

# On incomplete sampling under birth-death models and connections to the sampling-based coalescent

Tanja Stadler

*Institut für Integrative Biologie, ETH Zürich, Universitätsstr. 16, 8092 Zürich, Switzerland, Phone +41 44 632 45 48, Fax +41 44 632 12 71, tanja.stadler@env.ethz.ch*

---

## Abstract

The constant rate birth-death process is used as a stochastic model for many biological systems, for example phylogenies or disease transmission. As the biological data is usually not fully available, it is crucial to understand the effect of incomplete sampling. In this paper, we analyze the constant rate birth-death process with incomplete sampling. We derive the density of the bifurcation events for trees on  $n$  leaves which evolved under this birth-death-sampling process. This density is used for calculating prior distributions in Bayesian inference programs and for efficiently simulating trees. We show that the birth-death-sampling process can be interpreted as a birth-death process with reduced rates and complete sampling. This shows that joint inference of birth rate, death rate and sampling probability is not possible. The birth-death-sampling process is compared to the sampling-based population genetics model, the coalescent. It is shown that despite many similarities between these two models, the distribution of bifurcation times remains different even in the case of very large population sizes. We illustrate these findings on an Hepatitis C virus dataset from Egypt. We show that the transmission times estimates are significantly different – the widely used Gamma statistic even changes its sign from negative to positive when switching from the coalescent to the birth-death process.

*Key words:* Birth-death trees, random sampling, bifurcation times.

---

## 1. Introduction

The *constant rate birth-death process* [8, 14] is widely used in biology as a neutral model for reproduction. It is a branching process which starts with an initial individual (which might be a species, an individual within a population, or an infected host). At any point in time during the process, each individual has a constant rate  $\mu$  of dying, and a constant rate  $\lambda$  of branching (i.e. giving birth to an additional individual).

We will focus on three main applications in biology: The constant rate birth-death process has successfully been used in modeling speciation and extinction – it is used as a prior distribution when

inferring phylogenies using Bayesian methods [6] as well as a neutral model for testing evolutionary hypotheses, see e.g. [10, 13]. The birth-death process is also widely used as a population model – for example, in [27], the dynamics of a population within a species is analyzed assuming a birth-death process. In [25], the birth-death process is used as a model for the spread of an infectious disease (Tuberculosis) and key values of the disease (reproductive value, net transmission rate) are estimated.

In the above mentioned biological applications,  $n$  individuals are sampled, and the task is to recover the process which led to these  $n$  individuals. Thus it is crucial to understand the birth-death process conditioned on observing  $n$  individuals. This has been investigated in detail in [1, 11, 12, 18]. In these articles, it is assumed that all individuals which evolved under the birth-death process and survived to the present are sampled. When analyzing data however, usually only a fraction of the individuals which evolved under the process are sampled. In this paper, we account for incomplete sampling by including *random sampling* in the birth-death process model. We distinguish between two scenarios: Either (i) we sample each individual in the complete birth-death tree with probability  $\rho$  – conditioning that the tree after sampling has  $n$  leaves (*birth-death-sampling $_{\rho}$  process*), or (ii) given a tree on  $m$  individuals, we sample  $n$  of these individuals uniformly at random (*birth-death-sampling $_m$  process*).

In [24], it is shown that the distribution on discrete realizations of the birth-death-sampling processes – i.e. realizations where the time between bifurcation events is ignored – is invariant towards random sampling. In [26], it is shown that the distribution on discrete realizations and bifurcation times are independent. The distribution on discrete realizations is well known (uniform distribution, see e.g. [2]). The bifurcation times distribution will be investigated in this paper.

In the first part of the paper, we derive the joint density function for the  $n - 1$  bifurcation times in a birth-death-sampling $_{\rho}$  tree on  $n$  individuals. We will establish that the birth-death-sampling $_{\rho}$  process can be interpreted as a birth-death-sampling $_{\rho}$  process with higher (resp. smaller) sampling probability and smaller (resp. higher) birth and death rates. This has an important consequence for parameter estimation: One cannot estimate the three parameters simultaneously. One of the three parameters needs to be known. We will discuss the influence of the sampling parameter on the birth and death rates. In particular, we show that for estimated birth and death rates on a dataset under complete sampling, one can rescale the rates assuming a more realistic sampling probability.

The density function for the bifurcation times is implemented in the latest version of BEAST [6] as a prior for the birth-death-sampling $_{\rho}$  model. Further, it is used in the Python package CASS [22] for fast simulations of birth-death-sampling $_{\rho}$  trees on  $n$  individuals.

We further derive the density function for the time of the most recent common ancestor of the  $n$  sampled individuals under a birth-death-sampling $_m$  process. Note that the density function for the bifurcation events between the most recent common ancestor and the present can be calculated

recursively [24].

The second part of the paper compares the birth-death-sampling process to the coalescent [15, 16, 17] which is commonly used to model processes on the population level. Realizations under the birth-death-sampling process and under the coalescent (with arbitrary population size changes) have the same distribution when considering discrete realizations, i.e. ignoring timing between bifurcation events [24]. We will investigate the difference in bifurcation times in this paper. Under the coalescent, the population size changes are deterministic. Under the birth-death process, the population size fluctuates according to the stochastic process. The expected number of individuals under a birth-death process with  $\lambda > \mu$  grows exponentially. Therefore the birth-death process is compared to the coalescent with exponentially growing population size. Under a birth-death process with  $\lambda = \mu$ , each individual is expected to give birth to one new individual during its lifetime. Therefore this birth-death process is compared to the coalescent with constant population size.

We showed in [12] that the expected bifurcation times under the coalescent (constant population size) and a birth-death process with  $\lambda = \mu = n/2$  are equal. We further showed in [24] that the expected bifurcation times under the coalescent (constant population size) and a birth-death-sampling <sub>$m$</sub>  process with  $\lambda = \mu = m/2$  are equal. However the higher moments differ. As the coalescent assumes an infinite population size, we compare the coalescent to the birth-death-sampling process with  $m \rightarrow \infty$ . We show that in the limit the bifurcation times distribution under the coalescent and the birth-death-sampling process remain different.

Further, we establish that the distribution on bifurcation times under the coalescent and a birth-death-sampling <sub>$\rho$</sub>  process remain different in the limit  $\rho \rightarrow 0$ .

This shows that even though the coalescent and the birth-death-sampling process equal on the level of discrete realizations and – for certain parameter combinations – on the level of expected bifurcation times, they are fundamentally different, even in the limit of very sparse sampling. We illustrate these findings by reconstructing the transmission tree of an Hepatitis C virus (HCV) dataset from Egypt [20] using Bayesian methods. We use both a birth-death process prior and a coalescent prior. The posterior reconstructed tree shape (i.e. the reconstructed tree ignoring the edge lengths) is approximately the same under both priors. However, the transmission times are different: the Gamma statistic [19] – a popular statistic measuring relative timing of bifurcation events – changes its sign from negative to positive when switching from the coalescent to the birth-death process. This means that choosing one model over the other makes a difference towards bifurcation times, and care has to be taken in model selection.

The paper is organized as follows. In Section 2 we will introduce the birth-death model formally and state known results. Section 3 derives densities under the birth-death-sampling <sub>$\rho$</sub>  process, Section

4 derives densities under the birth-death-sampling<sub>m</sub> process. We compare the density functions to the coalescent in Section 5. In Section 6, we discuss the implications of the theoretical results for data analysis. For an HCV dataset, we reconstruct the transmission trees under the birth-death process and the coalescent prior and then investigate the differences in the reconstructed trees.

## 2. The birth-death process model

A constant rate birth-death process is a stochastic process which starts with an initial individual. An individual gives birth to a new individual after exponential (rate  $\lambda$ ) waiting times and dies after an exponential (rate  $\mu$ ) waiting time. We set  $0 \leq \mu \leq \lambda$ , otherwise the process is supercritical. In the following, time 0 is today and time  $\tau > 0$  the origin of the tree, so time is increasing going into the past. Special cases of the birth-death process are the Yule model [30] where  $\mu = 0$  and the critical branching process [1, 18] where  $\mu = \lambda$ .

