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]✉

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 and generate phylogenetic trees. It has recently 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 ensures that statistical methods that implement this model infer the true underlying temporal diversification or epidemiological dynamics from phylogenetic trees or directly from molecular or other comparative data. We further show that the time-dependent 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.

Significance Statement. Identifiability of a statistical model is crucial for consistent inference which guarantees that parameters can be estimated close to their true values with sufficient amount of data. Recent results on unidentifiability of birth-death sampling models have put into question many macroevolutionary and epidemiological studies and caused considerable concern among researchers using these models. A related class of models, fossilised birth-death (FBD) models, has recently become a major tool for inferring past speciation or infectious disease transmission dynamics and dated evolutionary (phylogenetic) trees. Our result showing identifiability of FBD models enables researchers to confidently use these models

[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

E-mail addresses: ✉sasha.gavryushkina@canterbury.ac.nz.

Date: February 9, 2024.

and ensures the large body of research in palaeontology, ecology, and epidemiology reliant on the previously estimated parameters remains valid.

Author Contributions. ASG initiated the idea of proving identifiablity. 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.

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 [1–6] to infectious disease transmission histories [7, 8]. A birth-death process describes how lineages bifurcate and die with corresponding birth and death rates, and results in a phylogenetic tree [9, 10]. 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 [11–13]. 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.

Inferring birth and death rates is challenging due to the fact that samples from which phylogenetic trees are reconstructed do not represent all lineages that are involved in the process, leaving some portions of the trees unobserved. This challenge has been approached by modelling the sampling proportion [14] and considering the distribution of reconstructed trees — the observed parts of the trees [2].

Temporal changes in the rates of species diversification or infectious disease transmission have been accounted for by time-dependent birth and death rates [8, 10]. The probability density functions for reconstructed trees, in this case, were derived for the simple case of piecewise constant rates [15] or more general case with arbitrary time-dependent rates [6], which enabled efficient inference of the speciation or transmission temporal dynamics [8, 16]. Importantly, such inference relies on statistical identifiability of the time-dependent birth-death models, meaning that, given sufficiently large trees, one can infer the true rates.

Despite the wide popularity of these models, it has only recently been shown 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 [17]. Therefore, rate parameter estimates for such trees can not be reliably obtained using statistical methods. Crucially, Louca and Pennell found that the number of such birth-death pairs is infinite [17], which rules out manual parameter selection from the class of equally likely parameter pairs as a mitigating strategy. This result has caused major concerns [18] among researchers which 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 [19, 20]. 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 [21]. For this reason, an additional parameter, sampling rate, was added to the birth-death process, introducing a birth-death sampling process [6, 8, 21–23]. 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.[24] 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 [25] 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 [3, 21]. This model is widely used to estimate fossilisation rates and date species phylogenies [26, 27], however, the identifiability of the time-dependent rates for the FBD model has not been established [25].

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 also show that the time-dependent birth-death sampling model [23], 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.

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 lineages 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 and demonstrate an important relationship between the distributions of complete and reconstructed trees under the time-dependent FBD model.

2. Preliminaries

Here we consider a time-dependent fossilised birth-death (FBD) process [3, 21, 23, 28]. 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 ρ_0, called the *sampling fraction*. This time-dependant FBD process is a special case of the time-dependent birth-death sampling process [23] 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 $\mathcal{T}$ (Figure 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. [23]. We also 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 [29–31], 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. [6, 21, 23]), 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".

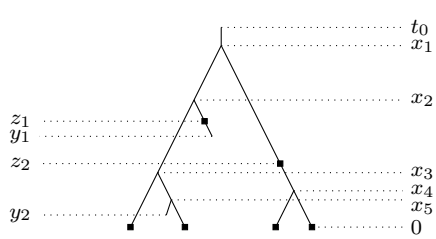
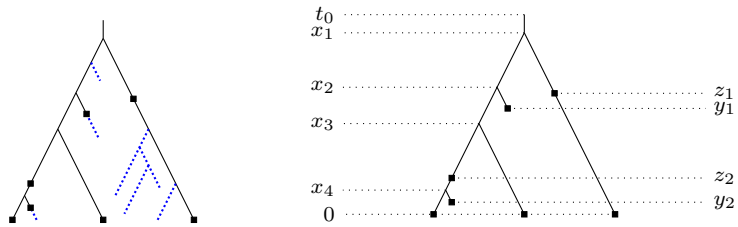


FIGURE 1. An example of a complete tree generated by a time-dependent FBD process. The stem of the tree is denoted t_0, and the present occurs at time 0. Birth events occur at times x_i, death events at times y_j, and samples at times z_k.

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*



(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 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).

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 ρ_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 $(x_1, x_2, \dots x_{M_0+n-1})$, the times of the terminal sampling events as $(y_1, y_2, \dots y_n)$, and the times of the remaining sampling events as $(z_1, z_2, \dots z_m)$ (Figure 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 [23] that 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 function 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 λ, μ, and ψ are continuous at zero.

Proof. See SI Section S2 □

In what follows we assume that λ, μ, and ψ satisfy these conditions.

The probability density function for reconstructed trees [23] 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 [6]. Then the probability density function for reconstructed trees, conditioning on the event of sampling at least one lineage (S), is:

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

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

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 θ 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 certain restrictions on the parameters that define set $\Theta_{\text{compl.}}$. The reconstructed tree model is defined by probability densities $f_\theta^{\text{rec.}}$ defined by equation (3) and has a slightly different set of allowable parameters $\Theta_{\text{rec.}}$.

3. 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 $f_{\theta_1}^{\text{compl.}} = f_{\theta_2}^{\text{compl.}}$ almost everywhere, then the following equalities hold:

$$\lambda_1(t)\Psi_1(t) = \lambda_2(t)\Psi_2(t) \text{ for almost all } t \in [0, t_0] \tag{4}$$

$$\lambda_1(t)\mu_1(t) = \lambda_2(t)\mu_2(t) \text{ for almost all } t \in [0, t_0] \tag{5}$$

$$\psi_1(t) = \psi_2(t) \text{ for almost all } t \in [0, t_0] \tag{6}$$

$$\Psi_1(t_0) = \Psi_2(t_0) \tag{7}$$

Proof. Equalities (4)–(6) are obtained by considering trees that differ by one event as follows. To show (4), 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

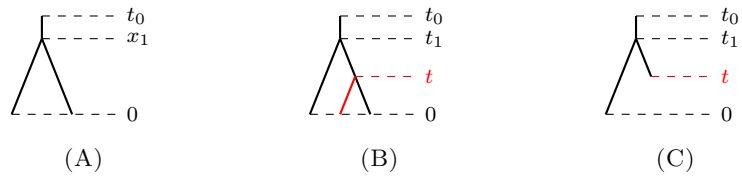


FIGURE 3. Tree modifications used in the proof of Lemma 3.1. Subfigure (A) depicts the original tree, while Subfigure (B) shows the addition of an extant edge born at time t and Subfigure (C) shows the contraction of an extant edge such that it dies at time t.

other by attaching an additional edge at time t (Figure 3A and 3B), their densities differ by one term, namely, $\lambda(t)\Psi(t)$, which implies equality (4).

By considering pairs of trees in which one edge switches from being extant to going extinct at time t (Figure 3A and 3C), we obtain $\frac{\mu_1(t)}{\Psi_1(t)} = \frac{\mu_2(t)}{\Psi_2(t)}$, which together with (4) implies (5). Similarly, considering trees with an additional sampling event, we obtain (6). Since all the individual terms except for $\Psi(t_0)$ are equal in the equal probability density functions, we conclude that the $\Psi(t_0)$ terms should also be equal (7). 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 (4)-(6) imply that γ must satisfy the following differential equation:

$$\frac{d\gamma}{dt} = \lambda_1\gamma^2 - (\lambda_1 + \mu_1)\gamma + \mu \quad \text{for almost all } t \in [0, t_0] \tag{8}$$

Definition of γ and equality (7) 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 λ, μ, ψ and ρ_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 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. Then we apply Theorem 3.1 to complete the proof.

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

$$\begin{aligned} \tilde{\lambda} &= \lambda(1 - E) \\ \tilde{\mu} &= \frac{\psi E}{1 - E} \\ \tilde{\psi} &= \psi \\ \tilde{\rho}_0 &= 1 \end{aligned} \tag{9}$$

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.

Lemma 3.2. 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). □

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 (λ, μ, or ψ) differ on at least a non-zero measure subset of $[0, t_0]$.

Lemma 3.3. Parameter transformation (9) is a one-to-one correspondence.

Proof. See SI Section S4.3. □

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 3.2 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 3.3 that $\theta_1 = \theta_2$. □

Note, that if ψ is equal to zero on some subset of $[0, t_0]$ then at every point of this subset the process can be described by only two pulled rates as shown in [17] and the model becomes unidentifiable. In other words, by choosing identical birth, death and sampling rates whenever ψ is greater than zero and congruent birth and death rates (as defined for unidentifiable birth-death models) whenever ψ is zero we obtain congruent rates for the reconstructed FBD model.

4. Unidentifiability of the birth-death model with removal at sampling

In this section we address the identifiability problem for a birth-death sampling model with removal at sampling [23]. 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:

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

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 $\psi^* > 0$ we define

$$\lambda^* = \frac{\lambda\psi}{\psi^*} \tag{11}$$

$$\mu^* = \lambda^* - \psi^* + \frac{1}{\lambda^*}\frac{\mathrm{d}\lambda^*}{\mathrm{d}t} - \lambda + \mu + \psi - \frac{1}{\lambda}\frac{\mathrm{d}\lambda}{\mathrm{d}t} \tag{12}$$

$$r^* = 1 - \frac{\psi(1-r)}{\psi^*} \tag{13}$$

Equation (13) implies $\tilde{\psi} = \tilde{\psi^*}$. Louca et al. [24] showed that (11) and (12) imply:

$$\lambda(1-E) = \lambda^*(1-E^*)$$
$$\frac{\psi}{1-E} = \frac{\psi^*}{1-E^*}$$

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 θ and an arbitrary ψ^* distinct from ψ we obtained another model defined by distinct parameters θ^* for which the pulled rates are equal and therefore the probability distributions over reconstructed trees are also equal. Therefore the time-dependant birth-death sampling model with removal at sampling is unidentifiable. However, if one fixes r to a function that is not equal to one almost everywhere on $[0, t_0]$, then the model becomes identifiable (SI section S5).

