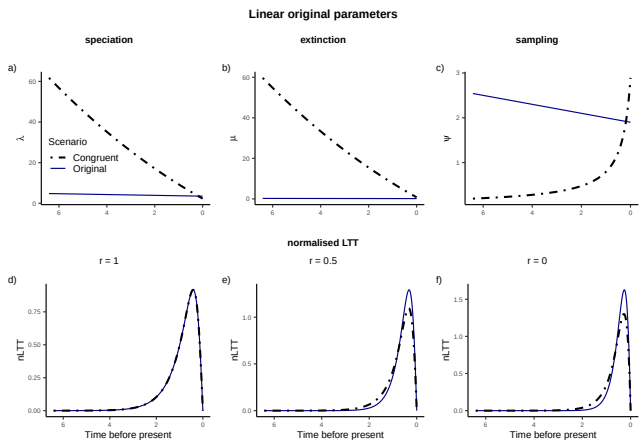



Figure 4: A set of congruent scenarios where $\theta_1$ has linear rates. Subfigures a–c) show the parameters of $\theta_1$ and $\theta_2$ where $\theta_1 = (0.2t + 3.5, 0.015t + 0.15, 0.1t + 1.9)$. Subfigures d–f) illustrate that when sampled ancestors are permitted, $\theta_1$ and $\theta_2$ no longer produce identical nLTT curves.

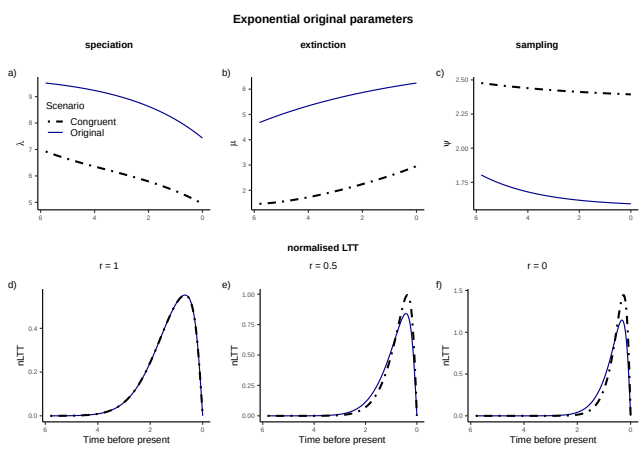



Figure 5: A set of congruent scenarios where $\theta_1$ has exponential rates which were randomly selected. Subfigures a–c) show the parameters of $\theta_1$ and $\theta_2$ where $\theta_1 \approx (9.84 - 2.41e^{-0.35t}, 7.16 - 0.92e^{0.17t}, 1.57 + 0.02e^{0.43t})$. Subfigures d–f) illustrate that when sampled ancestors are permitted, $\theta_1$ and $\theta_2$ no longer produce identical nLTT curves.

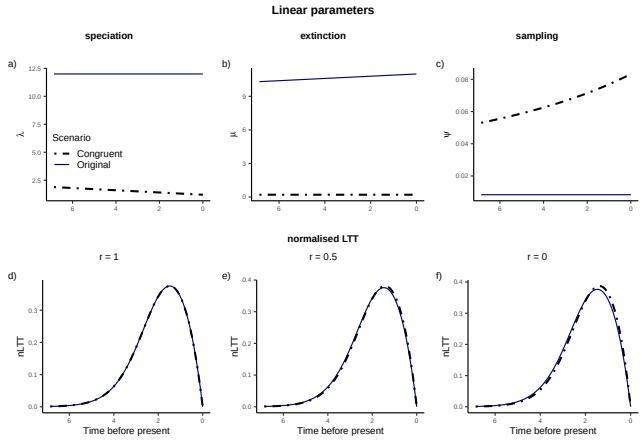



Figure 6: A set of congruent scenarios, where $\theta_1$ has linear rates. Subfigures a–c) show the parameters of $\theta_1 = \left(12, -0.1t + 11, \frac{1}{120}\right)$ and $\theta_2 = \left(0.1t + 1, \frac{1}{120}, \frac{0.1}{0.1t+1}\right)$. Subfigures d–f) illustrate that although reducing the removal probability results in different nLTT curves for $\theta_1$ and $\theta_2$, the difference in density is small for all times considered.

due to sampling, and thus the number of lineages in trees is largely unaffected by the removal probability.