Such birth-death processes are commonly used to model speciation and extinction in phylogenies or the spread of diseases. In these biological applications, the data consists of  $n$  extant species / infected individuals. We therefore condition the process to have  $n$  individuals today. Throughout this paper we refer to this conditioned constant rate birth-death process simply as a *birth-death process*. The age of the tree, i.e. the time since origin of the birth-death process is  $\tau$ ; since  $\tau$  is typically not known, we assume a uniform prior on  $(0, \infty)$  for the time of origin as it has been done in [1, 18]. A tree which evolved under a birth-death process includes dead individuals, see Figure 1, left. From this tree, the extinct lineages are deleted, see Figure 1, middle. This new tree is called the reconstructed tree. Note that when reconstructing a tree from (molecular) data, we see the reconstructed tree. Extinct lineages are only apparent when the fossils (or from a disease recovered individuals) are included. We therefore investigate the distribution on reconstructed trees. The reconstructed tree consists of two parts: (i) the discrete *ranked tree* and (ii) the  $n - 1$  *bifurcation times*  $x_1 > x_2, \dots, x_{n-1}$ , see Figure 1. A ranked tree is induced by a reconstructed tree as follows. The first bifurcation event in the tree (i.e. the root) is labeled with 1, the second bifurcation event is labeled with 2, etc. Then the edge lengths are dropped. Therefore a ranked tree is a purely discrete object – a tree topology together with an ordering on the bifurcation events, see Figure 1, right. The distribution on ranked trees and bifurcation times are independent [26]. The distribution on ranked trees is the uniform distribution [24]. The distribution on bifurcation times under the birth-death process is discussed in detail in [11]: The density function for  $x = (x_1, \dots, x_{n-1})$ , conditioned on the time of origin,  $\tau$ , is,

$$f^{\lambda, \mu}(x|\tau, n) = (n-1)! \prod_{i=1}^{n-1} \frac{(\lambda - \mu)^2 e^{-(\lambda - \mu)x_i}}{(\lambda - \mu e^{-(\lambda - \mu)x_i})^2} \frac{\lambda - \mu e^{-(\lambda - \mu)\tau}}{1 - e^{-(\lambda - \mu)\tau}}.$$

Further, with the uniform prior on the time of origin, the density for  $x$  given  $n$  individuals at the present,  $f(x|n)$ , is obtained,

$$f^{\lambda, \mu}(x|n) = n!(\lambda - \mu)\lambda^{n-1} \frac{e^{-(\lambda - \mu)x_1}}{\lambda - \mu e^{-(\lambda - \mu)x_1}} \prod_{i=1}^{n-1} \frac{(\lambda - \mu)^2 e^{-(\lambda - \mu)x_i}}{(\lambda - \mu e^{-(\lambda - \mu)x_i})^2}. \quad (1)$$

This density function is used in the Bayesian inference program BEAST [6].

## 3. The birth-death-sampling <sub>$\rho$</sub> process

We assumed complete sampling of leaves in the previous section. However, in many applications, we have incomplete sampling. In the following we assume that we sample each leaf with probability  $\rho$ ,

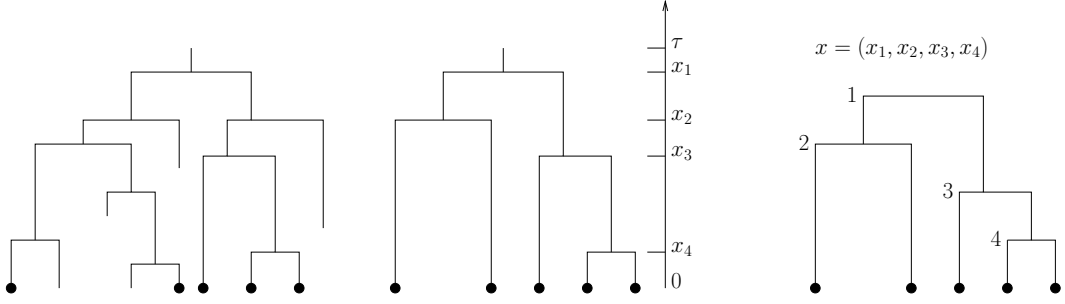


Figure 1: A tree including the extinct lineages and all leaves (left) and its reconstructed tree with five leaves sampled (middle). A reconstructed tree consists of continuous bifurcation times  $x = (x_1, \dots, x_{n-1})$  and the discrete ranked tree (right). The bifurcation times are independent of the ranked tree under the birth-death-sampling process. The distribution on ranked trees under the birth-death-sampling process is the uniform distribution. We will determine the distribution on bifurcation times  $x$ .

we call this process the birth-death-sampling $_{\rho}$  process. The reconstructed tree obtained after random sampling is conditioned to have  $n$  leaves.

In [26], the author established that for a reconstructed tree under a birth-death process, the distribution on the ranked trees is independent of the distribution on bifurcation times. The distribution on ranked trees – which is the uniform distribution – is invariant towards random sampling [24]. It therefore remains to investigate the distribution on bifurcation times for a reconstructed tree after random sampling, see Figure 1.

### 3.1. Deriving the density function for the bifurcation times

In this section we derive the density for the bifurcation times  $x$  under the birth-death-sampling $_{\rho}$  process,  $f_{\rho}^{\lambda, \mu}(x|n)$ . Conditioning on the most recent common ancestor of the sampled individuals being at time  $x_1$ , we obtain the density function of the bifurcation times  $x = (x_1, \dots, x_{n-1})$  [29],

$$f_{\rho}^{\lambda, \mu}(x|n, x_1) = (n-2)! \left( \prod_{i=2}^{n-1} \frac{(\lambda - \mu)^2 e^{-(\lambda - \mu)x_i}}{(\rho\lambda + (\lambda(1 - \rho) - \mu)e^{-(\lambda - \mu)x_i})^2} \right) \left( \frac{\rho\lambda + (\lambda(1 - \rho) - \mu)e^{-(\lambda - \mu)x_1}}{1 - e^{-(\lambda - \mu)x_1}} \right)^{n-2}.$$

Using the same approach as in [11], we obtain the density function of the bifurcation times  $x = (x_1, \dots, x_{n-1})$  conditioning on the time of origin,  $\tau$ , of the process:

$$f_{\rho}^{\lambda, \mu}(x|n, \tau) = (n-1)! \left( \prod_{i=1}^{n-1} \frac{(\lambda - \mu)^2 e^{-(\lambda - \mu)x_i}}{(\rho\lambda + (\lambda(1 - \rho) - \mu)e^{-(\lambda - \mu)x_i})^2} \right) \left( \frac{\rho\lambda + (\lambda(1 - \rho) - \mu)e^{-(\lambda - \mu)\tau}}{1 - e^{-(\lambda - \mu)\tau}} \right)^{n-1}. \quad (2)$$

Note that for  $\rho = 1$ , we have  $f_1^{\lambda, \mu}(x|n, \tau) = f^{\lambda, \mu}(x|n, \tau)$ .

In general, we do not know the time of origin,  $\tau$ . Commonly, a uniform prior on  $(0, \infty)$  is assumed. Let  $q_{\rho}^{\lambda, \mu}(\tau|n)$  be the density function for the time of origin of a birth-death-sampling $_{\rho}$  tree on  $n$  individuals (we derive this density in Lemma 3.1). The density of the bifurcation times being  $x = (x_1, \dots, x_{n-1})$  and the time of origin being  $\tau$  is,

$$f_{\rho}^{\lambda, \mu}(x, \tau|n) = f_{\rho}^{\lambda, \mu}(x|n, \tau) q_{\rho}^{\lambda, \mu}(\tau|n). \quad (3)$$

To obtain the density of  $x$  given  $n$ ,  $f_{\rho}^{\lambda, \mu}(x|n)$ , we have to integrate Equation (3) over  $\tau$  (from  $x_1$  to  $\infty$ ). In order to derive  $f_{\rho}^{\lambda, \mu}(x|n)$ , we first determine  $q_{\rho}^{\lambda, \mu}(\tau|n)$  (Lemma 3.1). In the second step we

integrate Equation (3) (Theorem 3.3). The following Lemma is proven in the Appendix.

**Lemma 3.1.** *Consider a birth-death-sampling $_{\rho}$  process with birth rate  $\lambda$  and death rate  $\mu$ . A uniform prior on  $(0, \infty)$  is assumed for the time of origin. We sample each individual in the birth-death tree with probability  $\rho$ . The density for a tree on  $n$  individuals to have time of origin  $\tau$  is,*

$$q_{\rho}^{\lambda, \mu}(\tau|n) = n(\lambda - \mu)^2 (\lambda \rho)^n \frac{(1 - e^{-(\lambda - \mu)\tau})^{n-1} e^{-(\lambda - \mu)\tau}}{(\rho\lambda + (\lambda(1 - \rho) - \mu)e^{-(\lambda - \mu)\tau})^{n+1}}. \quad (4)$$

**Remark 3.2.** The software package CASS [22] uses Equation (4) and (2) to sample trees under the birth-death-sampling $_{\rho}$  process. This sampling procedure is very efficient: we sample a ranked tree on  $n$  leaves uniformly at random. We then sample the time of origin from the distribution given in Equation (4). Last, the bifurcation times are sampled from Equation (2). In particular, the simulation approach does not require a sample of trees with more than  $n$  leaves from which to subsample. We incorporate the sampling directly into the bifurcation times which makes the simulation algorithm very fast. For a detailed explanation of the simulation procedure see [23], where we describe the method under complete sampling, i.e. for the special case  $\rho = 1$ .