5. nLTT curves

In unidentifiable models (birth-death or birth-death sampling with removal at sampling cases) there are sets of birth-death sampling rates for which the probability density functions and therefore reconstructed tree distributions are identical [17, 24]. Such parameters form congruence classes, where any pair of parameters are congruent if they produce the same distributions of reconstructed trees. Such congruent rates also produce the same normalised deterministic lineage through time (nLTT) curves. This curve is the expected number of lineages at times t generated in the stochastic birth-death sampling processes [32]. The stochastic process describes the number of lineages present, given diversification events which are realisations of birth-death sampling process. Deterministic calculations assume that the timing of events is solely determined by the diversification rates, with no random component. Louca *et al.* [24] used the R package *castor* [33] to illustrate the congruency of the birth-death model by comparing the nLTT curves produced using different congruent rate parameters.

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 parameters, θ_1 and θ_2. Given $\theta_1 = (\lambda_1, \mu_1, \psi_1)$, $\theta_2 = (\lambda_2, \mu_2, \psi_2)$ is chosen such that θ_1 and θ_2 result in the same nLTT curve under the time-dependent birth-death sampling model with a removal probability $r = 1$.

We then use θ_1 and θ_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. The top half of Figure 4 illustrates the diversification rates and deterministic curves for a scenario with exponential diversification rates for θ_1, while

the bottom half depicts a scenario with linear forms for θ_1. Note that restricting θ_1 to linear or exponential rates does not place the same restriction on θ_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 5. 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 due to sampling, and thus the number of lineages in trees is largely unaffected by the removal probability.

6. Discussion

In this study, we have shown that the time-dependant FBD model is identifiable and therefore eliminated the concern related to the use of such models for parameter inference [18] 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 closely [19], 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 [22]; 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 [20, 34]. While theoretically unidentifiable models cannot be resolved with any amount of data [17], 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 [34].

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

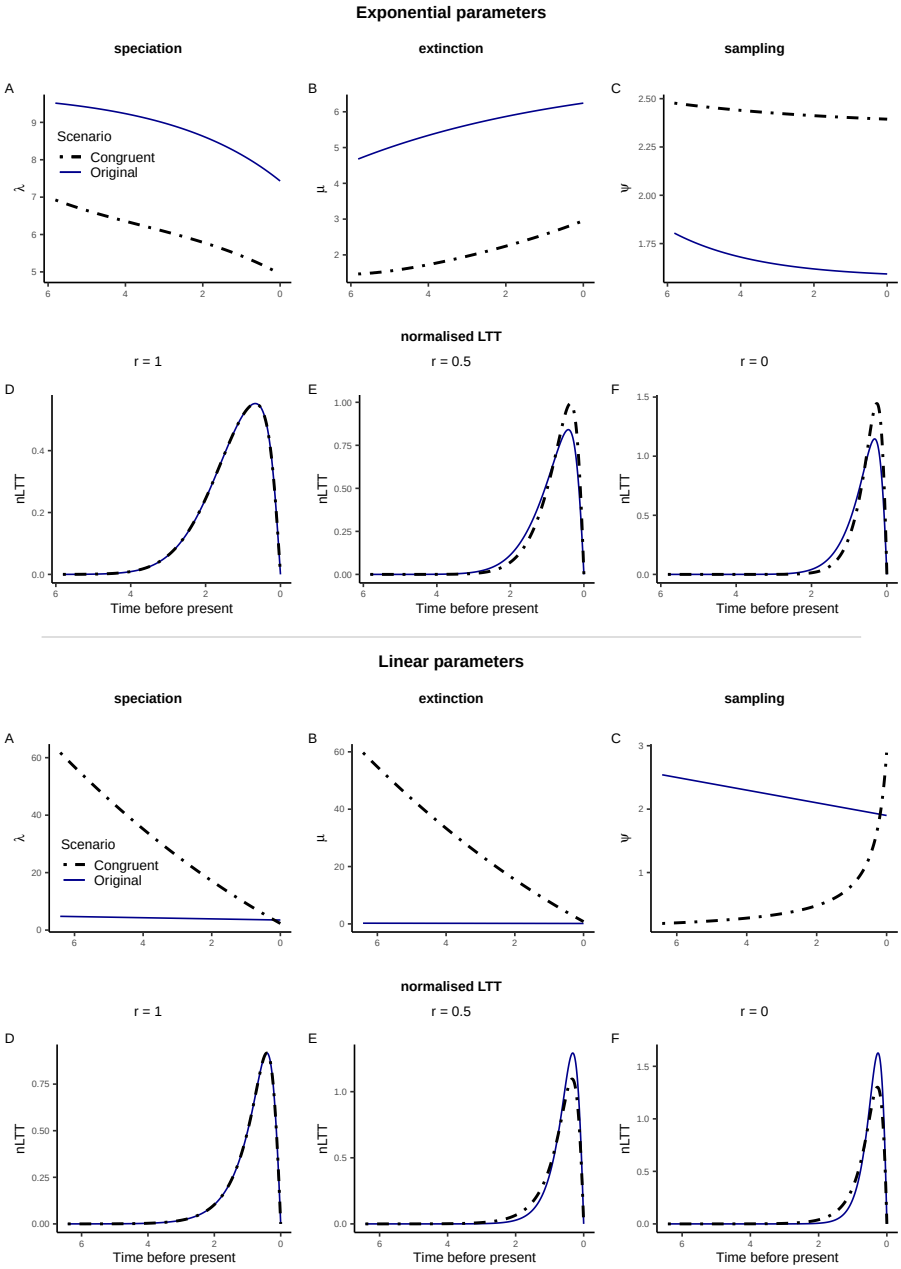


FIGURE 4. Two sets of congruent scenarios, one where θ_1 has exponential rates, and one where θ_1 has linear rates. In each set of scenarios, subfigures A–C show the parameters of θ_1 and θ_2. In both cases, subfigures D–F illustrate that when sampled ancestors are permitted, θ_1 and θ_2 no longer produce identical nLTT curves.

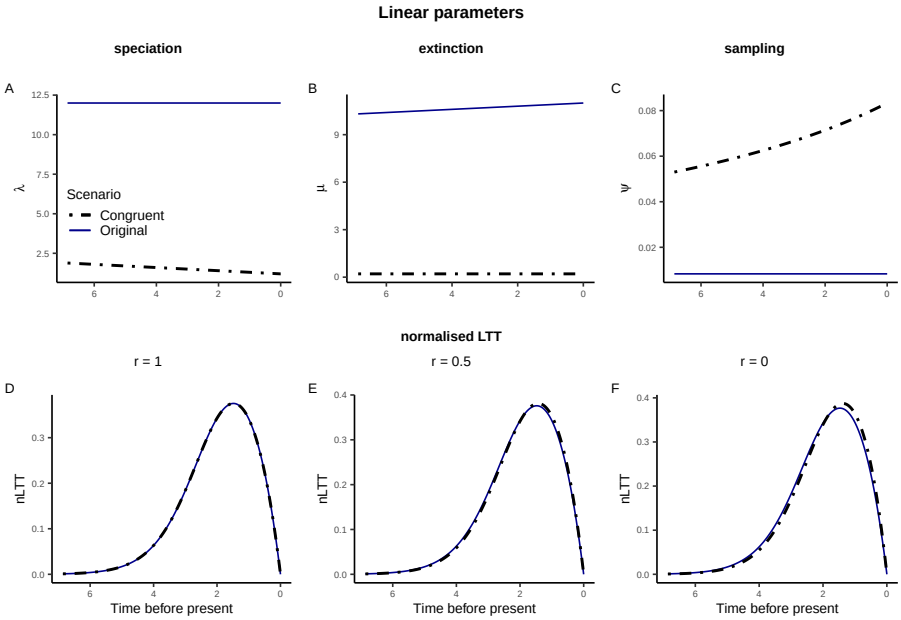


FIGURE 5. A set of congruent scenarios, where θ_1 has linear rates. Subfigures A–C show the parameters of $\theta_1 = \left(12, -0.1t + 11, \frac{1}{120}\right)$ and $\theta_2 = \left(0.1t + 1, \frac{1}{120}, \frac{0.1}{0.1t+1}\right)$. Subfigures D–F illustrate that although reducing the removal probability results in different nLTT curves for θ_1 and θ_2, the difference in density is small for all times considered.

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 [22] (a tree in which fossil samples can be sampled ancestors). Including sampled ancestor fossils has been shown to be very important in reducing biases in estimated extinction rates [22, 35]. Here, we have shown that allowing fossils to be sampled ancestors is crucial for theoretical identifiability of the diversification rates. An additional benefit of including sampled through time data in the analysis is that the sampling rate itself can be inferred from such data, even if the sampling scheme was not constant through time. 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 initial rates from pulled rates. Given that species' physical structure or habitat may make them unsuitable for fossilisation [36], 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 ψ is equal to zero we can apply the results of Legried and Terhorst [19, 20] to obtain the identifiability of the reconstructed FBD model.

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 [24], as has been previously shown for piecewise constant rates [22]. Here we have also shown that the time-dependent birth-death sampling model remains identifiable, provided r is fixed to a probability 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 [34] or explore congruence classes for patterns that are conserved within them [37, 38].

By defining pulled rates introduced for birth-death sampling models [24] for the FBD model, we discovered 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. The initial birth rate is adjusted by the probability of the newly born lineage to be observed. The initial 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. Note that this property extends the pulled rates

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 [39], age-dependent rates [40], or different speciation types [41]. While the identifiability of the pure birth multi-type model with some restrictions on the transition rates has been recently shown [42], 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 ψ 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.