A direct comparison of nLTT curves under different linear removal probabilities is shown in Figure 7. In this example scenario, the speciation, extinction and sampling rates are fixed. The linear removal probability functions with different slops reflect scenarios ranging from rapid growth of removal probability with time to constant removal probability. We observe that the steeper the removal probability function, the greater the deviation in the nLTT curve compared to the case where the removal probability is constant (shown as a solid line).

# 4 Discussion

In this study, we have shown that the time-dependent FBD model is identifiable and therefore eliminated the concern related to the use of such models for parameter inference (Pagel, 2020) caused by the previous unidentifiability results. Given the identifiability, the FBD model is an absolutely suitable model for inferring speciation and extinction rates from dated phylogenies, as well as epidemiological parameters in scenarios where removal upon diagnosis is not implied.

Previous studies showed identifiability of the birth-death model without sampling for piecewise constant and piecewise polynomial rates. The FBD model is identifiable in an arbitrary class of rate functions. While a piecewise constant model can be defined to approximate an unidentifiable model arbitrarily

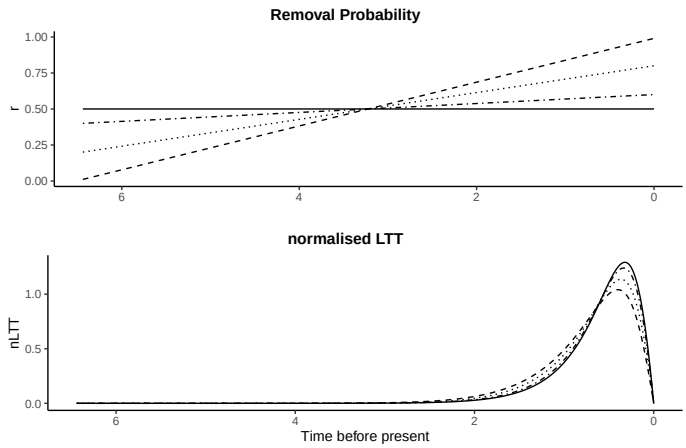



Figure 7: An example of differences in nLTT curves caused by changing the removal probability $r$ while keeping speciation, extinction and sampling rates fixed: $\theta = (0.2t + 3.5, 0.015t + 0.15, 0.1t + 1.9)$.

closely (Legried and Terhorst, 2022), it is unclear how much a particular piecewise constant model that maximises the likelihood function deviates from the true non-piecewise constant set of rates. Although computational methods do not allow us to infer arbitrary functions, and we still need to restrict the form of the estimated rates, we can now directly model, for example, exponential growth, which would reduce the number of estimated parameters and required tree size compared to piecewise constant or polynomial approximations.

Previous implementations of the FBD model have relied on simulation studies to show identifiability (Gavryushkina et al., 2014); our theoretical results make this time-consuming step of investigating parameter identifiability via simulation unnecessary, whether one wishes to use constant rates, piecewise-constant, piecewise-polynomial, or more complex scenarios. Such simulations can still be useful for assessing the precision of estimated parameters, robustness of the model to assumptions violation or validation of software implementations.

It is important to note, that although the FBD model is theoretically identifiable, a given dataset may not be of a sufficient size for rate selection (Morlon et al., 2022; Legried and Terhorst, 2023). While theoretically unidentifiable models cannot be resolved with any amount of data (Louca and Pennell, 2020b), there is no fixed finite quantity of data for theoretically identifiable scenarios which guarantees a desired precision of the inferred rates. Small trees or samples may lead to wide confidence or credible intervals for parameters such as extinction rate even in the case when the true rates have simple forms (Morlon et al., 2022). A further caveat to applying theoretical identifiability results is that they unrealistically rely on knowing the true phylogeny without an error. With limited molecular, morphological, or other data, simplified evolutionary models and otherwise non-perfect inference methods, knowing the true phylogeny

without an error is not possible.