We now establish the main theorem of this section. The proof is found in the Appendix.

**Theorem 3.3.** *Consider a birth-death-sampling $_{\rho}$  process with birth rate  $\lambda$  and death rate  $\mu$ . A uniform prior on  $(0, \infty)$  is assumed for the time of origin. We sample each individual in the birth-death tree with probability  $\rho$ . Conditioning the tree on having  $n$  individuals after random sampling, the density of bifurcation times  $x = (x_1, \dots, x_{n-1})$  is,*

$$f_{\rho}^{\lambda, \mu}(x|n) = n!(\lambda - \mu)(\lambda \rho)^{n-1} \frac{e^{-(\lambda - \mu)x_1}}{\rho\lambda + (\lambda(1 - \rho) - \mu)e^{-(\lambda - \mu)x_1}} \prod_{i=1}^{n-1} \frac{(\lambda - \mu)^2 e^{-(\lambda - \mu)x_i}}{(\rho\lambda + (\lambda(1 - \rho) - \mu)e^{-(\lambda - \mu)x_i})^2}. \quad (5)$$

For  $\mu \rightarrow \lambda$ , we get in the limit,

$$f_{\rho}^{\lambda, \lambda}(x|n) = n!(\lambda \rho)^{n-1} \frac{1}{1 + \rho\lambda x_1} \prod_{i=1}^{n-1} \frac{1}{(1 + \rho\lambda x_i)^2}. \quad (6)$$

**Remark 3.4.** Note that a birth-death-sampling $_{\rho}$  process with birth rate  $\lambda$ , death rate  $\mu$  and sampling probability  $\rho$  equals a birth-death-sampling $_{\rho'}$  process with sampling probability  $\rho'$ , birth rate  $\lambda' = \rho\lambda/\rho'$  and death rate  $\mu' = \mu - \lambda(1 - \rho/\rho')$ . This follows directly from (5): First note that  $\lambda' - \mu' = \lambda - \mu$ . Further,  $\rho'\lambda' = \rho\lambda$  and  $\lambda'(1 - \rho') - \mu' = \lambda(1 - \rho) - \mu$ . Therefore, Equation (5) yields,

$$f_{\rho'}^{\lambda', \mu'}(x|n) = f_{\rho}^{\lambda, \mu}(x|n).$$

This has a major impact on estimating the birth and death rates and the sampling probability. The three parameters cannot be estimated simultaneously. We will discuss this issue further in Section 6.

### 3.1.1. Sampling probability $\rho \rightarrow 0$

Under the birth-death-sampling $_{\rho}$  process for  $\rho \rightarrow 0$ , with  $\lambda - \mu = d$ ,  $\lambda\rho = c$ , we obtain from Equation (5),

$$f_0^{\lambda, \mu}(x|n) = n!dc^{n-1} \frac{e^{-dx_1}}{c + (d - c)e^{-dx_1}} \prod_{i=1}^{n-1} \frac{d^2 e^{-dx_i}}{(c + (d - c)e^{-dx_i})^2}. \quad (7)$$

Note that, for constants  $c > d > 0$ ,

$$f_0^{\lambda, \mu}(x|n) = f_1^{c, c-d}(x|n),$$

i.e. the bifurcation times are equal to a birth-death process with birth rate  $c$  and death rate  $c - d$ . Under the birth-death-sampling $_{\rho}$  process for  $\rho \rightarrow 0$ ,  $\lambda = \mu$  and  $\rho\lambda = c$  where  $c$  is a constant, we obtain

with Equation (6),

$$f_0^{\lambda,\lambda}(x|n) = n!c^{n-1} \frac{1}{1+cx_1} \prod_{i=1}^{n-1} \frac{1}{(1+cx_i)^2}. \quad (8)$$

Note again that,

$$f_0^{\lambda,\lambda}(x|n) = f_1^{c,c}(x|n).$$

## 4. The birth-death-sampling<sub>m</sub> process

### 4.1. Sampling two leaves

We now derive the density that two randomly chosen leaves in a tree of size  $m$  have bifurcated at time  $x$ ,  $f_m^{\lambda,\mu}(x|2)$ . In Section 4.2, we extend the calculations for the case when sampling  $n$  lineages in a tree of  $m$  lineages. The following Theorem for sampling 2 leaves is proven in the Appendix.

**Theorem 4.1.** *Consider a birth-death-sampling<sub>m</sub> process with birth rate  $\lambda$  and death rate  $\mu$ . A uniform prior on  $(0, \infty)$  is assumed for the time of origin. We sample two individuals from a birth-death tree on  $m$  individuals. The density of the time of the most recent common ancestor of the two sampled individuals is,*

$$f_m^{\lambda,\mu}(x|2) = \frac{2\lambda e^{(\lambda-\mu)x}}{m-1} \left(\frac{1}{a+1}\right)^{m+1} \left[ (a+1)^{m+1} - \binom{m+1}{2} a^2 - (m+1)a - 1 \right] \quad (9)$$

where  $a := \frac{\lambda-\mu}{\lambda} \frac{e^{-(\lambda-\mu)x}}{(1-e^{-(\lambda-\mu)x})}$ . For the limit  $\mu \rightarrow \lambda$ , we have,

$$f_m^{\lambda,\mu}(x|2) = \frac{2\lambda}{m-1} \left(\frac{\lambda x}{1+\lambda x}\right)^{m+1} \left[ \left(1 + \frac{1}{\lambda x}\right)^{m+1} - \binom{m+1}{2} \frac{1}{(\lambda x)^2} - \frac{m+1}{\lambda x} - 1 \right]. \quad (10)$$

#### 4.1.1. Sampling two leaves in the limit $m \rightarrow \infty$

We will now determine the density for the time of the most recent common ancestor of two leaves sampled from a tree on  $m$  individuals where  $m \rightarrow \infty$ . From Equation (9) with  $\lambda - \mu = d$ ,  $\lambda = cm$ , we obtain,

$$f_m^{\lambda,\mu}(x|2) = \frac{2cme^{dx}}{m-1} - \frac{2cme^{dx}}{m-1} \left( \frac{1-e^{-dx}}{1-e^{-dx} + \frac{d}{cm}e^{-dx}} \right)^{m+1} \times \left[ \binom{m+1}{2} \left( \frac{d}{cm} \frac{e^{-dx}}{1-e^{-dx}} \right)^2 + (m+1) \frac{d}{cm} \frac{e^{-dx}}{1-e^{-dx}} + 1 \right].$$

In the limit, when  $m \rightarrow \infty$ , we get, with the property  $\lim_{n \rightarrow \infty} (1+x/n)^n = e^x$ ,

$$\begin{aligned} f_\infty^{\lambda,\mu}(x|2) &= 2ce^{dx} - 2ce^{dx} e^{-\frac{d}{c} \frac{e^{-dx}}{1-e^{-dx}}} \times \\ &\quad \left[ \frac{1}{2} \left( \frac{d}{c} \frac{e^{-dx}}{1-e^{-dx}} \right)^2 + \frac{d}{c} \frac{e^{-dx}}{1-e^{-dx}} + 1 \right] \\ &= 2ce^{dx} - \frac{d^2}{c} e^{dx} \left( \frac{1}{e^{dx}-1} \right)^2 e^{-\frac{d}{c} \frac{e^{-dx}}{1-e^{-dx}}} - 2d \frac{e^{dx}}{e^{dx}-1} e^{-\frac{d}{c} \frac{e^{-dx}}{1-e^{-dx}}} - 2ce^{dx} e^{-\frac{d}{c} \frac{e^{-dx}}{1-e^{-dx}}}. \end{aligned} \quad (11)$$

For  $\lambda = \mu$ , when  $m \rightarrow \infty$ , we obtain from Equation (10) with  $\lambda = cm$ ,

$$f_\infty^{\lambda,\lambda}(x|2) = 2c - \frac{e^{-1/cx}}{cx^2} - \frac{2e^{-1/cx}}{x} - 2ce^{-1/cx}. \quad (12)$$