Materials and Methods. To plot nLTT curves under congruent scenarios, Louca et al. [24] randomly selected diversification rates, which were generated under either an exponential or an Ornstein-Uhlenbeck stochastic process.

Here, we randomly select exponential diversification rates, as in Louca et al. [24], while selecting linear diversification rates manually to avoid technical difficulties with obtaining congruent scenarios with positive rates. We generate a tree with between 100,000 and 200,000 tips, given θ_1. This provides us with tree properties necessary to compare deterministic properties for different removal probabilities over a consistent time period and extant lineages. We use the castor function **congruent_hbds_model** and calculations based on the pulled diversification rate to obtain θ_2 values, and then calculate deterministic values using the **simulate_deterministic_hbds** function. Rather than utilising the removal probability r, the package functions use the retention probability κ, where $\kappa = 1 - r$.

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.

Code availability. The R code from Louca et al. [24] 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.

Acknowledgements. We thank Dmitry Berdinsky, Lars Berling and Kerry Manson for their helpful discussions throughout this research. 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). KT was supported by UC Aho Hīnātore, UC Accelerator Scholarship.

References

1. Raup, D. M., Gould, S. J., Schopf, T. J. & Simberloff, D. S. Stochastic models of phylogeny and the evolution of diversity. *The Journal of Geology* **81,** 525–542 (1973).
2. Nee, S., Holmes, E. C., May, R. M., *et al.* Estimating extinction from molecular phylogenies. *Extinction* (1995).
3. Heath, T. A., Huelsenbeck, J. P. & Stadler, T. The fossilized birth–death process for coherent calibration of divergence-time estimates. *Proceedings of the National Academy of Sciences* **111,** E2957–E2966 (2014).
4. Nee, S. Birth-death models in macroevolution. *Annu. Rev. Ecol. Evol. Syst.* **37,** 1–17 (2006).
5. Silvestro, D., Salamin, N. & Schnitzler, J. PyRate: a new program to estimate speciation and extinction rates from incomplete fossil data. *Methods in Ecology and Evolution* **5,** 1126–1131 (2014).
6. Morlon, H., Parsons, T. L. & Plotkin, J. B. Reconciling molecular phylogenies with the fossil record. en. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 16327–16332 (Sept. 2011).

7. Stadler, T., Kouyos, R., von Wyl, V., Yerly, S., Böni, J., Bürgisser, P., Klimkait, T., Joos, B., Rieder, P., Xie, D., Günthard, H. F., Drummond, A. J., Bonhoeffer, S. & Swiss HIV Cohort Study. Estimating the basic reproductive number from viral sequence data. en. *Mol. Biol. Evol.* **29,** 347–357 (Jan. 2012).
8. Stadler, T., Kühnert, D., Bonhoeffer, S. & Drummond, A. J. Birth–death skyline plot reveals temporal changes of epidemic spread in HIV and hepatitis C virus (HCV). *Proceedings of the National Academy of Sciences* **110,** 228–233 (2013).
9. Feller, W. Die grundlagen der volterraschen theorie des kampfes ums dasein in wahrscheinlichkeitstheoretischer behandlung. *Acta Biotheoretica* **5,** 11–40 (1939).
10. Kendall, D. G. On the generalized" birth-and-death" process. *The annals of mathematical statistics* **19,** 1–15 (1948).
11. Yang, Z. & Rannala, B. Bayesian phylogenetic inference using DNA sequences: a Markov Chain Monte Carlo Method. en. *Mol. Biol. Evol.* **14,** 717–724 (July 1997).
12. Höhna, S., Landis, M. J., Heath, T. A., Boussau, B., Lartillot, N., Moore, B. R., Huelsenbeck, J. P. & Ronquist, F. RevBayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification language. *Systematic biology* **65,** 726–736 (2016).
13. Bouckaert, R., Vaughan, T. G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., Heled, J., Jones, G., Kühnert, D., De Maio, N., *et al.* BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS computational biology* **15,** e1006650 (2019).
14. Stadler, T. On incomplete sampling under birth-death models and connections to the sampling-based coalescent. en. *J. Theor. Biol.* **261,** 58–66 (Nov. 2009).
15. Stadler, T. Mammalian phylogeny reveals recent diversification rate shifts. en. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 6187–6192 (Apr. 2011).
16. Louca, S. & Pennell, M. W. A General and Efficient Algorithm for the Likelihood of Diversification and Discrete-Trait Evolutionary Models. en. *Syst. Biol.* **69,** 545–556 (May 2020).
17. Louca, S. & Pennell, M. W. Extant timetrees are consistent with a myriad of diversification histories. en. *Nature* **580,** 502–505 (Apr. 2020).
18. Pagel, M. Evolutionary trees can't reveal speciation and extinction rates. *Nature* **580,** 461–462 (Apr. 2020).
19. Legried, B. & Terhorst, J. A class of identifiable phylogenetic birth-death models. en. *Proc. Natl. Acad. Sci. U. S. A.* **119,** e2119513119 (Aug. 2022).
20. Legried, B. & Terhorst, J. Identifiability and inference of phylogenetic birth–death models. *J. Theor. Biol.* **568,** 111520 (July 2023).
21. Stadler, T. Sampling-through-time in birth–death trees. *J. Theor. Biol.* **267,** 396–404 (Dec. 2010).
22. Gavryushkina, A., Welch, D., Stadler, T. & Drummond, A. J. Bayesian inference of sampled ancestor trees for epidemiology and fossil calibration. *PLoS Comput. Biol.* **10,** e1003919 (Dec. 2014).
23. MacPherson, A., Louca, S., McLaughlin, A., Joy, J. B. & Pennell, M. W. Unifying Phylogenetic Birth-Death Models in Epidemiology and Macroevolution. en. *Syst. Biol.* **71,** 172–189 (Dec. 2021).

24. Louca, S., McLaughlin, A., MacPherson, A., Joy, J. B. & Pennell, M. W. Fundamental Identifiability Limits in Molecular Epidemiology. en. *Mol. Biol. Evol.* **38,** 4010–4024 (Aug. 2021).
25. Andréoletti, J. & Morlon, H. Exploring congruent diversification histories with flexibility and parsimony. en. *Methods Ecol. Evol.* (Oct. 2023).
26. Stadler, T., Gavryushkina, A., Warnock, R. C. M., Drummond, A. J. & Heath, T. A. The fossilized birth-death model for the analysis of stratigraphic range data under different speciation modes. en. *J. Theor. Biol.* **447,** 41–55 (June 2018).
27. Wright, A. M., Bapst, D. W., Barido-Sottani, J. & Warnock, R. C. M. Integrating Fossil Observations Into Phylogenetics Using the Fossilized Birth–Death Model. *Annu. Rev. Ecol. Evol. Syst.* **53,** 251–273 (Nov. 2022).
28. Nee, S., May, R. M. & Harvey, P. H. The reconstructed evolutionary process. en. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **344,** 305–311 (May 1994).
29. Steel, M. *Phylogeny: Discrete and Random Processes in Evolution* en (SIAM-Society for Industrial and Applied Mathematics, Sept. 2016).
30. Gavryushkin, A. & Drummond, A. J. The space of ultrametric phylogenetic trees. en. *J. Theor. Biol.* **403,** 197–208 (Aug. 2016).
31. Gavryushkin, A., Whidden, C. & Matsen 4th, F. A. The combinatorics of discrete time-trees: theory and open problems. en. *J. Math. Biol.* **76,** 1101–1121 (Apr. 2018).
32. Kubo, T. & Iwasa, Y. Inferring the rates of branching and extinction from molecular phylogenies. en. *Evolution* **49,** 694–704 (Aug. 1995).
33. Louca, S. & Doebeli, M. Efficient comparative phylogenetics on large trees. en. *Bioinformatics* **34,** 1053–1055 (Mar. 2018).
34. Morlon, H., Robin, S. & Hartig, F. Studying speciation and extinction dynamics from phylogenies: addressing identifiability issues. en. *Trends Ecol. Evol.* **37,** 497–506 (June 2022).
35. O'Reilly, J. E. & Donoghue, P. C. J. The Effect of Fossil Sampling on the Estimation of Divergence Times with the Fossilized Birth–Death Process. en. *Syst. Biol.* **69,** 124–138 (May 2019).
36. Shaw, J. O., Briggs, D. E. G. & Hull, P. M. Fossilization potential of marine assemblages and environments. *Geology* **49,** 258–262 (Mar. 2021).
37. Höhna, S., Kopperud, B. T. & Magee, A. F. CRABS: Congruent rate analyses in birth–death scenarios. en. *Methods Ecol. Evol.* **13,** 2709–2718 (Dec. 2022).
38. Kopperud, B. T., Magee, A. F. & Höhna, S. Rapidly changing speciation and extinction rates can be inferred in spite of nonidentifiability. en. *Proc. Natl. Acad. Sci. U. S. A.* **120,** e2208851120 (Feb. 2023).
39. Maddison, W. P., Midford, P. E. & Otto, S. P. Estimating a binary character's effect on speciation and extinction. *Systematic biology* **56,** 701–710 (2007).
40. Hagen, O., Hartmann, K., Steel, M. & Stadler, T. Age-dependent speciation can explain the shape of empirical phylogenies. *Systematic biology* **64,** 432–440 (2015).
41. Stadler, T., Gavryushkina, A., Warnock, R. C., Drummond, A. J. & Heath, T. A. The fossilized birth-death model for the analysis of stratigraphic range data under different speciation modes. *Journal of theoretical biology* **447,** 41–55 (2018).

42. Dragomir, D., Allman, E. S. & Rhodes, J. A. Parameter Identifiability of a Multitype Pure-Birth Model of Speciation. *Journal of Computational Biology* **30,** 277–292 (2023).