The results of our study show the importance of including fossils when estimating speciation and extinction rates. The general case when only a dated phylogeny is available is consistent with many congruent rates. Once fossils are placed on the phylogeny, the congruence class collapses to a single set of rates that are consistent with the given sampled ancestor phylogenetic tree (a tree in which fossil samples can be sampled ancestors, Gavryushkina et al., 2014). Including sampled ancestor fossils has been shown to be very important in reducing biases in estimated extinction rates (Gavryushkina et al., 2014; O'Reilly and Donoghue, 2019). Here, we have shown that allowing fossils to be sampled ancestors is crucial for theoretical identifiability of the diversification rates.

It is intuitive that allowing removal probabilities of less than one provides more information about the sampling rate, which in turn helps determine the birth and death rates. As a lineage under the time-dependent FBD model may survive being sampled, there is the potential for each lineage to be sampled on multiple occasions. These additional sampling events define the time intervals between sampling on continuing lineages, which are informative for determining the sampling rate. Whereas if sampling guarantees removal, the sampling rate is confounded by the unknown birth and death rates. It becomes impossible to distinguish between scenarios with low sampling rate and many unobserved lineages and high sampling rate with few unobserved lineages.

The ability to infer the sampling rate itself, even if the sampling scheme was not constant through time, is an additional benefit of the FBD model. Inferred fossilisation rates could for example be further used in consecutive studies as prior information.

As we demonstrated in the nLTT examples, the sampling rate has to be high enough to introduce noticeable differences in the distributions of the reconstructed trees and therefore enable precise maximum likelihood or Bayesian inference of the rates. Our identifiability finding cannot be applied to scenarios which allow past time periods with no probability of sampling lineages. The proven one-to-one correspondence between complete and reconstructed trees relies upon a non-zero past sampling rate to determine original rates from pulled rates. Given that species' physical structure or habitat may make them unsuitable for fossilisation (Shaw et al., 2021), the FBD model might not be suitable for scenarios requiring $\psi(t) = 0$ for significantly long periods of time. However if the birth and death rates are piecewise polynomial where $\psi$ is equal to zero we can apply the results of Legried and Terhorst (Legried and Terhorst, 2022, 2023) to obtain the identifiability of the reconstructed FBD model.

It is important to note, that although the inclusion of fossils helps identify the speciation and extinction rates, the age uncertainty and biased sampling associated with fossil data could still propagate to the speciation and extinction rates estimates. Modelling fossil age uncertainty and using stratigraphic ranges of species as opposed to single fossil representatives (Stadler et al., 2018) could improve the accuracy of credible intervals and reduce the bias.

The identifiability of the time-dependent FBD model arises from allowing sampled fossils to remain in the process (removal probability $r$ fixed to zero)

and no longer holds once the removal probability is set to one (Louca et al., 2021), as has been previously shown for piecewise constant rates (Gavryushkina et al., 2014). Here we have also shown that the time-dependent birth-death sampling model remains identifiable, provided $r$ is fixed to a function not equal to one. Such a model could be used in epidemiological studies for diseases where diagnosis does not imply that the patient prevents new transmissions, for example, influenza or coronavirus (where isolation strategies have not been applied). However, if one wants to estimate the removal probability from the data, the model becomes unidentifiable. In this case, one can impose strong prior assumptions on the rates (Morlon et al., 2022) or explore congruence classes for patterns that are conserved within them (Höhna et al., 2022; Kopperud et al., 2023).

While we have considered the impact of different fixed constant and linear removal probabilities on nLTT curves in several examples, the question of how misspecified removal probabilities impact the inference is yet to be explored. The difference in nLTT curves for different linear removal probabilities (Figure 7) demonstrates how this parameter affects the distribution of reconstructed trees, highlighting that a misspecified fixed removal probability could potentially influence diversification rate estimates. Potential avenues of inquiry include describing how the estimates of selected parameter rates are affected by the magnitude of error in removal probabilities.