#### 4.2. Sampling $n$ leaves

In this section, we determine the time of the most recent common ancestor of  $n$  individuals sampled from a tree on  $m$  individuals. When randomly sampling  $n$  individuals, from a tree  $\mathcal{T}_m$  on  $m$  individuals, the probability that the  $k$ -th bifurcation event is the the most recent common ancestor of the  $n$  individuals can be calculated as follows. Let  $s_k$  be the  $k$ -th bifurcation event in  $\mathcal{T}_m$ . Let  $s_k^l$  be the event that  $s_k$  has  $l$  descendant leaves. Let  $s_k^{l,l_1}$  be the event that  $s_k$  has  $l$  descendant leaves,  $l_1$  leaves in the left subtree and  $l - l_1$  leaves in the right subtree. Let  $mrca(s_k)$  be the event that  $s_k$  is the most recent common ancestor of the  $n$  sampled leaves. Then,

$$p_k^n := \mathbb{P}[mrca(s_k)] = \sum_{l=n}^{m-k+1} \mathbb{P}[s_k^l] \sum_{l_1=1}^l \mathbb{P}[s_k^{l,l_1} | s_k^l] \mathbb{P}[mrca(s_k) | s_k^{l,l_1}]$$

In [24] we established,

$$\mathbb{P}[s_k^l] = k(k-1) \frac{(l-1)!(m-l-1)!}{(m-1)!} \binom{m-k-1}{l-2}.$$

Further, from [21],

$$\mathbb{P}[s_k^{l,l_1} | s_k^l] = \frac{1}{l-1}.$$

Counting the number of ways sampling  $n$  leaves such that  $s_k$  is the most recent common ancestor yields,

$$\mathbb{P}[mrca(s_k) | s_k^{l,l_1}] = \frac{\binom{l}{n} - \binom{l_1}{n} - \binom{l-l_1}{n}}{\binom{m}{n}}.$$

Combining these probabilities yields,

$$p_k^n = \sum_{l=n}^{m-k+1} \binom{k}{l} \frac{\binom{m-k-1}{l-2}}{\binom{m-1}{l}} \sum_{l_1=1}^l \frac{\binom{l}{n} - \binom{l_1}{n} - \binom{l-l_1}{n}}{\binom{m}{n}}.$$

Note that for  $n = 2$ , the formula simplifies to Equation (17). With this formula for  $p_k^n$ , we can calculate the density of the time of the most recent common ancestor of a sample of  $n$  individuals in a tree of  $m$  individuals,

$$f_m^{\lambda,\mu}(x|n) = \sum_{k=1}^{m-1} p_k^n g_k^{\lambda,\mu}(x|m)$$

with  $g_k^{\lambda,\mu}(x|m)$  as introduced in Equation (18).

The distribution for the  $k$ -th bifurcation event in a tree of size  $n$  obtained from a birth-death-sampling $_m$  process where the complete tree has size  $m$  can be obtained via a recursion [24],

$$g_{m,k}^{\lambda,\mu}(x|n) = \left(1 - \frac{(k+1)k}{(n+1)n}\right) g_{m,k}^{\lambda,\mu}(x|n+1) + \frac{(k+1)k}{(n+1)n} g_{m,k+1}^{\lambda,\mu}(x|n+1)$$

with  $g_{m,k}^{\lambda,\mu}(x|m) = g_k^{\lambda,\mu}(x|m)$  (Equation (18)).

## 5. Comparison with the coalescent

When birth-death models are used within populations (e.g. for modeling transmission of diseases) sampling is very incomplete. In Equations (7, 8) (resp. (11, 12)) we established the density function of the birth-death-sampling $_\rho$  (resp. birth-death-sampling $_m$ ) process for  $\rho \rightarrow 0$  (resp.  $m \rightarrow \infty$ ). We will now compare these distributions to the coalescent - one of the standard models in population genetics.

### 5.1. The coalescent

The coalescent [15, 16, 17] is a continuous time approximation of the discrete Wright-Fisher model [9, 28]. Under the Wright-Fisher model, non-overlapping generations are assumed, and each individual

in the current generation chooses an individual uniformly at random from the previous generation. The population size might vary in different ways. Let  $N_0$  be the initial population size, and  $N_t$  the population size after time  $t$ . In this paper we consider a constant population size, i.e.  $N_t = N_0$  for all  $t$ , and an exponentially increasing population size, i.e.  $N_t = N_0 e^{rt}$  for some constant  $r > 0$ . The genealogy of  $n$  individuals which are sampled from the current generation is traced back to their most recent common ancestor. This means that under the Wright-Fisher model, we do not consider a sampling probability, but we sample  $n$  individuals from a population of  $m$  individuals. The coalescent is the continuous time approximation of the Wright-Fisher model for  $m \rightarrow \infty$ . Kingman [15, 16, 17] obtained the density for the  $n - 1$  coalescent times of  $n$  individuals under the coalescent when the population size is constant,

$$f_{const}(x|n) = \frac{n!(n-1)!}{2^{n-1}} \prod_{i=2}^n e^{-\binom{i}{2}(x_{i-1}-x_i)}$$

with  $x_n := 0$ .

In the case of an exponential increasing population size with growth parameter  $r$ , the density for the  $n - 1$  coalescent times of  $n$  individuals is [5],

$$f_{exp}^r(x|n) = \frac{n!(n-1)!}{2^{n-1}} \prod_{i=2}^n e^{rx_{i-1}} e^{-\binom{i}{2} \frac{e^{rx_i} - 1 - e^{rx_i}}{r}}$$

again with  $x_n := 0$ .

Note that under the birth-death-sampling process with  $\lambda > \mu$ , we expect an exponential growth of the population. We therefore compare this process to the coalescent with exponentially increasing population size.

Under the birth-death-sampling process with  $\lambda = \mu$ , we expect each individual to have one offspring, so the population stays constant in expectation, before going extinct. We compare this process to the coalescent with constant population size.

Note that under the birth-death process, the population size is fluctuating, whereas under the coalescent, the population size is changing deterministically.

## 5.2. Constant population size

In [24], we obtained that the expectation of the  $k$ -th bifurcation time in a tree with  $n$  individuals and  $\lambda = \mu = \frac{m}{2}$  under the birth-death-sampling <sub>$m$</sub>  process equals the expected coalescent time under the coalescent with constant population size. However, the higher moments are different for finite  $m$ . We will now investigate how the density  $f_{\rho}^{\lambda, \mu}(x|2)$  for  $\rho \rightarrow 0$  and  $f_m^{\lambda, \mu}(x|2)$  for  $m \rightarrow \infty$  compare to the most recent common ancestor of two individuals under the coalescent. The time of the most recent common ancestor under the coalescent is

$$f_{const}(x|2) = e^{-x}$$

with expectation 1. We established in Equation (8),

$$f_0^{\lambda, \lambda}(x|2) = \frac{2c}{(1+cx)^3}.$$

Note that this density function has expectation 1 for  $c = 1$ . Further, we established in Equation (12) the density function,

$$f_{\infty}^{\lambda, \lambda}(x|2) = 2c - \frac{e^{-1/cx}}{cx^2} - \frac{2e^{-1/cx}}{x} - 2ce^{-1/cx}.$$

Its expectation is 1 for  $c = 1/2$ . Clearly, we can see from the equations for these three densities that,

$$f_0^{\lambda, \lambda}(\cdot|2) \neq f_{const}(\cdot|2) \neq f_{\infty}^{\lambda, \lambda}(\cdot|2),$$

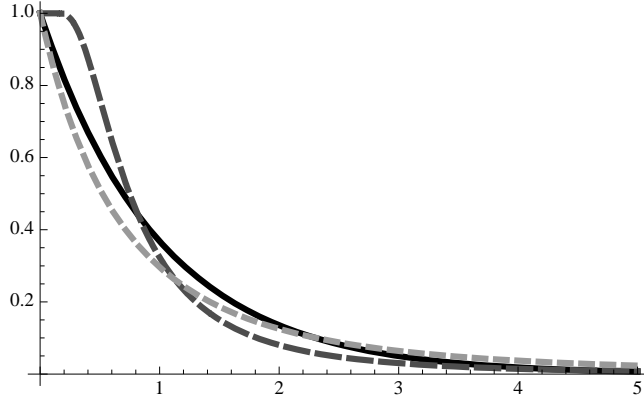


Figure 2: Density of the coalescent time of two individuals under the coalescent,  $f_{const}(x|2)$ , (solid black), under the birth-death-sampling $_m$  process,  $f_{\infty}^{\lambda,\lambda}(x|2)$  with  $m \rightarrow \infty, \lambda = m/2$  (dashed dark) and under the birth-death-sampling $_{\rho}$  process,  $f_0^{\lambda,\lambda}(x|2)$  with  $\rho = 0, \lambda = 1/(2\rho)$  (dashed light).

i.e. the coalescent and the birth-death process remain different, even for large populations and sparse sampling. The distributions with parameters  $c = 1/2$  – which corresponds to  $f_0^{\lambda,\lambda}(0|2) = f_{const}(0|2) = f_{\infty}^{\lambda,\lambda}(0|2)$  – are shown in Figure 2.