The Fossilised Birth-Death model is identifiable — Supporting Information

Kate Truman, Timothy G Vaughan, Alex Gavryushkin, and Alexandra "Sasha" Gavryushkina

February 9, 2024

S1 Preliminaries

For convenience, we restate the notation used in the main article.

S1.1 Rate and present sampling probability parameters

The time dependent speciation, extinction, and sampling rates of the FBD process are denoted $\lambda(t)$, $\mu(t)$, and $\psi(t)$ respectively. The probability that an extant lineage is sampled, known as the sampling fraction, is ρ_0. We denote $\theta = (\lambda(t), \mu(t), \psi(t), \rho_0)$. The measure of a set X is denoted $\boldsymbol{\mu}(X)$ to differentiate measures from extinction rates.

S1.2 Complete tree

The realisation of the time-dependent FBD process is a *complete* tree $\mathcal{T}$. This tree has N_0 present day taxa which are each sampled at present. The n death events occur at times $(y_1, y_2, \dots y_n)$, the $N_0 + n - 1$ branching events at $(x_1, x_2, \dots x_{N_0+n-1})$, and the m sampling events at $(z_1, z_2, \dots z_m)$. The first branching event at x_1 is called the root of the tree. A tree $\mathcal{T}$ can be split in its discrete and continuous components and can be written as a pair $(T, \bar{t})$ where T is the ranked ordered tree topology 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 $t_0 > t_1 > \dots > t_{N_0+2n+m-1} > 0$.

The probability density function for the complete trees under the time-dependent FBD process is:

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

where $\Psi(t) = e^{-\int_0^t \lambda(u)+\mu(u)+\psi(u)\, du}$ (Figure S1). The function can be derived using the unified model approach from MacPherson et al.[1].

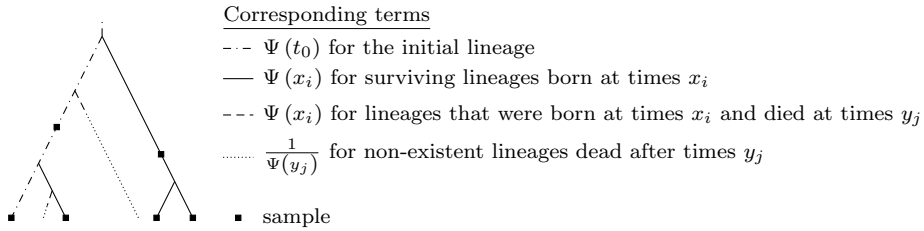


Figure S1: Complete tree integral mapping

S1.3 Reconstructed tree

In the 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.

We define the function $E(t)$, the probability of a lineage alive at time t to not leave sampled descendants. This function is described by the differential equation [1]:

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

with initial condition $E(0) = 1 - \rho_0$. 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.

Denote

$$\Phi(t) = e^{\int_0^t 2\lambda(u)E(u) - \lambda(u) - \mu(u) - \psi(u)\, du}$$

Let M_0 be the number of extant species sampled. Then the probability density function of reconstructed trees conditioning on (S), the event of sampling at least one lineage, is [1]:

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

S2 Probability of not observing a lineage

The probability density function for the reconstructed trees (3) includes function E, which is defined as a solution to the differential equation (2). Therefore we can only consider the distribution of reconstructed trees when such a function

exists and is unique. We first formulate conditions for the rates that guarantee that any extended solution of equation (2) is bounded.

Lemma S2.1. Let $\lambda \geq 0$, $\mu \geq 0$, and $\psi \geq 0$ be Lebesgue-integrable on $[0, t_0]$, then any extended solution $E(t)$ of equation (2) will satisfy: $E(t) \geq 0$ on $[0, t_0]$.

Proof. Let $F(t) = \frac{E(t)}{\Psi(t)}$, taking the derivative of F and using equation (2):

$$\frac{\mathrm{d}F}{\mathrm{d}t} = \lambda(t)\Psi(t)F^2 + \frac{\mu(t)}{\Psi(t)} \text{ for a.a. } t \in [0, t_0]$$

with $F(0) = 1 - \rho_0$. Note that the derivative of F is non-negative for all $t \in [0, t_0]$ where it exists. Given that $F(0) \geq 0$ we have that $F(t) \geq 0$ for all $t \in [0, t_0]$. Therefore $E(t) \geq 0$ for all $t \in [0, t_0]$. □

Lemma S2.2. Let $\rho_0 > 0$,

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

then for any extended solution E: $E(t) < 1$ on $[0, t_0]$.

In the case when $\rho_0 = 0$, additional conditions:

- $\psi(0) > 0$
- λ, μ, and ψ are continuous in zero

guarantee $E(t) < 1$ on $(0, t_0]$.

Proof. Let $H(t) = [1 - E(t)]e^{-\int_0^t \lambda - \mu - \psi}$ then

$$\frac{\mathrm{d}H}{\mathrm{d}t} = -\lambda e^{\int_0^t (\lambda - \mu - \psi)} H^2 + \psi e^{-\int_0^t (\lambda - \mu - \psi)} \text{ for a.a. } t \in [0, t_0] \tag{4}$$

with $H(0) = \rho_0$. Note that $H(t) > 0$ implies $E(t) < 1$. We are going to prove that if $H(t^*) > 0$ for some t^* then $H(t) > 0$ for all $t \in [t^*, t_0]$.

Suppose there is $\hat{t} \in (t^*, t_0]$ such that $H(\hat{t}) < 0$ then due to continuity of H and $H(t^*) > 0$ there have to be $t' < t^*$ such that $H(t) > 0$ for all $t \in [t^*, t')$ and $H(t') = 0$. Then we have:

$$\frac{\mathrm{d}H}{\mathrm{d}t} \geq -\lambda e^{\int_0^t (\lambda - \mu - \psi)} H^2 \text{ for a. a. } t \in [t^*, t')$$

$$-\frac{\mathrm{d}\frac{1}{H}}{\mathrm{d}t} \geq -\lambda e^{\int_0^t (\lambda - \mu - \psi)} \text{ for a. a. } t \in [t^*, t')$$

$$\frac{\mathrm{d}\frac{1}{H}}{\mathrm{d}t} \leq \lambda e^{\int_0^t (\lambda - \mu - \psi)} \text{ for a. a. } t \in [t^*, t')$$

$$\frac{1}{H(t)} - \frac{1}{H(t^*)} \leq \int_{t^*}^t \lambda(s) e^{-\int_0^s \lambda(u) - \mu(u) - \psi(u) du} ds \text{ for a. a. } t \in [t^*, t')$$

$$H(t) \geq \frac{H(t^*)}{1 + H(t^*)\int_{t^*}^{t} \lambda(s) e^{-\int_0^s \lambda(u)-\mu(u)-\psi(u)du} ds} \text{ for a. a. } t \in [t^*, t')$$

$$H(t) \geq \frac{H(t^*)}{1 + H(t^*)\int_{t^*}^{t'} \lambda(s) e^{-\int_0^s \lambda(u)-\mu(u)-\psi(u)du} ds} = c > 0 \text{ for a. a. } t \in [t^*, t')$$

Due to the continuity of H and $H(t') = 0$, there has to be a neighbourhood of t' where H is smaller than c which contradicts the above inequality. Therefore $H(t) > 0$ for all $t \in [t^*, t_0]$.

Now if $\rho_0 > 0$, then $H(0) > 0$, and therefore $H(t) > 0$ for all $t \in [0, t_0]$. If $\rho_0 = 0$ and λ, μ and ψ are continuous at zero, then E is differentiable at zero and therefore H is differentiable at zero. Furthermore due to $\psi(0) > 0$, the derivative of H is strictly positive at zero. This implies that in the neighbourhood of zero H increases and therefore $H(t) > 0$ in some neighbourhood of zero, which again implies that $H(t) > 0$ for all $t \in (0, t_0]$. □

We now show that the conditions of lemma S2.2 are sufficient for the existence and uniqueness of the solution to equation (2). Consider

$$f(t, E) = \lambda(t)E^2 - (\lambda(t) + \mu(t) + \psi(t))\, E + \mu(t)$$

on $[0, t_0) \times (0, 1)$. f is Lebesgue-integrable in t for any fixed E and continuous in E for any fixed t. From $E(t) \leq 1$ we have:

$$|f(t, E)| \leq \lambda(t)E^2 + (\lambda(t) + \mu(t) + \psi(t))\, E + \mu(t) \leq$$

$$2\lambda(t) + 2\mu(t) + \psi(t) \text{ for } (t, E) \in [0, t_0) \times (0, 1)$$

As the right-hand side of the inequality is Lebesgue-integrable, f satisfies the conditions of the Carathéodory's theorem and therefore there is a local solution (in the extended sense) to equation (2), which can be extended on the whole interval $[0, t_0]$ due to E being bounded.

f is locally lipschitzian on $[0, t_0) \times (0, 1)$:

$$|f(t, E_1) - f(t, E_2)| = \Big|\big[\lambda(t)(E_1 + E_2) - (\lambda(t) + \mu(t) + \psi(t))\big](E_1 - E_2)\Big| \leq$$

$$(2\lambda(t) + \mu(t) + \psi(t))|E_1 - E_2|$$

for any $t \in [0, t_0]$ with $2\lambda(t) + \mu(t) + \psi(t)$ being Lebesgue-integrable. Therefore the solution is unique.

The conditions of lemma S2.2 guarantee that we can define a probability distribution over the reconstructed trees. We could not relax these conditions and did not attempt to show that they are necessary conditions. Milder conditions could possibly be found.