By extending pulled rates introduced for birth-death sampling models (Louca et al., 2021) to the case of the FBD model, we also showed that the distributions of reconstructed and complete trees generated by related FBD processes are equivalent. This equivalence provides a different view on the nature of reconstructed trees that can be seen as complete trees generated by an FBD model with adjusted or pulled rates, as was previously noted by Louca et al. (2021) for trees without sampled ancestors. The original birth rate is adjusted by the probability of the newly born lineage to be observed. The original sampling rate is split in two: pulled death rate and pulled sampling rate, depending on the type of the occurred sampling event. Events that are followed by lineage disappearance from the observer become death events and events that are followed by further sampling events (sampled ancestors) become sampling events.

Despite the identifiability of the FBD model in the arbitrary class of rates, the inference methods require fixing the class of inferred rates. An avenue for future research would be to quantify how well different classes of functions that are available for inference (e.g., the widely used piece-wise constant functions) approximate the true rates of different function classes which imply identifiability of the corresponding FBD model.

Another direction for future research is to investigate identifiability for variations of birth-death models that account for lineage specific birth and death rates (Maddison et al., 2007), age-dependent rates (Hagen et al., 2015), or different speciation types (Stadler et al., 2018). While the identifiability of the pure birth multi-type model with some restrictions on the transition rates has been recently shown (Dragomir et al., 2023), the identifiability of the parameters in other models has not yet been generally established.

Finally, we would like to emphasise that this study completely answers the question of the time-dependent reconstructed FBD model identifiability. We have shown the identifiability for arbitrary rate functions (subject to the conditions that guarantee the probability distribution over the reconstructed trees can be defined) as long as the sampling rate is strictly positive. The last requirement is essential for identifiability, because allowing $\psi$ to be equal to zero produces congruent scenarios and therefore implies unidentifiability.

In conclusion, the FBD model is an excellent candidate for modelling evolutionary processes. It shares computational conveniences of other birth-death models, describes sampling process with more realistic assumptions, and as we have shown here, possesses an essential statistical property of identifiability.

## Supplementary material

Drayd link: https://doi.org/10.5061/dryad.sj3tx96ch


## Acknowledgements

We thank Dmitry Berdinsky, Lars Berling, and Kerry Manson for their helpful discussions throughout this research.

## Funding

AG was partially supported by Royal Society Te Apārangi through a Rutherford Discovery Fellowship (UOC1702) and a Marsden grant (21-UOC-057), and a Data Science Programmes grant (UOAX1932). AG and ASG were supported by Ministry of Business, Innovation, and Employment of New Zealand through an Endeavour Smart Ideas grant (UOOX1912). TGV was supported by funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme grant agreement no. 101001077. KT was supported by UC Aho Hīnātore, UC Accelerator Scholarship.


## Author Contributions

ASG initiated the idea of proving identifiability. KT, TGV, AG, ASG extended the idea of transformed parameters to FBD case. KT, AG, ASG worked on the proofs. KT investigated nLTT plots. KT, AG, ASG wrote and edited the manuscript. TGV edited the manuscript.

# References


Jérémy Andréoletti and Hélène Morlon. Exploring congruent diversification histories with flexibility and parsimony. *Methods Ecol. Evol.*, October 2023.

Remco Bouckaert, Timothy G Vaughan, Joëlle Barido-Sottani, Sebastián Duchêne, Mathieu Fourment, Alexandra Gavryushkina, Joseph Heled, Graham Jones, Denise Kühnert, Nicola De Maio, et al. Beast 2.5: An advanced software platform for bayesian evolutionary analysis. *PLoS Comput. Biol.*, 15 (4):e1006650, 2019.

Dakota Dragomir, Elizabeth S Allman, and John A Rhodes. Parameter identifiability of a multitype pure-birth model of speciation. *J. Comput. Biol.*, 30(3): 277–292, 2023.

Willy Feller. Die grundlagen der volterraschen theorie des kampfes ums dasein in wahrscheinlichkeitstheoretischer behandlung. *Acta Biotheor.*, 5(1):11–40, 1939.