### 5.3. Exponentially growing population size

For a sample of size two, the density for the coalescent of these two individuals, assuming an exponentially increasing population size, is,

$$f_{exp}^r(x|2) = e^{rx} e^{-\frac{e^{rx}-1}{r}}.$$

For the birth-death-sampling $_{\rho}$  process, the density of the coalescence of two individuals ( $\rho \rightarrow 0, \lambda - \mu = d, \lambda\rho = c$ ), is derived in Equation (7),

$$f_0^{\lambda,\mu}(x|2) = \frac{2cd^3 e^{-2dx}}{(c + (d-c)e^{-dx})^3}. \quad (13)$$

Note that, for constants  $c > d > 0$ ,

$$f_0^{\lambda,\mu}(x|2) = f_1^{c,c-d}(x|2).$$

For the birth-death-sampling $_{\rho}$  process, the density of the coalescence of two individuals ( $\rho \rightarrow 0, \lambda - \mu = d, \lambda = cm$ ), is derived in Equation (11),

$$f_{\infty}^{\lambda,\mu}(x|2) = 2ce^{dx} - \frac{d^2}{c} e^{dx} \left( \frac{1}{e^{dx}-1} \right)^2 e^{-\frac{d}{c} \frac{e^{-dx}}{1-e^{-dx}}} - 2d \frac{e^{dx}}{e^{dx}-1} e^{-\frac{d}{c} \frac{e^{-dx}}{1-e^{-dx}}} - 2ce^{dx} e^{-\frac{d}{c} \frac{e^{-dx}}{1-e^{-dx}}}.$$

This shows that even in the limit  $\rho \rightarrow 0$  and  $m \rightarrow \infty$ , we have

$$f_0^{\lambda,\mu}(\cdot|2) \neq f_{exp}^r(\cdot|2) \neq f_{\infty}^{\lambda,\mu}(\cdot|2),$$

i.e. again the coalescent and the birth-death process remain different, even for large populations and sparse sampling. The distributions with parameters  $c = 1/2, d = 1/4, r = 1$  – which correspond to  $f_0^{\lambda,\lambda}(0|2) = f_{exp}^r(0|2) = f_{\infty}^{\lambda,\lambda}(0|2)$  – are shown in Figure 3.

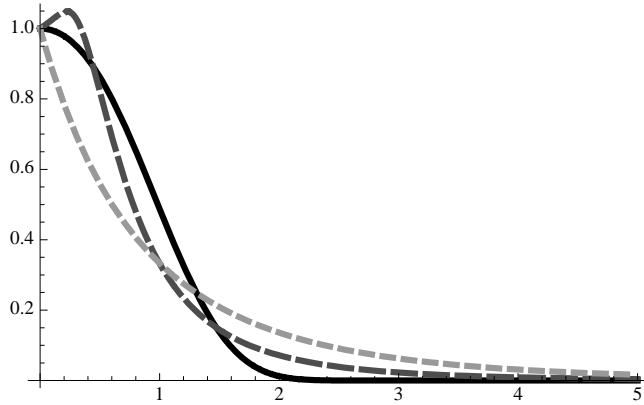


Figure 3: Density of the coalescent time of two individuals under the coalescent (exponentially growing population size),  $f_{exp}^r(x|2)$  with  $r = 1$  (solid black), under the birth-death-sampling $_m$  process,  $f_{\infty}^{\lambda,\mu}(x|2)$  with  $m \rightarrow \infty, \lambda = m/2, \lambda - \mu = 1/4$  (dashed dark) and under the birth-death-sampling $_{\rho}$  process,  $f_0^{\lambda,\mu}(x|2)$  with  $\rho = 0, \lambda = 1/(2\rho), \lambda - \mu = 1/4$  (dashed light).

## 6. Impact on data analysis

### 6.1. Bayesian phylogenetic tree inference under a birth-death-sampling prior

In Bayesian phylogenetic tree inference, Equation (5) is the default prior distribution on bifurcation times. With  $\theta$  being the parameters of the sequence evolution model and  $\tau$  being a reconstructed tree, the posterior given the data (i.e. an alignment) is,

$$\mathbb{P}[\tau, \theta, \lambda, \mu, \rho | data] \propto \mathbb{P}[data | \tau, \theta] \mathbb{P}[\theta] \mathbb{P}[\tau | \lambda, \mu, \rho] \mathbb{P}[\lambda, \mu] \mathbb{P}[\rho].$$

Bayesian tree inference methods sample from this posterior distribution. Note that on the right-hand side, only the term  $\mathbb{P}[\tau | \lambda, \mu, \rho] \mathbb{P}[\lambda, \mu] \mathbb{P}[\rho]$  depends on  $\lambda, \mu, \rho$ .

As under the birth-death-sampling $_{\rho}$  model, each ranked tree is equally likely [24], and there are  $\frac{n!(n-1)!}{2^{n-1}}$  ranked trees on  $n$  leaves [7], we have,

$$\mathbb{P}[\tau | \lambda, \mu, \rho] = f_{\rho}^{\lambda,\mu}(x|n) \frac{2^{n-1}}{n!(n-1)!}.$$

With Remark 3.4, we have  $f_{\rho}^{\lambda'/\rho, \mu' - \lambda'(1-1/\rho)}(x|n) = f_1^{\lambda', \mu'}(x|n)$ . If we assume a uniform prior for  $(\lambda, \mu)$  and  $\rho$ , we therefore have,

$$\mathbb{P}[\tau, \theta, \lambda'/\rho, \mu' - \lambda'(1-1/\rho), \rho | data] = \mathbb{P}[\tau, \theta, \lambda', \mu', 1 | data]. \quad (14)$$

This means that we cannot estimate  $\lambda, \mu$  and  $\rho$  together.

In BEAST, sampling is assumed to be complete and the default is a uniform prior on  $[0, 10^6]$  for  $\lambda - \mu$  and a uniform prior on  $[0, 1]$  for  $\mu/\lambda$ . In particular, that implies a uniform prior on the pair

$(\lambda, \mu)$  (see Lemma A.1 in the Appendix). When including incomplete sampling, we assume a uniform prior on  $[0, 1]$  for  $\rho$ .

Note that the posterior reconstructed trees do not change when including incomplete sampling (Equation (14)). One can therefore do the Bayesian tree inference assuming complete sampling. Let the birth and death rate estimates under complete sampling be  $\lambda', \mu'$ . If the sampling probability  $\rho$  can be obtained from a different consideration, we then get the actual birth and death rate estimates as

$$\lambda = \lambda'/\rho, \quad \mu = \mu' - \lambda'(1 - 1/\rho). \quad (15)$$

We suggest to transform the birth and death rates in reconstructed trees according to this equation. Note that for decreasing sampling probability  $\rho$ , the birth rate and death rate increase. The difference of birth and death rate stays constant. The ratio of death rate over birth rate tends to 0 for  $\rho \rightarrow 0$ .

### 6.2. Maximum likelihood birth and death rates

When estimating maximum likelihood birth and death rates on a given reconstructed tree  $\tau$ , we maximize the density

$$\mathbb{P}[\tau|\lambda, \mu, \rho] = f_{\rho}^{\lambda, \mu}(x|n) \frac{2^{n-1}}{n!(n-1)!}.$$

When estimating birth and death rates for a given reconstructed tree, we have with Remark 3.4:

$$\mathbb{P}[\tau|\lambda'/\rho, \mu' - \lambda'(1 - 1/\rho), \rho] = \mathbb{P}[\tau|\lambda', \mu', 1]$$

It is therefore again impossible to get estimates for the three parameters simultaneously. Again, we can estimate the birth and death rates under complete sampling and then get the actual birth and death rates under incomplete sampling from Equation (15).