S3 Elements of complete FBD identifiability proof

S3.1 Probability space on trees

We first define a measure on the space of trees. Consider a tree $\mathcal{T}$ that is obtained under the time-dependent birth-death model without present sampling (guaranteed complete sampling) or sampling through time. Its discrete structure is represented by a ranked oriented topology [2]. Ranked means that each branching or tip node has a distinct ranking, except the extant tips, which all have rank $N_0 + n$, and the ranking is consistent with the ranking induced by ancestor-descendant relationship in the tree. Oriented means that every branch after a branching event is designated by a label *left* or *right*. Each ranked node is mapped to a time point on the interval $[0, t_0]$ preserving the ranking. Thus, $\mathcal{T} = (T, t_1, \ldots, t_{N_0+n-1})$, where a node with rank i has time t_i. In the notation introduced above, $t_i = x_{\pi_x(i)}$ and $t_i = y_{\pi_y(i)}$, with corresponding functions π_x and π_y that map indices of branching and death events to the rankings on the tree. Thus, we can see the space of trees as a disjoint union of sets $S_T = \{(t_1, \ldots, t_{N_0+n-1}) \mid t_0 > t_1 > \cdots > t_{N_0+n-1} > 0\}$ for each distinct ranked topology T. Then the probability measure of a set $\mathbb{T}$ of trees is $\mu(\mathbb{T}) = \sum_{T \in \mathbf{T}} \mu(S_T \bigcap \mathbb{T})$ where $\mathbf{T}$ is the set of all ranked topologies.

S3.2 Parameter constraints for equal probability densities

The identifiability of the time-dependant FBD model for complete trees relies on Lemma 3.1 from the main text. Here we provide a more detailed prove of the lemma.

Proof. Remember that t_0 is fixed, $\theta_1 = (\lambda_1, \mu_1, \psi_1, t_0)$, and $\theta_2 = (\lambda_2, \mu_2, \psi_2, t_0)$. Assume that for all ranked topologies T,

$$f^{\text{compl.}}(T, \bar{t}|\theta_1) = f^{\text{compl.}}(T, \bar{t}|\theta_2)$$

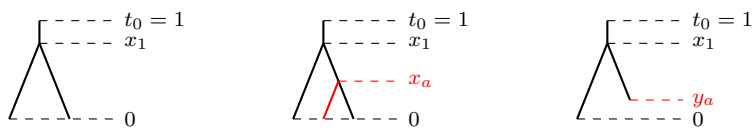
for all $\bar{t} \in S_T \setminus \mathbf{0}_T$ where $\boldsymbol{\mu}(\mathbf{0}_T) = 0$.

Consider all trees with a single birth event and no death or sampling events. An example of such a tree is given in Figure S2a. All such trees will have the same ranked topology G and only differ by the time of the branching event x_1. For each such tree (G, x_1) we define a tree which is produced by adding one extant lineage at a time point x_a after the first birth event on the right branch, e.g., $x_a < x_1$ (Figures S2b). Denote the obtained tree (G^+, x_1, x_a). We also define a tree (G^*, x_1, y_a) (Figures S2c) formed by contracting the rigth extant edge in (G, x_1) to become extinct at time y_a after the birth event.

Note $S_G = [0, t_0]$ and $S_{G^+} = \{(x_1, x_a) \mid t_0 > x_1 > x_a > 0\}$. Since the probability densities for model 1 with parameters θ_1 and model 2 with parameters θ_2 are equal almost everywhere, we have:

$$f^{\text{compl.}}(G, x_1|\theta_1) = f^{\text{compl.}}(G, x_1|\theta_2)$$

for all $x_1 \in S_G \setminus \mathbf{0}_G$ where $\boldsymbol{\mu}(\mathbf{0}_G) = 0$.



(a) G (b) G^+ (c) G^*

Figure S2: Tree variants. Subfigure (a) depicts the original tree G, while Subfigure (b) shows the addition of an extant edge at time x_a and Subfigure (c) shows the contraction of an extant edge such that it dies at time y_a.

We also have:

$$f^{\text{compl.}}(G^+, x_1, x_a|\theta_1) = f^{\text{compl.}}(G^+, x_1, x_a|\theta_2)$$

for all $(x_1, x_a) \in S_{G^+} \setminus \mathbf{0}_{G^+}$ and $\boldsymbol{\mu}(\mathbf{0}_{G^+}) = 0$.

Using equation (1) for the probability density of complete trees, we can rewrite the last two equations as follows:

$$\Psi_1(t_0)\lambda_1(x_1)\Psi_1(x_1) = \Psi_2(t_0)\lambda_2(x_1)\Psi_2(x_1)$$
$$\Psi_1(t_0)\lambda_1(x_1)\Psi_1(x_1)\lambda_1(x_a)\Psi_1(x_a) = \Psi_2(t_0)\lambda_2(x_1)\Psi_2(x_1)\lambda_2(x_a)\Psi_2(x_a)$$

Therefore

$$\lambda_1(x_a)\,\Psi(x_a) = \lambda_2(x_a)\,\Psi(x_a) \text{ for all } x_a \in [0, t_0] \setminus \mathbf{0}_1 \tag{5}$$

where

$$\mathbf{0}_1 = \{x \mid \text{ for all } x' \in (x, t_0] \setminus \mathbf{0}_G : (x', x) \in \mathbf{0}_{G^+}\}$$

Now we show that $\boldsymbol{\mu}(\mathbf{0}_1) = 0$. Suppose $\boldsymbol{\mu}(\mathbf{0}_1) > 0$ then for some $b < t_0$: $\boldsymbol{\mu}(\mathbf{0}_1 \cap [0, b]) > 0$. The product of non-zero measure sets is non-zero: $\boldsymbol{\mu}((\mathbf{0}_1 \cap [0, b]) \times ((b, t_0] \setminus \mathbf{0}_G)) > 0$. Note that $(\mathbf{0}_1 \cap [0, b]) \times ((b, t_0] \setminus \mathbf{0}_G) \subset \mathbf{0}_{G^+}$, and therefore $\boldsymbol{\mu}(\mathbf{0}_{G^+}) > 0$, which contradicts the definition of $\mathbf{0}_{G^+}$, therefore $\boldsymbol{\mu}(\mathbf{0}_1) = 0$.

Applying a similar argument to trees G and G^* we obtain:

$$\frac{\mu_1(y_a)}{\Psi(y_a)} = \frac{\mu_2(y_a)}{\Psi(y_a)} \text{ for all } y_a \in [0, t_0] \setminus \mathbf{0}_2 \tag{6}$$

Then (5) and (6) imply:

$$\lambda_1(t)\mu_1(t) = \lambda_2(t)\mu_2(t) \text{ for all } t \in [0, t_0] \setminus (\mathbf{0}_1 \cup \mathbf{0}_2) \tag{7}$$

A similar argument applied to trees where one sampling event is added after the birth event on one of the branches would lead to $\psi_1(t) = \psi_2(t)$ for all $t \in [0, t] \setminus \mathbf{0}_3$ with $\boldsymbol{\mu}(\mathbf{0}_3) = 0$.

□

S3.3 Uniqueness of gamma

Denote $\gamma(t) = \frac{\Psi_1(t)}{\Psi_2(t)}$. Note that due to λ_1, λ_2, μ_1, μ_2, ψ_1 and ψ_2 being positive and Lebesgue-integrable on $[0, t_0]$, $\gamma(t)$ is continuous on the interval $[0, t_0]$ and therefore it is bounded by some constant K on this interval.

The parameter constraints (see Lemma (3.1) in the main text) imply that:

$$\begin{aligned}
\lambda_2(t) &= \gamma(t)\,\lambda_1(t) \text{ for a. a. } t \in [0, t_0] \\
\mu_2(t) &= \frac{\mu_1(t)}{\gamma(t)} \text{ for a. a. } t \in [0, t_0] \\
\psi_2(t) &= \psi_1(t) \text{ for a. a. } t \in [0, t_0]
\end{aligned}$$

Using these equations and the definition of γ we obtain:

$$\gamma(t) = e^{-\int_0^t \lambda_1 + \mu_1 - \gamma\lambda_1 - \frac{\mu_1}{\gamma}\, du} \text{ for a. a. } t \in [0, t_0]$$

Taking the derivative of both sides we have:

$$\frac{d\gamma}{dt} = \lambda_1\gamma^2 - (\lambda_1 + \mu_1)\gamma + \mu_1 \text{ for a. a. } t \in [0, t_0]$$

From the definition of γ we obtain the initial condition $\gamma(0) = 1$. Moreover, from Equation (7) in the main text we have that $\gamma(t_0) = 1$. An extended solution to this equation is $\gamma(t) = 1$.

The following inequality holds for the right hand side of the equation:

$$|\lambda_1\gamma_1^2 - (\lambda_1+\mu_1)\gamma_1 + \mu_1 - (\lambda_1\gamma_2^2 - (\lambda_1+\mu_1)\gamma_2 + \mu_1)| \leq |\gamma_1 + \gamma_2 - \lambda_1 - \mu_1||\gamma_1 - \gamma_2| \leq$$

$$(2K_U + \lambda_1 + \mu_1)|\gamma_1 - \gamma_2|$$

where $K = sup_{t \in U}\gamma(t)$ for a given compact set U. Since $2K + \lambda_1 + \mu_1$ is Lebesgue-integrable the solution is unique.

S4 Transformed (pulled) rates

In the main text, we introduced a parameter transformation (repeated from equation (9) in the main text) that defines pulled rates:

$$\begin{aligned}
\tilde{\lambda} &= \lambda(1 - E) \\
\tilde{\mu} &= \frac{\psi E}{1 - E} \\
\tilde{\psi} &= \psi
\end{aligned} \tag{8}$$

Note that in the conditions of lemma S2.2, $E(t) > 1$ and $\tilde{\mu}$ is defined for all $t \in (0, t_0]$.

S4.1 Reconstructed and complete tree distributions

In this section, we show that the distribution of reconstructed trees is equivalent to the distribution of complete trees generated by the FBD process with transformed parameters.

Let $f^{\text{compl.}}(\mathcal{T}|\tilde{\lambda},\tilde{\mu},\tilde{\psi},t_0)$ be the probability density of complete trees under the time-dependent FBD model with $\tilde{\rho}_0 = 1$. Let $f^{\text{rec.}}(\mathcal{T}|\lambda,\mu,\psi,\rho_0,t_0,S)$ be the probability density of reconstructed trees under the time-dependent FBD model, conditioned on sampling of at least one lineage (S).