Alex Gavryushkin and Alexei J Drummond. The space of ultrametric phylogenetic trees. *J. Theor. Biol.*, 403:197–208, August 2016.

Alex Gavryushkin, Chris Whidden, and Frederick A Matsen, 4th. The combinatorics of discrete time-trees: theory and open problems. *J. Math. Biol.*, 76(5): 1101–1121, April 2018.

Alexandra Gavryushkina, David Welch, Tanja Stadler, and Alexei J Drummond. Bayesian inference of sampled ancestor trees for epidemiology and fossil calibration. *PLoS Comput. Biol.*, 10(12):e1003919, December 2014.

Oskar Hagen, Klaas Hartmann, Mike Steel, and Tanja Stadler. Age-dependent speciation can explain the shape of empirical phylogenies. *Syst. Biol.*, 64(3): 432–440, 2015.

Tracy A Heath, John P Huelsenbeck, and Tanja Stadler. The fossilized birth–death process for coherent calibration of divergence-time estimates. *Proc. Natl. Acad. Sci. U. S. A.*, 111(29):E2957–E2966, 2014.

Sebastian Höhna, Michael J Landis, Tracy A Heath, Bastien Boussau, Nicolas Lartillot, Brian R Moore, John P Huelsenbeck, and Fredrik Ronquist. Revbayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification language. *Syst. Biol.*, 65(4):726–736, 2016.

Sebastian Höhna, Bjørn T Kopperud, and Andrew F Magee. CRABS: Congruent rate analyses in birth–death scenarios. *Methods Ecol. Evol.*, 13(12):2709–2718, December 2022.

Thijs Janzen, Sebastian Höhna, and Rampal S Etienne. Approximate bayesian computation of diversification rates from molecular phylogenies: introducing a new efficient summary statistic, the nltt. *Methods in ecology and evolution*, 6(5):566–575, 2015.

David G Kendall. On the generalized" birth-and-death" process. *Ann. Math. Stat.*, 19(1):1–15, 1948.

Bjørn T Kopperud, Andrew F Magee, and Sebastian Höhna. Rapidly changing speciation and extinction rates can be inferred in spite of nonidentifiability. *Proc. Natl. Acad. Sci. U. S. A.*, 120(7):e2208851120, February 2023.

Takuya Kubo and Yoh Iwasa. Inferring the rates of branching and extinction from molecular phylogenies. *Evolution*, 49(4):694–704, August 1995.

Brandon Legried and Jonathan Terhorst. A class of identifiable phylogenetic birth-death models. *Proc. Natl. Acad. Sci. U. S. A.*, 119(35):e2119513119, August 2022.

Brandon Legried and Jonathan Terhorst. Identifiability and inference of phylogenetic birth–death models. *J. Theor. Biol.*, 568:111520, July 2023.

Stilianos Louca and Michael Doebeli. Efficient comparative phylogenetics on large trees. *Bioinformatics*, 34(6):1053–1055, March 2018.

Stilianos Louca and Matthew W Pennell. A general and efficient algorithm for the likelihood of diversification and Discrete-Trait evolutionary models. *Syst. Biol.*, 69(3):545–556, May 2020a.

Stilianos Louca and Matthew W Pennell. Extant timetrees are consistent with a myriad of diversification histories. *Nature*, 580(7804):502–505, April 2020b.

Stilianos Louca, Angela McLaughlin, Ailene MacPherson, Jeffrey B Joy, and Matthew W Pennell. Fundamental identifiability limits in molecular epidemiology. *Mol. Biol. Evol.*, 38(9):4010–4024, August 2021.

Ailene MacPherson, Stilianos Louca, Angela McLaughlin, Jeffrey B Joy, and Matthew W Pennell. Unifying phylogenetic Birth-Death models in epidemiology and macroevolution. *Syst. Biol.*, 71(1):172–189, December 2021.

Wayne P Maddison, Peter E Midford, and Sarah P Otto. Estimating a binary character's effect on speciation and extinction. *Syst. Biol.*, 56(5):701–710, 2007.