### 6.3. Coalescent tree prior vs. birth-death process tree prior

In order to show that different prior distributions on trees result in different posterior distributions, we investigate transmission trees of an infectious disease, Hepatitis C. For infectious diseases, it is not clear which prior to use: The birth-death process models each infected host having an exponential (rate  $\mu$ ) distributed lifetime during which it infects other hosts with a constant rate  $\lambda$ . The coalescent assumes discrete generations, i.e. the time from infection until infecting new hosts is fixed. Reality should be somewhere in between.

We investigated an HCV dataset from Egypt [20] consisting of 71 HCV-infected individuals. The transmission history of this dataset is reconstructed using a Bayesian approach. Note that the posterior distribution of trees is invariant towards changes in the sampling probability (as discussed above – only the birth and death rates change according to Equation (15) under incomplete sampling). This means

that the posterior distribution on trees under incomplete sampling and complete sampling is the same. We therefore did the Bayesian analysis under complete sampling.

We ran BEAST both with the birth-death process tree prior and the coalescent tree prior on our HCV dataset. Under the coalescent, we chose an exponentially growing population size, since under the birth-death process, the population size is in expectation growing exponentially as well (unless  $\lambda = \mu$ ). Under both tree prior distributions, we chose the HKY model as a substitution model. Four categories were chosen in the gamma + invariant site model. We used the relaxed clock model with an uncorrelated lognormal prior in order to avoid confounding temporal information with molecular rate variation. The remaining settings were left default. Both runs were carried out for 100 million generations, and each 1 million-th state was sampled. The first 10% of the trees were removed as a burn-in. Convergence was then checked with the BEAST program Tracer. The effective sampling size was at least 240, in most cases higher. The xml input files for BEAST as well as the posterior trees are available from the author upon request.

First, we investigated the difference in tree shape (i.e. the trees without edge lengths). The tree shape distribution under the birth-death process and the coalescent are the same [2]. This suggests also the same tree shape distribution in the posterior. For the 900 trees in each posterior, we calculated the majority rule consensus tree. The two consensus trees were almost identical, there is only one edge in the birth-death consensus tree which induces a bipartition of the tip labels which is not found in the coalescent consensus tree. All the bipartitions of the tips induced by the edges in the coalescent consensus tree are also found in the birth-death consensus tree.

We then did a more detailed analysis of the tree shape. The Colless index [4] is a measure of tree shape – it measures how balanced a tree is. We calculated the Colless value for each tree and compared the distribution of Colless values under the birth-death prior and the coalescent prior. The mean was 444.10 with standard deviation 62.33 under the birth-death prior and the mean was 445.95 with standard deviation 67.28 under the coalescent prior. This shows that the tree shape distribution of the posterior is about the same.

Next, we investigated the time of transmission events. The Gamma statistic [19] is a standard statistic summarizing the relative timing of bifurcation events in a tree. A Gamma value smaller than 0 indicates that most transmission events occur earlier than expected under a pure birth model (i.e. a birth-death process with  $\mu = 0$ ). A Gamma value bigger than 0 indicates that most transmission events occur later than expected under a pure birth model, see also [19].

We calculated the Gamma statistic for all posterior trees. Under the birth-death prior, the mean was 0.35 with standard deviation 0.94. Under the coalescent, the mean was  $-0.56$  with standard deviation 1.56. This means that under the birth-death model prior, the times of transmission events

are clustered closer to the tips than expected under a pure birth process. Under the coalescent, the transmission events are clustered closer to the root than expected under the pure birth model. The standard deviation is significantly bigger under the coalescent than under the birth-death model. A difference in transmission times changes the infectious disease dynamics. Therefore care has to be taken in model selection.

Last, we looked at the expected growth rate of the number of infected individuals. For the birth-death process, we find  $\lambda - \mu = 8.24$  with standard deviation 2.64. For the coalescent, we find the growth rate  $r = 9.35$  with standard deviation 8.18. Therefore, the growth rate estimates are in the same order of magnitude. Again, the standard deviation under the coalescent is much higher than the standard deviation under the birth-death process.

## 7. Summary

We derived the distribution of bifurcation times for a tree which evolved under the birth-death-sampling $_{\rho}$  process. We showed that the birth-death-sampling $_{\rho}$  process can be interpreted as a birth-death process (complete sampling) with reduced rates. This means in particular that we cannot estimate the sampling probability together with the bifurcation and extinction rates. One of the three rates has to be known in order to infer the other two rates. The derived distribution on bifurcation times for trees on  $n$  leaves is implemented in the Bayesian inference program BEAST as a prior distribution. It is further used in the Python package CASS to simulate trees efficiently.

Secondly, we derived the distribution of the time of the most recent common ancestor of a sample of  $n$  individuals which evolved under the birth-death-sampling $_m$  process.

We compare the two sampling scenarios to the sampling-based coalescent model. It is well known that the discrete realizations (i.e. the ranked trees) have the same distribution. Further, the coalescent with constant population size and the birth-death-sampling $_m$  process with  $\lambda = \mu = m/2$  have the same expected bifurcation times. However, the higher moments are different. We show that the distribution of bifurcation times of the coalescent and the two sampling scenarios are also different in the limit of infinite tree size.

We show, using an HCV dataset, that the theoretical differences also become apparent when applying the models to data. In particular, the Gamma statistic measuring the relative timing of bifurcation (i.e. transmission) events changes its sign from negative to positive when switching from a coalescent prior to a birth-death prior.

## 8. Acknowledgements

The author thanks Alexei Drummond for very helpful discussion and Joseph Heled for integrating the results as a prior into BEAST. T.S. is funded by the ETH Zurich.

## References

- [1] D. Aldous and L. Popovic. A critical branching process model for biodiversity. *Adv. in Appl. Probab.*, 37(4):1094–1115, 2005.
- [2] D. J. Aldous. Stochastic models and descriptive statistics for phylogenetic trees, from Yule to today. *Statist. Sci.*, 16(1):23–34, 2001.
- [3] I. N. Bronstein, K. A. Semendjajew, G. Musiol, and H. Mühlig. *Taschenbuch der Mathematik*. Verlag Harri Deutsch, Thun, expanded edition, 2001.
- [4] D. H. Colless. Phylogenetics: the theory and practice of phylogenetic systematics. *Syst. Zool.*, 31(1):100–104, 1982.
- [5] A. Drummond, G. Nicholls, A. Rodrigo, and W. Solomon. Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. *Genetics*, 161(3):1307–1320, 2002.
- [6] A. Drummond and A. Rambaut. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.*, 7:214, 2007.
- [7] A. W. F. Edwards. Estimation of the branch points of a branching diffusion process. (With discussion.). *J. Roy. Statist. Soc. Ser. B*, 32:155–174, 1970.
- [8] W. Feller. *An introduction to probability theory and its applications. Vol. I*. Third edition. John Wiley & Sons Inc., New York, 1968.
- [9] R. Fisher. *The Genetical Theory of Natural Selection*. Clarendon, Oxford, 1930.
- [10] D. Ford, T. Gernhard, and E. Matsen. A method for investigating relative timing information on phylogenetic trees. *Syst. Biol.*, in press, 2009.
- [11] T. Gernhard. The conditioned reconstructed process. *J. Theo. Biol.*, 253(4):769–778, 2008.
- [12] T. Gernhard. New analytic results for speciation times in neutral models. *Bull. Math. Biol.*, 70(4):1082–1097, 2008.