The probability density of the complete trees under the pulled rates is (in what follows we omit t_0 to simplify notation):

$$\begin{aligned} f^{\text{compl.}}\left(\mathcal{T} \mid \tilde{\lambda},\tilde{\mu},\tilde{\psi}\right) =& e^{-\int_0^{t_0}\tilde{\lambda}(u)+\tilde{\mu}(u)+\tilde{\psi}(u)\,du} \\ &\times \prod_{i=1}^{M_0+n-1}\tilde{\lambda}\left(x_i\right)e^{-\int_0^{x_i}\tilde{\lambda}(u)+\tilde{\mu}(u)+\tilde{\psi}(u)\,du} \\ &\times \prod_{j=1}^{n}\tilde{\mu}\left(y_j\right)e^{\int_0^{y_j}\tilde{\lambda}(u)+\tilde{\mu}(u)+\tilde{\psi}(u)\,du}\prod_{k=1}^{m}\tilde{\psi}\left(z_k\right) \end{aligned}$$

Substituting the values for the pulled rates in the exponential terms we obtain:

$$\begin{aligned} e^{-\int_0^t\tilde{\lambda}+\tilde{\mu}+\tilde{\psi}\,dt} =& e^{-\int_0^t\lambda-\lambda E+\frac{\psi E}{1-E}+\psi\,dt} \\ =& e^{-\int_0^t\lambda E-\mu+\frac{\psi E}{1-E}\,dt}\left(e^{-\int_0^t\lambda+\mu+\psi-2\lambda E\,dt}\right) \end{aligned}$$

Assuming $\rho_0 > 0$ and using equation (2), we have:

$$\begin{aligned} e^{\int_0^t\lambda E-\mu+\frac{\psi E}{1-E}\,dt} =& e^{\int_0^t\frac{\frac{dE}{dt}}{E-1}\,du} = e^{\ln|E(t)-1|-\ln|E(0)-1|} \\ =& e^{\ln\frac{1-E(t)}{\rho_0}} = \frac{1-E\left(t\right)}{\rho_0} \end{aligned}$$

which implies:

$$e^{-\int_0^t\tilde{\lambda}+\tilde{\mu}+\tilde{\psi}\,dt} = \frac{\rho_0}{1-E\left(t\right)}\left(e^{-\int_0^t\lambda+\mu+\psi-2\lambda E\,dt}\right)$$

We thus re-write the complete tree probability density using exponential terms of the same form as the reconstructed tree probability density.

$$\begin{aligned} f^{\text{compl.}}\left(\mathcal{T} \mid \tilde{\lambda},\tilde{\mu},\tilde{\psi}\right) =& \frac{\rho_0}{1-E\left(t_0\right)}e^{-\int_0^{t_0}\lambda+\mu+\psi-2\lambda E\,du} \\ &\times \prod_{i=1}^{M_0+n-1}\tilde{\lambda}\left(x_i\right)\frac{\rho_0}{1-E\left(x_i\right)}e^{-\int_0^{x_i}\lambda+\mu+\psi-2\lambda E\,du} \\ &\times \prod_{j=1}^{n}\tilde{\mu}\left(y_j\right)\frac{\left(1-E\left(y_j\right)\right)}{\rho_0}e^{\int_0^{y_j}\lambda+\mu+\psi-2\lambda E\,du}\prod_{k=1}^{m}\tilde{\psi}\left(z_k\right) \end{aligned}$$

We substitute the pulled rates and use Equation (3) in the main article:

$$
\begin{aligned}
&f^{\text{compl.}}\left(\mathcal{T} \mid \tilde{\lambda}, \tilde{\mu}, \tilde{\psi}\right) = \\
&\frac{\rho_0}{1-E\left(t_0\right)} e^{-\int_0^{t_0} \lambda+\mu+\psi-2\lambda E \, du} \\
&\times \prod_{i=1}^{M_0+n-1} \lambda\left(x_i\right)\left(1-E\left(x_i\right)\right) \frac{\rho_0}{1-E\left(x_i\right)} e^{-\int_0^{x_i} \lambda+\mu+\psi-2\lambda E \, du} \\
&\times \prod_{j=1}^{n}\left(\psi\left(y_j\right) \frac{E\left(y_j\right)}{\rho_0}\right) e^{\int_0^{y_j} \lambda+\mu+\psi-2\lambda E \, du} \prod_{k=1}^{m} \psi\left(z_k\right) \\
&= f^{\text{rec.}}\left(\mathcal{T} \mid \lambda, \mu, \psi, S\right)
\end{aligned}
$$

When $\rho_0 = 0$, similar derivations imply $e^{\int_y^x \lambda E-\mu+\frac{\psi E}{1-E} \, dt} = \frac{1-E(x)}{1-E(y)}$. Without present day samples, each branching event can be matched with a death event and exponential terms in both complete and reconstructed densities can be written as: $e^{-\int_{y_j}^{x_i} \tilde{\lambda}+\tilde{\mu}+\tilde{\psi} \, du}$ or $e^{-\int_{y_j}^{x_i} \lambda+\mu+\psi-2\lambda E \, du}$. Then substitution of pulled rates in the complete tree density will again produce the reconstructed tree density conditioned on sampling of at least one lineage. Note that in this border case, $E(0) = 1$ and the pulled birth rate becomes zero while pulled death rate becomes infinite at time zero. This guarantees there are no lineages surviving until present in the complete tree case.

S4.2 Connection to deterministic FBD process

A deterministic birth-death sampling model provides an equation for the lineage through time (LTT) plot curve based on the birth, death, and sampling rates. This LTT plot curve corresponds to the expected number of lineages through time in the stochastic birth-death sampling model [3] with the same rates and some sampling fraction. Considering the deterministic approach, the pulled rates describe the expected number of observed lineages (occurring in the reconstructed tree) in the corresponding stochastic model. In what follows we derive the pulled rates using the deterministic approach.

Probability of unobserved lineages

The pulled birth rate $\tilde{\lambda} = \lambda\left(1-E\right)$ has been derived in Louca and Pennell's unidentifiability result for reconstructed timetrees. This expression falls naturally from the definition of $E\left(t\right)$ as "the fraction of lineages extant at time t that will be missing from the timetree (either due to extinction or not having been sampled)" [4]. Then, only the birth events that produce surviving lineages will contribute to the pulled birth rate, hence the expression for this rate.

The differential equation (2) that describes E is not guaranteed to have an analytical solution, and therefore it is useful to also view $E\left(t\right)$ from an alternative perspective. Namely, $1-E\left(t\right)$ is also the percentage of lineages alive at time t

that we will find in the sample under the deterministic model based on birth, death and sampling rates. It is easy to define a pulled death rate similarly to the pulled birth rate, and produce an expression describing the number of observed lineages alive at time t, which allows the proportion of unobserved lineages to be calculated.

From Louca and Pennell [4], we have the expected number of lineages, both observed and unobserved, denoted $N(t)$. M_0 is the number of lineages sampled at the present and ρ_0 is the sampling fraction.

$$N(t) = \frac{M_0}{\rho_0} e^{\int_0^t \mu(u) - \lambda(u)\, du}$$

Let $V(t)$ be the number of observed lineages. Note that the number of observed lineages under the original model is equal to the number of all lineages in the complete trees under the model with pulled rates:

$$V(t) = M_0 e^{\int_0^t \tilde{\mu}(u) - \tilde{\lambda}(u)\, du}$$

We subtract the number of observed lineages from the number of overall lineages to obtain the number of unobserved lineages, $Q(t)$.

$$\begin{aligned} Q(t) =& N(t) - V(t) \\ =& \frac{M_0}{\rho_0} e^{\int_0^t \mu(u) - \lambda(u)\, du} - M_0 e^{\int_0^t \tilde{\mu}(u) - \tilde{\lambda}(u)\, du} \end{aligned}$$

The probability that a lineage will be missing from the reconstructed timetree is equal to the proportion of lineages at that time that are unobserved.

$$\begin{aligned} E(t) =& \frac{Q(t)}{N(t)} \\ =& 1 - \rho_0 e^{\int_0^t \tilde{\mu}(u) - \tilde{\lambda}(u) - \mu(u) + \lambda(u)\, du} \end{aligned} \tag{9}$$

Pulled death rate

From equation (9) and the expression for the pulled birth rate, we obtain:

$$E(t) = 1 - \rho_0 e^{\int_0^t \tilde{\mu}(u) + \lambda(u) E(u) - \mu(u)\, du}$$

By calculating the derivative of both sides and combining it with Equation (2), we find

$$\tilde{\mu} + \lambda E - \mu = \lambda E - \mu - \frac{\psi E}{E - 1}$$

which implies

$$\tilde{\mu} = \frac{\psi E}{1 - E}$$

Pulled sampling rate

As the sampling rate on observed lineages is the same as in the initial process, we set $\tilde{\psi} = \psi$.

S4.3 One-to-one correspondence between pulled and normal rates

Now we show that Equation (8) defines a one-to-one correspondence between sets of birth, death, and strictly positive sampling rates and corresponding pulled rates.

We say that two sets of rates $F_1 = (\lambda_1, \mu_1, \psi_1, \rho_{0,1})$ and $F_2 = (\lambda_2, \mu_2, \psi_2, \rho_{0,2})$ are not equal ($F_1 \neq F_2$) if there is a pair of rates (birth, death, or sampling) that are not equal everywhere on $[0, t_0]$ or $\rho_{0,1} \neq \rho_{0,2}$. We also write $\alpha_1 \overset{\text{a.e.}}{=} \alpha_2$ for functions that are equal almost everywhere on $[0, t_0]$ and $\alpha_1 \overset{\text{a.e.}}{\neq} \alpha_2$ for functions that differ on at least a non-zero measure subset of $[0, t_0]$.