Hélène Morlon, Todd L Parsons, and Joshua B Plotkin. Reconciling molecular phylogenies with the fossil record. *Proc. Natl. Acad. Sci. U. S. A.*, 108(39): 16327–16332, September 2011.

Hélène Morlon, Stéphane Robin, and Florian Hartig. Studying speciation and extinction dynamics from phylogenies: addressing identifiability issues. *Trends Ecol. Evol.*, 37(6):497–506, June 2022.

S Nee, R M May, and P H Harvey. The reconstructed evolutionary process. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, 344(1309):305–311, May 1994.

S Nee, E C Holmes, R M May, and others. Estimating extinction from molecular phylogenies. *Extinction*, 1995.

Sean Nee. Birth-death models in macroevolution. *Annu. Rev. Ecol. Evol. Syst.*, 37:1–17, 2006.

Joseph E O'Reilly and Philip C J Donoghue. The effect of fossil sampling on the estimation of divergence times with the fossilized Birth–Death process. *Syst. Biol.*, 69(1):124–138, May 2019.

Mark Pagel. Evolutionary trees can't reveal speciation and extinction rates. *Nature*, 580:461–462, April 2020.

David M Raup, Stephen Jay Gould, Thomas JM Schopf, and Daniel S Simberloff. Stochastic models of phylogeny and the evolution of diversity. *J. Geol.*, 81(5): 525–542, 1973.

Jack O Shaw, Derek E G Briggs, and Pincelli M Hull. Fossilization potential of marine assemblages and environments. *Geology*, 49(3):258–262, March 2021.

Daniele Silvestro, Nicolas Salamin, and Jan Schnitzler. Pyrate: a new program to estimate speciation and extinction rates from incomplete fossil data. *Methods Ecol. Evol.*, 5(10):1126–1131, 2014.

Tanja Stadler. On incomplete sampling under birth-death models and connections to the sampling-based coalescent. *J. Theor. Biol.*, 261(1):58–66, November 2009.

Tanja Stadler. Sampling-through-time in birth–death trees. *J. Theor. Biol.*, 267 (3):396–404, December 2010.

Tanja Stadler. Mammalian phylogeny reveals recent diversification rate shifts. *Proc. Natl. Acad. Sci. U. S. A.*, 108(15):6187–6192, April 2011.

Tanja Stadler, Roger Kouyos, Viktor von Wyl, Sabine Yerly, Jürg Böni, Philippe Bürgisser, Thomas Klimkait, Beda Joos, Philip Rieder, Dong Xie, Huldrych F Günthard, Alexei J Drummond, Sebastian Bonhoeffer, and Swiss HIV Cohort Study. Estimating the basic reproductive number from viral sequence data. *Mol. Biol. Evol.*, 29(1):347–357, January 2012.

Tanja Stadler, Denise Kühnert, Sebastian Bonhoeffer, and Alexei J Drummond. Birth–death skyline plot reveals temporal changes of epidemic spread in HIV and hepatitis C virus (HCV). *Proc. Natl. Acad. Sci. U. S. A.*, 110(1):228–233, 2013.

Tanja Stadler, Alexandra Gavryushkina, Rachel CM Warnock, Alexei J Drummond, and Tracy A Heath. The fossilized birth-death model for the analysis of stratigraphic range data under different speciation modes. *J. Theor. Biol.*, 447:41–55, 2018.

Mike Steel. *Phylogeny: Discrete and Random Processes in Evolution.* SIAM-Society for Industrial and Applied Mathematics, September 2016.

April M Wright, David W Bapst, Joëlle Barido-Sottani, and Rachel C M Warnock. Integrating fossil observations into phylogenetics using the fossilized Birth–Death model. *Annu. Rev. Ecol. Evol. Syst.*, 53(1):251–273, November 2022.

Z Yang and B Rannala. Bayesian phylogenetic inference using DNA sequences: a markov chain monte carlo method. *Mol. Biol. Evol.*, 14(7):717–724, July 1997.