- [13] P. H. Harvey, R. M. May, and S. Nee. Phylogenies without fossils. *Evolution*, 48:523–529, 1994.
- [14] D. G. Kendall. On the generalized “birth-and-death” process. *Ann. Math. Statist.*, 19(1):1–15, 1948.
- [15] J. F. C. Kingman. The coalescent. *Stochastic Processes and Their Applications*, 13:235–248, 1982.
- [16] J. F. C. Kingman. Exchangeability and the evolution of large populations. *Exchangeability in Probability and Statistics*, pages 97–112, 1982.
- [17] J. F. C. Kingman. On the genealogy of large populations. *J. Appl. Probab.*, 19A:27–43, 1982.
- [18] L. Popovic. Asymptotic genealogy of a critical branching process. *Ann. Appl. Probab.*, 14(4):2120–2148, 2004.
- [19] O. G. Pybus and P. H. Harvey. Testing macro-evolutionary models using incomplete molecular phylogenies. *Proc. R. Soc. London*, 267(1459):2267–2272, 2000.
- [20] S. C. Ray, R. R. Arthur, A. Carella, J. Bukh, and D. L. Thomas. Genetic epidemiology of hepatitis C virus throughout Egypt. *J. Infect. Dis.*, 182(3):698–707, Sep 2000.
- [21] J. Slowinski. Probabilities on  $n$ -trees under two models: a demonstration that asymmetrical interior nodes are not improbable. *Systematic Zoology*, 39:89–94, 1990.
- [22] T. Stadler. Cass. <http://www.tb.ethz.ch/people/tstadler>, 2006–2008.
- [23] T. Stadler. Evolving trees - models for speciation and extinction in phylogenetics. *PhD thesis, Technical University of Munich*, 2008.
- [24] T. Stadler. Lineages-through-time plots of neutral models for speciation. *Mathematical Biosciences*, 216:163–171, 2008.
- [25] M. Tanaka, A. Francis, F. Luciani, and S. Sisson. Using approximate Bayesian computation to estimate tuberculosis transmission parameters from genotype data. *Genetics*, 173(3):1511, 2006.
- [26] E. A. Thompson. *Human evolutionary trees*. Cambridge University Press, 1975.
- [27] I. Volkov, J. Banavar, S. Hubbell, and A. Maritan. Neutral theory and relative species abundance in ecology. *Nature*, 424(6952):1035–1037, 2003.
- [28] S. Wright. Evolution in mendelian populations. *Genetics*, 16:97–159, 1931.
- [29] Z. Yang and B. Rannala. Bayesian phylogenetic inference using DNA sequences: A markov chain monte carlo method. *Mol. Biol. Evol.*, 17(7):717–724, 1997.

[30] G. U. Yule. A mathematical theory of evolution: based on the conclusions of Dr. J.C. Willis. *Philos. Trans. Roy. Soc. London Ser. B*, 213:21–87, 1924.

## A. Appendix

*Proof of Lemma 3.1.* For complete sampling and conditioning the tree on having  $n$  individuals today, the time of origin has density function [11],

$$q^{\lambda, \mu}(\tau|n) = n\lambda^n(\lambda - \mu)^2 \frac{(1 - e^{-(\lambda - \mu)\tau})^{n-1} e^{-(\lambda - \mu)\tau}}{(\lambda - \mu e^{-(\lambda - \mu)\tau})^{n+1}}.$$

We now assume for each individual a constant sampling probability  $\rho$ . Therefore, given we sample  $n$  individuals in a tree of age  $\tau$ , the probability that the original tree has  $m$  individuals, is, with Bayes' law and  $\int_0^\infty \mathbb{P}[m|\tau]d\tau = \frac{1}{m\lambda}$  [11],

$$\begin{aligned} \mathbb{P}[m|n] &= \mathbb{P}[n|m] \frac{\int_0^\infty \mathbb{P}[m|\tau]\mathbb{P}[\tau]d\tau}{\int_0^\infty \mathbb{P}[n|\tau]\mathbb{P}[\tau]d\tau} \\ &= \binom{m}{n} \rho^n (1 - \rho)^{m-n} \frac{\int_0^\infty \mathbb{P}[m|\tau]d\tau}{\int_0^\infty \mathbb{P}[n|\tau]d\tau} \\ &= \binom{m}{n} \rho^n (1 - \rho)^{m-n} \frac{1}{m\lambda \int_0^\infty \mathbb{P}[n|\tau]d\tau} \\ &= C \frac{\binom{m}{n} (1 - \rho)^m}{m} \end{aligned}$$

where  $C$  is the normalizing constant. Using the property from [3],

$$\sum_{m=0}^{\infty} \binom{m+n}{n} a^m = \frac{1}{(1-a)^{n+1}}. \quad (16)$$

we obtain,

$$\begin{aligned} \frac{1}{C} &= \sum_{m=n}^{\infty} \frac{\binom{m}{n} (1 - \rho)^m}{m} \\ &= \frac{1}{n} \sum_{m=n}^{\infty} \binom{m-1}{n-1} (1 - \rho)^m \\ &= \frac{1}{n} (1 - \rho)^n \sum_{m=0}^{\infty} \binom{n-1+m}{n-1} (1 - \rho)^m \\ &= \frac{1}{n} (1 - \rho)^n \rho^{-n}. \end{aligned}$$

This yields,

$$\mathbb{P}[m|n] = \binom{m-1}{n-1} (1 - \rho)^{m-n} \rho^n.$$

The density for the time of origin under random sampling for a tree on  $n$  individuals is therefore,

$$\begin{aligned} q_\rho^{\lambda, \mu}(\tau|n) &= \sum_{m=n}^{\infty} q^{\lambda, \mu}(\tau|m) \mathbb{P}[m|n] \\ &= n(\lambda - \mu)^2 (1 - \rho)^{-n} \rho^n \sum_{m=n}^{\infty} \binom{m}{n} (1 - \rho)^m \lambda^m \frac{(1 - e^{-(\lambda - \mu)\tau})^{m-1} e^{-(\lambda - \mu)\tau}}{(\lambda - \mu e^{-(\lambda - \mu)\tau})^{m+1}} \\ &= n(\lambda - \mu)^2 (1 - \rho)^{-n} \rho^n \sum_{m=0}^{\infty} \binom{m+n}{n} (1 - \rho)^{m+n} \lambda^{m+n} \frac{(1 - e^{-(\lambda - \mu)\tau})^{m+n-1} e^{-(\lambda - \mu)\tau}}{(\lambda - \mu e^{-(\lambda - \mu)\tau})^{m+n+1}} \\ &= n(\lambda - \mu)^2 (\lambda \rho)^n \frac{(1 - e^{-(\lambda - \mu)\tau})^{n-1} e^{-(\lambda - \mu)\tau}}{(\lambda - \mu e^{-(\lambda - \mu)\tau})^{n+1}} \sum_{m=0}^{\infty} \binom{m+n}{n} (1 - \rho)^m \lambda^m \frac{(1 - e^{-(\lambda - \mu)\tau})^m}{(\lambda - \mu e^{-(\lambda - \mu)\tau})^m}. \end{aligned}$$

With Equation (16) we simplify  $q_\rho^{\lambda,\mu}(\tau|n)$ ,

$$\begin{aligned}
q_\rho^{\lambda,\mu}(\tau|n) &= n(\lambda - \mu)^2(\lambda\rho)^n \frac{(1 - e^{-(\lambda-\mu)\tau})^{n-1} e^{-(\lambda-\mu)\tau}}{(\lambda - \mu e^{-(\lambda-\mu)\tau})^{n+1} (1 - (1 - \rho)\lambda \frac{1 - e^{-(\lambda-\mu)\tau}}{\lambda - \mu e^{-(\lambda-\mu)\tau}})^{n+1}} \\
&= n(\lambda - \mu)^2(\lambda\rho)^n \frac{(1 - e^{-(\lambda-\mu)\tau})^{n-1} e^{-(\lambda-\mu)\tau}}{(\lambda - \mu e^{-(\lambda-\mu)\tau} - (1 - \rho)\lambda(1 - e^{-(\lambda-\mu)\tau}))^{n+1}} \\
&= n(\lambda - \mu)^2(\lambda\rho)^n \frac{(1 - e^{-(\lambda-\mu)\tau})^{n-1} e^{-(\lambda-\mu)\tau}}{(\rho\lambda + (\lambda(1 - \rho) - \mu)e^{-(\lambda-\mu)\tau})^{n+1}}
\end{aligned}$$

which establishes the lemma.