Lemma S4.1. Let $F_1 = (\lambda_1, \mu_1, \psi_1, \rho_{0,1})$ and $F_2 = (\lambda_2, \mu_2, \psi_2, \rho_{0,2})$ and $F_1 \neq F_2$. Let $\tilde{F}_1 = \left(\tilde{\lambda}_1, \tilde{\mu}_1, \tilde{\psi}_1\right)$ and $\tilde{F}_2 = \left(\tilde{\lambda}_2, \tilde{\mu}_2, \tilde{\psi}_2\right)$ be the pulled rates obtained from F_1 and F_2 using Equation (8). Then $\tilde{F}_1 \neq \tilde{F}_2$.

Proof. We consider two cases.

Case 1: $\psi_1 \overset{\text{a.e.}}{\neq} \psi_2$

The case where $\psi_1 \overset{\text{a.e.}}{\neq} \psi_2$ is trivial, as we know $\tilde{\psi}_1 \overset{\text{a.e.}}{=} \psi_1$ and $\tilde{\psi}_2 \overset{\text{a.e.}}{=} \psi_2$, implying $\tilde{\psi}_1 \overset{\text{a.e.}}{\neq} \tilde{\psi}_2$ and therefore $\tilde{F}_1 \neq \tilde{F}_2$.

Case 2: $\psi_1 \overset{\text{a.e.}}{=} \psi_2$

We now consider the case where $\psi_1 \overset{\text{a.e.}}{=} \psi_2$. Suppose $\tilde{F}_1 = \tilde{F}_2$. Then from $\tilde{\mu}_1 \overset{\text{a.e.}}{=} \tilde{\mu}_2$:

$$\frac{\psi_1 E_1}{1 - E_1} \overset{\text{a.e.}}{=} \frac{\psi_2 E_2}{1 - E_2}$$

As we assume that $\psi_1(t), \psi_2(t) > 0$ for almost all $t \in [0, t_0]$, we have $E_1(t), E_2(t) < 1$ for all t except possibly for $t = 0$ in the case when $\rho_0 = 0$.

$$\begin{aligned}
\frac{1 - E_1}{E_1} &\overset{\text{a.e.}}{=} \frac{1 - E_2}{E_2} \\
\frac{1}{E_1} - 1 &\overset{\text{a.e.}}{=} \frac{1}{E_2} - 1 \\
E_1 &\overset{\text{a.e.}}{=} E_2
\end{aligned}$$

From continuity of E_1 and E_2 we have $E_1 = E_2$ for all $t \in [0, 1]$ and therefore $\rho_{0,1} = 1 - E_1(0) = 1 - E_2(0) = \rho_{0,2}$.

From $\tilde{\lambda}_1 = \tilde{\lambda}_2$:

$$\begin{aligned}
\lambda_1 (1 - E_1) &\overset{\text{a.e.}}{=} \lambda_2 (1 - E_2) \\
\lambda_1 (1 - E_1) &\overset{\text{a.e.}}{=} \lambda_2 (1 - E_1) \\
\lambda_1 &\overset{\text{a.e.}}{=} \lambda_2
\end{aligned}$$

From $E_1 = E_2$, we know that their derivatives exist almost everywhere and are also equal when they exist:

$$\begin{aligned}
\frac{\mathrm{d}E_1}{\mathrm{d}t} &\overset{\text{a.e.}}{=} \frac{\mathrm{d}E_2}{\mathrm{d}t} \\
-\left(\lambda_1 + \mu_1 + \psi_1\right) E_1 + \lambda_1 E_1^2 + \mu_1 &\overset{\text{a.e.}}{=} -\left(\lambda_2 + \mu_2 + \psi_2\right) E_2 + \lambda_2 E_2^2 + \mu_2 \\
-\left(\lambda_1 + \mu_1 + \psi_1\right) E_1 + \lambda_1 E_1^2 + \mu_1 &\overset{\text{a.e.}}{=} -\left(\lambda_1 + \mu_2 + \psi_1\right) E_1 + \lambda E_1^2 + \mu_2 \\
\mu_1 \left(1 - E_1\right) &\overset{\text{a.e.}}{=} \mu_2 \left(1 - E_1\right) \\
\mu_1 &\overset{\text{a.e.}}{=} \mu_2
\end{aligned}$$

This contradicts the assumption of $F_1 \neq F_2$, therefore $\tilde{F}_1 \neq \tilde{F}_2$ □

S4.4 Reverse transformation

Equation (8) (also equation (9) in the main text) defines the transformation of the FBD process rates to pulled rates. Here we derive the reverse transformation, where rates of the process producing reconstructed trees are defined through pulled rates (rates of the process that produces complete trees with the same distribution). Assume the pulled rates are non-negative for all t and $\tilde{\psi}(0) > 0$ for a. a. t. We also assume that μ and ψ are piecewise differentiable. Then the reverse transformation is:

$$\begin{aligned}
\lambda &= \frac{\tilde{\lambda}(\tilde{\mu} + \tilde{\psi})}{\tilde{\psi}} \\
\mu &= \frac{\mathrm{d}\frac{\tilde{\mu}}{\tilde{\psi}}}{\mathrm{d}t} \frac{\tilde{\psi}}{\tilde{\mu} + \tilde{\psi}} + \tilde{\mu} \frac{\tilde{\lambda} + \tilde{\psi}}{\tilde{\psi}} \\
\psi &= \tilde{\psi} \\
\rho_0 &= \frac{\tilde{\psi}(0)}{\tilde{\psi}(0) + \tilde{\mu}(0)}
\end{aligned} \tag{10}$$

Now we show that this is indeed the reverse transformation of the pulled rates into original rates. Firstly, $\tilde{\psi}$ determines ψ. To determine $\tilde{\lambda}$, we can write E in terms of the pulled death and sampling rates:

$$\begin{aligned}
E &= \frac{\tilde{\mu}}{\tilde{\psi}} \left(1 - E\right) \\
E &= \frac{\frac{\tilde{\mu}}{\tilde{\psi}}}{1 + \frac{\tilde{\mu}}{\tilde{\psi}}} \\
E &= \frac{\tilde{\mu}}{\tilde{\mu} + \tilde{\psi}}
\end{aligned}$$

and therefore:

$$\lambda = \frac{\tilde{\lambda}\left(\tilde{\mu}+\tilde{\psi}\right)}{\tilde{\psi}}$$

Furthermore as $\rho_0 = 1 - E(0)$, we obtain

$$\rho_0 = \frac{\tilde{\psi}(0)}{\tilde{\psi}(0)+\tilde{\mu}(0)}$$

To determine μ, we first write $\frac{\mathrm{d}E}{\mathrm{d}t}$ with respect to the pulled death and sampling rates.

$$\begin{aligned}
\frac{\mathrm{d}E}{\mathrm{d}t} =& \frac{\mathrm{d}\frac{\tilde{\mu}}{\tilde{\mu}+\tilde{\psi}}}{\mathrm{d}t} \\
=& \frac{\frac{\mathrm{d}\tilde{\mu}}{\mathrm{d}t}\left(\tilde{\mu}+\tilde{\psi}\right) - \mu\left(\frac{\mathrm{d}\tilde{\mu}}{\mathrm{d}t}+\frac{\mathrm{d}\tilde{\psi}}{\mathrm{d}t}\right)}{\left(\tilde{\mu}+\tilde{\psi}\right)^2} \\
=& \frac{\frac{\mathrm{d}\tilde{\mu}}{\mathrm{d}t}\tilde{\psi} - \mu\frac{\mathrm{d}\tilde{\psi}}{\mathrm{d}t}}{\left(\tilde{\mu}+\tilde{\psi}\right)^2} \\
=& \frac{\tilde{\psi}^2}{\left(\tilde{\mu}+\tilde{\psi}\right)^2}\frac{\mathrm{d}\frac{\tilde{\mu}}{\tilde{\psi}}}{\mathrm{d}t}
\end{aligned}$$

From equation (2) and $\tilde{\lambda}+\tilde{\psi} = \lambda(1-E)+\psi$ we have:

$$\begin{aligned}
\mu\left(1-E\right) =& \frac{\mathrm{d}E}{\mathrm{d}t} + (\tilde{\lambda}+\tilde{\psi})E \\
\frac{\tilde{\psi}}{\tilde{\mu}+\tilde{\psi}}\mu =& \frac{\tilde{\psi}^2}{\left(\tilde{\mu}+\tilde{\psi}\right)^2}\frac{\mathrm{d}\frac{\tilde{\mu}}{\tilde{\psi}}}{\mathrm{d}t} + \left(\tilde{\lambda}+\tilde{\psi}\right)\frac{\tilde{\mu}}{\tilde{\mu}+\tilde{\psi}} \\
\mu =& \frac{\tilde{\psi}}{\tilde{\mu}+\tilde{\psi}}\frac{\mathrm{d}\frac{\tilde{\mu}}{\tilde{\psi}}}{\mathrm{d}t} + \left(\tilde{\lambda}+\tilde{\psi}\right)\frac{\tilde{\mu}}{\tilde{\psi}}
\end{aligned}$$

S5 Adding removal at sampling probability

Here we consider the time-dependant birth-death sampling model with an extra parameter, removal at sampling probability r. In such a process a lineage sampled at time t is removed from the process with probability $r\left(t\right)$. In the main text we showed that if r is an arbitrary time-dependant function then the time-dependant birth-death sampling model with removal at sampling is unidentifiable. Here we show that when r is a fixed function and $r(t) > 1$ for almost all $t \in [0, t_0]$, the parameter transformation (equation 10 from the main

text) again defines a one-to-one correspondence between the distributions of reconstructed trees generated by time-dependant birth-death sampling models with removal at sampling and distributions of complete trees generated by the FBD model with pulled rates.

S5.1 Pulled rates

The pulled rates for the reconstructed tree are as follows. Note that the pulled birth rate does not differ from the $r\,(t) = 0$ case.

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

The probability $E(t)$ of a lineage alive at time t to not leave sampled descendants is independent of what happens to lineages after being sampled. Therefore E and differential equation (2) remain unchanged.