*Proof of Theorem 3.3.* Equations (2), (3) and (4) yield,

$$\begin{aligned}
f_\rho^{\lambda,\mu}(x|n) &= \int_{x_1}^{\infty} f_\rho^{\lambda,\mu}(x, \tau|n) d\tau = \int_{x_1}^{\infty} f_\rho^{\lambda,\mu}(x|n, \tau) q_\rho^{\lambda,\mu}(\tau|n) d\tau \\
&= n!(\lambda - \mu)^2(\lambda\rho)^n \left( \prod_{i=1}^{n-1} \frac{(\lambda - \mu)^2 e^{-(\lambda-\mu)x_i}}{(\rho\lambda + (\lambda(1 - \rho) - \mu)e^{-(\lambda-\mu)x_i})^2} \right) \times \\
&\quad \int_{x_1}^{\infty} \left( \frac{\rho\lambda + (\lambda(1 - \rho) - \mu)e^{-(\lambda-\mu)\tau}}{1 - e^{-(\lambda-\mu)\tau}} \right)^{n-1} \frac{(1 - e^{-(\lambda-\mu)\tau})^{n-1} e^{-(\lambda-\mu)\tau}}{(\rho\lambda + (\lambda(1 - \rho) - \mu)e^{-(\lambda-\mu)\tau})^{n+1}} d\tau \\
&= n!(\lambda - \mu)(\lambda\rho)^n \left( \prod_{i=1}^{n-1} \frac{(\lambda - \mu)^2 e^{-(\lambda-\mu)x_i}}{(\rho\lambda + (\lambda(1 - \rho) - \mu)e^{-(\lambda-\mu)x_i})^2} \right) \times \\
&\quad \int_0^{e^{-(\lambda-\mu)x_1}} \frac{1}{(\rho\lambda + (\lambda(1 - \rho) - \mu)x)^2} dx \\
&= n!(\lambda - \mu)(\lambda\rho)^n \left( \prod_{i=1}^{n-1} \frac{(\lambda - \mu)^2 e^{-(\lambda-\mu)x_i}}{(\rho\lambda + (\lambda(1 - \rho) - \mu)e^{-(\lambda-\mu)x_i})^2} \right) \times \\
&\quad \left( \frac{1}{(\lambda(1 - \rho) - \mu)\rho\lambda} - \frac{1}{(\lambda(1 - \rho) - \mu)(\rho\lambda + (\lambda(1 - \rho) - \mu)e^{-(\lambda-\mu)x_1})} \right) \\
&= n!(\lambda - \mu)(\lambda\rho)^{n-1} \frac{e^{-(\lambda-\mu)x_1}}{\rho\lambda + (\lambda(1 - \rho) - \mu)e^{-(\lambda-\mu)x_1}} \left( \prod_{i=1}^{n-1} \frac{(\lambda - \mu)^2 e^{-(\lambda-\mu)x_i}}{(\rho\lambda + (\lambda(1 - \rho) - \mu)e^{-(\lambda-\mu)x_i})^2} \right).
\end{aligned}$$

The distribution for  $\mu \rightarrow \lambda$  is obtained from the density  $f_\rho^{\lambda,\mu}(x|n)$  by using the property  $1 - x \sim e^{-x}$  for  $x \rightarrow 0$ . This completes the proof of the theorem.  $\square$

*Proof of Theorem 4.1.* Consider a birth-death tree on  $m$  individuals. Randomly sample 2 individuals. The probability that they coalesce in the  $k$ -th bifurcation event is,

$$\begin{aligned}
p_k^2 &= \left(1 - \frac{1}{\binom{m}{2}}\right) \left(1 - \frac{1}{\binom{m-1}{2}}\right) \cdots \left(1 - \frac{1}{\binom{k+2}{2}}\right) \frac{1}{\binom{k+1}{2}} \\
&= \frac{((\binom{m}{2} - 1) (\binom{m-1}{2} - 1) \cdots (\binom{k+2}{2} - 1) \binom{k}{2} \binom{k-1}{2} \cdots \binom{2}{2})}{\prod_{k=2}^m \binom{k}{2}} \\
&= \frac{\left(\prod_{i=k+2}^m \frac{(i+1)(i-2)}{2}\right) \frac{k!(k-1)!}{2^{k-1}}}{\frac{m!(m-1)!}{2^{m-1}}} \\
&= \frac{\frac{(m+1)!(m-2)!}{(k+2)!(k-1)! 2^{m-k-1}} \frac{k!(k-1)!}{2^{k-1}}}{\frac{m!(m-1)!}{2^{m-1}}} \\
&= \frac{2(m+1)}{(m-1)(k+1)(k+2)}. \tag{17}
\end{aligned}$$

Let  $g_k^{\lambda,\mu}(x|m)$  be the density of the time of the  $k$ -th bifurcation event ( $k = 1, \dots, m-1$ ) in a tree

on  $m$  leaves [11],

$$g_k^{\lambda, \mu}(x|m) = (k+1) \binom{m}{k+1} \lambda^{m-k} (\lambda - \mu)^{k+2} e^{-(\lambda - \mu)(k+1)x} \frac{(1 - e^{-(\lambda - \mu)x})^{m-k-1}}{(\lambda - \mu e^{-(\lambda - \mu)x})^{m+1}}. \quad (18)$$

Therefore the time of coalescence of two randomly picked individuals in a tree with  $m$  individuals is,

$$\begin{aligned} f_m^{\lambda, \mu}(x|2) &= \sum_{k=1}^{m-1} p_k^2 g_k^{\lambda, \mu}(x|m) \\ &= \sum_{k=1}^{m-1} \frac{2(m+1)}{(m-1)(k+1)(k+2)} (k+1) \binom{m}{k+1} \lambda^{m-k} (\lambda - \mu)^{k+2} \times \\ &\quad e^{-(\lambda - \mu)(k+1)x} \frac{(1 - e^{-(\lambda - \mu)x})^{m-k-1}}{(\lambda - \mu e^{-(\lambda - \mu)x})^{m+1}} \\ &= \frac{2\lambda^{m+2} (1 - e^{-(\lambda - \mu)x})^{m+1}}{(m-1)e^{-(\lambda - \mu)x} (\lambda - \mu e^{-(\lambda - \mu)x})^{m+1}} \sum_{k=1}^{m-1} \binom{m+1}{k+2} \left( \frac{\lambda - \mu}{\lambda} \frac{e^{-(\lambda - \mu)x}}{(1 - e^{-(\lambda - \mu)x})} \right)^{k+2} \\ &= \frac{2\lambda e^{(\lambda - \mu)x}}{m-1} \left( \frac{1}{a+1} \right)^{m+1} \left[ (a+1)^{m+1} - \binom{m+1}{2} a^2 - (m+1)a - 1 \right] \end{aligned} \quad (19)$$

where  $a := \frac{\lambda - \mu}{\lambda} \frac{e^{-(\lambda - \mu)x}}{(1 - e^{-(\lambda - \mu)x})}$ .

For  $\mu \rightarrow \lambda$ , we obtain with the property  $1 - x \sim e^{-x}$  for  $x \rightarrow 0$ ,

$$\begin{aligned} f_m^{\lambda, \mu}(x|2) &= \sum_{k=1}^{m-1} p_k f_k^{\lambda, \mu}(x|m) \\ &= \sum_{k=1}^{m-1} \frac{2(m+1)}{(m-1)(k+1)(k+2)} (k+1) \binom{m}{k+1} \lambda^{n-k} \frac{x^{n-k-1}}{(1 + \lambda x)^{n+1}} \\ &= \frac{2\lambda^{m+2} x^{m+1}}{(m-1)(1 + \lambda x)^{m+1}} \sum_{k=1}^{m-1} \binom{m+1}{k+2} \frac{1}{(\lambda x)^{k+2}} \\ &= \frac{2\lambda}{m-1} \left( \frac{\lambda x}{1 + \lambda x} \right)^{m+1} \left[ \left( 1 + \frac{1}{\lambda x} \right)^{m+1} - \binom{m+1}{2} \frac{1}{(\lambda x)^2} - \frac{m+1}{\lambda x} - 1 \right]. \end{aligned}$$

This establishes the theorem.  $\square$

**Lemma A.1.** *Assuma a uniform prior on  $[0, C]$  for  $\lambda - \mu$  and a uniform prior on  $[0, 1]$  for  $\mu/\lambda$ . This yields a uniform prior on  $[0, \infty) \times [0, \infty)$  for the pair  $(\lambda, \mu)$  with  $0 \leq \lambda - \mu \leq C$ .*

*Proof.* First note that for a given  $d, f$  with  $\lambda - \mu = d$  and  $\mu/\lambda = f$ , we have the unique solutions  $\lambda = \frac{d}{1-f}$  and  $\mu = \frac{df}{1-f}$ , i.e. for a given  $d, f$ , there is only one corresponding pair  $\lambda, \mu$ . Vice versa, for given  $\lambda, \mu$ , the values  $d, f$  are uniquely defined as  $d = \lambda - \mu$  and  $f = \mu/\lambda$ . We have,

$$\begin{aligned} \mathbb{P}[\lambda = x, \mu = y] &= \mathbb{P}\left[\frac{\mu}{\lambda} = \frac{y}{x}, \lambda - \mu = x - y\right] \\ &= \mathbb{P}\left[f = \frac{y}{x}, d = x - y\right] \\ &= \mathbb{P}\left[f = \frac{y}{x}\right] \mathbb{P}[d = x - y] \\ &= 1 \cdot \frac{1}{C} = \frac{1}{C} \end{aligned}$$

which is the uniform distribution.  $\square$