$N\,(t)$, the total number of lineages both observed and unobserved, is affected by $r\,(t)$, while $V\,(t)$, the total number of observed lineages, is not affected.

$$N\,(t) = \frac{M_0}{\rho_0} e^{\int_0^t \mu+\psi r-\lambda\, dt}$$

$$V\,(t) = M_0 e^{\int_0^t \tilde{\mu}-\tilde{\lambda}\, dt}$$

Therefore, we obtain a new expression for $\frac{\mathrm{d}E}{\mathrm{d}t}$:

$$\begin{aligned}E\,(t) =& \frac{N\,(t)-V\,(t)}{N\,(t)}\\ =& 1-\rho_0 e^{\int_0^t \tilde{\mu}+\lambda E-\mu-\psi r\, dt}\\ \frac{\mathrm{d}E}{\mathrm{d}t} =& (E-1)\left(\tilde{\mu}+\lambda E-\mu-\psi r\right)\end{aligned}$$

We can use our two expressions for $\frac{\mathrm{d}E}{\mathrm{d}t}$ to determine the pulled death rate.

$$\tilde{\mu}+\lambda E-\mu-\psi r = \lambda E-\mu+\frac{\psi E}{1-E}$$

$$\tilde{\mu} = \frac{\psi E}{1-E}+\psi r$$

S5.2 Probability density equality

MacPherson et al. [1] gives the probability density function for reconstructed trees generated by the birth-death sampling process with removal at sampling, conditioned on S, the event of sampling at least one lineage, as:

$$\begin{aligned} f^{\text{rec.}}\left(\mathcal{T}|\lambda,\mu,\psi,r,t_0,S\right) =& \frac{\rho_0^{M_0}}{1-E\left(t_0\right)}\Phi\left(t_0\right)\prod_{k=1}^{m}\psi\left(z_k\right)\left(1-r\left(z_k\right)\right) \\ &\times \prod_{i=1}^{M_0+n-1}\lambda\left(x_i\right)\Phi\left(x_i\right) \\ &\times \prod_{j=1}^{n}\frac{\psi\left(y_j\right)}{\Phi\left(y_j\right)}\left(\left(1-r\left(y_j\right)\right)E\left(y_j\right)+r\left(y_j\right)\right) \end{aligned}$$

To show that this distribution is the same as the distribution of complete trees under the FBD model with pulled rates we again substitute pulled rates in the expression for the probability density $f^{\text{compl.}}\left(\mathcal{T}|\tilde{\lambda},\tilde{\mu},\tilde{\psi},t_0\right)$ (equation 1) with expressions for the pulled rates defined by equations (11). The contribution of speciation events in these densities remains unchanged compared to the case when $r(t) = 0$. The equality of sampling event terms is obvious and it only remains to consider death events and affected edges to show that the complete and reconstructed tree probability densities remain equal.

We confirm the equality connecting the exponential terms previously shown for $r\left(t\right) = 0$:

$$\begin{aligned} e^{-\int_0^t \tilde{\lambda}+\tilde{\mu}+\tilde{\psi}\,dt} =& e^{-\int_0^t \lambda-\lambda E+\frac{\psi E}{1-E}+\psi r+\psi(1-r)\,dt} \\ =& \left(e^{-\int_0^t -\mu+\lambda E+\frac{\psi E}{1-E}\,dt}\right)e^{-\int_0^t \lambda+\mu+\psi-2\lambda E\,dt} \\ =& \frac{\rho_0}{1-E}e^{-\int_0^t \lambda+\mu+\psi-2\lambda E\,dt} \end{aligned}$$

$$\begin{aligned} \tilde{\mu}e^{\int_0^t \tilde{\lambda}+\tilde{\mu}+\tilde{\psi}\,dt} =& \left(\frac{\psi E}{1-E}+\psi r\right)\left(\frac{1-E}{\rho_0}\right)\frac{1}{\Phi\left(t\right)} \\ =& \frac{\psi}{\rho_0}\left(E+r-rE\right)\frac{1}{\Phi\left(t\right)} \\ =& \frac{\psi}{\rho_0\Phi\left(t\right)}\left(\left(1-r\right)E+r\right) \end{aligned}$$

Thus, the distributions are equal.

S5.3 One-to-one correspondence between pulled and normal rates

Lemma S5.1. Let $F_1 = (\lambda_1, \mu_1, \psi_1, \rho_{0,1}, r)$ and $F_2 = (\lambda_2, \mu_2, \psi_2, \rho_{0,2}, r)$ where r is a fixed function such that $r(t) \neq 0$ for almost all $t \in [0, t_0]$, and $F_1 \neq F_2$.

Let $\tilde{F}_1 = \left(\tilde{\lambda}_1, \tilde{\mu}_1, \tilde{\psi}_1\right)$ and $\tilde{F}_2 = \left(\tilde{\lambda}_2, \tilde{\mu}_2, \tilde{\psi}_2\right)$ be the pulled rates obtained from F_1 and F_2 using Equation (8). Then $\tilde{F}_1 \neq \tilde{F}_2$.

The proof of Lemma S5.1 closely follows the proof of Lemma S4.1. This lemma implies the identifiability of the time-dependent birth-death-sampling model with fixed removal probability.

S5.4 Original rates from pulled rates

As previously, we derive expression for $(E, \lambda, \mu, \psi, \rho_0)$ with respect to the pulled rates and r:

$$\begin{aligned}
\lambda &= \frac{\tilde{\lambda}\left(1-r\right)\left(\tilde{\psi}+\tilde{\mu}\right)}{\tilde{\psi}} \\
\mu &= \tilde{\mu}\left(\frac{\tilde{\psi}+\tilde{\lambda}-\tilde{\lambda}r}{\tilde{\psi}}\right) - \tilde{\psi}r\left(\tilde{\lambda}+\frac{1}{1-r}\right) + \frac{\mathrm{d}\frac{\tilde{\mu}}{\tilde{\psi}}}{\mathrm{d}t}\frac{\tilde{\psi}}{\tilde{\mu}+\tilde{\psi}} \\
\psi &= \frac{\tilde{\psi}}{1-r} \\
\rho_0 &= \frac{\tilde{\psi}\left(0\right)}{\left(1-r\right)\left(\tilde{\psi}\left(0\right)+\tilde{\mu}\left(0\right)\right)}
\end{aligned} \tag{12}$$

The sampling rate now depends upon the removal rate:

$$\psi = \frac{\tilde{\psi}}{1-r}$$

E, and thus also ρ_0, is fully dependent on the removal rate, pulled death and pulled sampling rate.

$$\left(\tilde{\mu}-\psi r\right)\left(1-E\right) = \psi E$$

$$E\left(\psi+\tilde{\mu}-\psi r\right) = \tilde{\mu}-\psi r$$

$$E = 1 - \frac{\tilde{\psi}}{\left(1-r\right)\left(\tilde{\mu}+\tilde{\psi}\right)}$$

$$\rho_0 = \frac{\tilde{\psi}\left(0\right)}{\left(1-r\right)\left(\tilde{\psi}\left(0\right)+\tilde{\mu}\left(0\right)\right)}$$

We then obtain the birth rate:

$$\lambda = \frac{\tilde{\lambda}\left(1-r\right)\left(\tilde{\psi}+\tilde{\mu}\right)}{\tilde{\psi}}$$

We write the derivative of E with respect to r and the pulled rates.

$$\begin{aligned}\frac{\mathrm{d}E}{\mathrm{d}t} =& \frac{\mathrm{d}}{\mathrm{d}t}\left(1 - \frac{\tilde{\psi}}{(1-r)\left(\tilde{\mu}+\tilde{\psi}\right)}\right)\\ =& \frac{\mathrm{d}\frac{\tilde{\mu}}{\tilde{\psi}}}{\mathrm{d}t}\frac{\tilde{\psi}^2}{\left(\tilde{\mu}+\tilde{\psi}\right)^2(1-r)}\end{aligned}$$

This allows us to obtain μ.

$$\begin{aligned}\mu =& \tilde{\mu} + \lambda E - \psi r + \frac{\frac{\mathrm{d}E}{\mathrm{d}t}}{1-E}\\ \mu =& \tilde{\mu} + \tilde{\lambda}\left(\frac{\tilde{\mu} - \tilde{\psi}r - \tilde{\mu}r}{\tilde{\psi}}\right) - \frac{\tilde{\psi}r}{1-r} + \frac{\mathrm{d}\frac{\tilde{\mu}}{\tilde{\psi}}}{\mathrm{d}t}\frac{\tilde{\psi}}{\tilde{\mu}+\tilde{\psi}}\\ \mu =& \tilde{\mu}\left(\frac{\tilde{\psi} + \tilde{\lambda} - \tilde{\lambda}r}{\tilde{\psi}}\right) - \tilde{\psi}r\left(\frac{\tilde{\lambda}}{\tilde{\psi}} + \frac{1}{1-r}\right) + \frac{\mathrm{d}\frac{\tilde{\mu}}{\tilde{\psi}}}{\mathrm{d}t}\frac{\tilde{\psi}}{\tilde{\mu}+\tilde{\psi}}\end{aligned}$$

We can confirm that the previous solutions for non-removal sampling are special cases of these general solutions.

References

1. MacPherson, A., Louca, S., McLaughlin, A., Joy, J. B. & Pennell, M. W. Unifying Phylogenetic Birth-Death Models in Epidemiology and Macroevolution. en. *Syst. Biol.* **71,** 172–189 (Dec. 2021).
2. Steel, M. *Phylogeny: Discrete and Random Processes in Evolution* en (SIAM-Society for Industrial and Applied Mathematics, Sept. 2016).
3. Kubo, T. & Iwasa, Y. Inferring the rates of branching and extinction from molecular phylogenies. en. *Evolution* **49,** 694–704 (Aug. 1995).
4. Louca, S. & Pennell, M. W. Extant timetrees are consistent with a myriad of diversification histories. en. *Nature* **580,** 502–505 (Apr. 2020).