ESTIMATING SPECIATION AND EXTINCTION RATES USING PHYLOGENIES: DEVELOPMENT AND IMPLEMENTATION OF A PROBABILISTIC MODEL

August 14, 2012

Prince Peprah Osei
`prince-peprah.osei@polytechnique.edu`

Supervisors:
Hélène Morlon, École polytechnique, CMAP, Paris
Amandine Veber, École polytechnique, CMAP, Paris

Author Summary

Is estimation of speciation and extinction rates improved when informations on both branching times and clade size/age are used? Is the use of either the branching times or clade size/age sufficient to infer the true rates of birth-and-death process? These are questions about the true rates of the underlying birth-and-death process. Understanding these dynamics is central to our knowledge of how species diversify across groups and regions. We have build on the approaches adopted by previous authors to infer diversification rates.

In particular, we develop and implement a probabilistic model using these two informations together. With simulations, we demonstrate that our approach is robust in estimating speciation and extinction rates. Applying our approach to the well-studied phylogeny of the cetaceans, we found that estimation of speciation and extinction rates is improved when informations on both branching times and the age/size of clades are used simultaneously.

Abstract

Abstract

Diversification rate, i.e. the speed at which lineages speciate or go extinct, is one of the most important metric in ecology and evolutionary biology. Approaches have been developed to estimate rates of speciation and extinction using the molecular phylogenies of extant species (tree describing the evolutionary relationships among species). The general approach consists in deriving the likelihood of birth-death models of cladogenesis given a phylogenetic tree, and to estimate the corresponding birth (speciation) and death (extinction) rates by maximum likelihood. So far, these likelihood approaches have focused on using either the branching times in phylogenies, or the age and size of clades, but not both simultaneously. Here, we develop an approach that uses information on both branching times and clade age/size to improve our estimation of speciation and extinction rates from phylogenies. We derive the joint likelihood of branching times and clade age/size, test the performance of the approach on simulated trees, and apply the approach to the empirical phylogeny of the cetaceans.

Contents

1 Introduction

Diversification rate, i.e. the speed at which lineages speciate or go extinct, is one of the most important metric in ecology and evolutionary biology. In particular, estimating these rates opens the possibility to study the dynamics of biodiversity over geological time scales. Historically, paleobiologists have estimated speciation and extinction rates using fossils. However, many groups of species (e.g. terrestrial insects, birds, plants, etc.) have no or very few fossils. Thus, methods have been developed to estimate diverisifation rates using contemporary data.

In particular, the phylogenies of extant species (trees describing the evolutionary relationships among species) have been used to estimate speciation and extinction rates. Two approaches have been developed: one uses information on the branching times separating nodes in a phylogeny, while the other one uses information on the age and size of clades. Approaches based on branching times were introduced by Nee et al., (1994) and further developed by Rabosky (2007), T. Stadler (2009) and Morlon et al., (2011). They consist in deriving the probability of observing the branching times observed in a phylogeny under various birth-death models of cladogenesis. Approaches based on the age and size of clades were introduced by Magallón and Sanderson (2001). They consist in deriving the probability of observing a given number of species today in a clade of a given age, assuming a birth-death models of cladogenesis underlies the diversification of this clade. Here we build on these stochastic birth-death models and likelihood approaches in order to infer rates of speciation and extinction using information on both branching times and the age/size of clades. We consider the simple situation in which the birth and death rates are constant over time, but this can easily be generalized to the case where the rates are time-dependent.

This report is organised as follows. In Section 2, we explain the two approaches adopted by previous authors to estimate the rates of diversification, and derive the joint likelihood of branching times and clade size as a basis for estimating diversification rates. In section 3, we test this approach using simulations and apply the approach to the well-studied phylogeny of the cetaceans.

2 Derivation of the joint likelihood

2.1 Estimation of speciation and extinction rates

Using molecular data, we can estimate the evolutionary tree joining species that are alive today. This tree, which does not contain any information on the extinct species is called a reconstructed phylogeny. A seminal paper by Nee et al., (1994) showed how such a reconstructed phylogeny-in particular the branching times separating nodes in the phylogeny–can be used to estimate rates of speciation and extinction. An other seminal paper by Magallón and Sanderson (2001) showed how to use the clade size and clade age estimated from molecular data to estimate net diversification rates.

In both approaches, the underlying idea is to model cladogenesis using the generalized birth-and-death process originally introduced by Kendall (1948). Each lineage gives rise to a new lineage (i.e. branches into 2 lineages) at a fixed rate λ (speciation) and go extinct at a fixed rate μ (extinction).

Here, I detail the two approaches by Nee et al., (1994) and Magallón and Sanderson (2001).

2.2 The branching times approach

Given a molecular phylogeny, Neet et al., (1994) constructed a likelihood function for the reconstructed phylogeny. The data set has the form $\{t_2, t_3, \ldots, t_N\}$, reporting the times when the second, third, ... N-th lineages first appear, where N is the total number of lineages in the phylogeny (see Figure 1). We define be $x_n \equiv T - t_n$, where T is the time of origin of the process: thus, x_n is the amount of time between the present and the birth of the nth lineage (Figure 1 has the schematic illustration of the branching times). What contributes to the likelihood are the birth events at t_3 and t_4 and the total amount of time during which the lineages do not give birth.
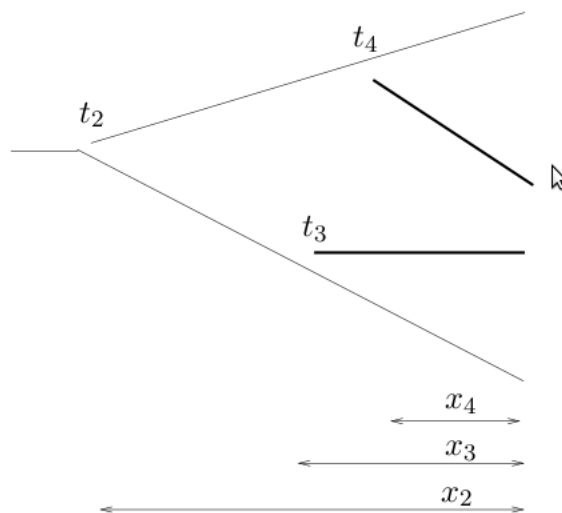

Figure 1: Schematic figure illustrating the branch times, $\{t_2, t_3, t_4\}$. The t_i's are the actual dates of the nodes and the x_i's are the length of time elapsed between the nodes and the present day.

Using this information, Nee et al., (1994) computed the likelihood function as

$$\mathcal{L}_B(t_2, \ldots, t_N) = (N-1)! \lambda^{N-2} \left\{ \prod_{i=3}^{N} P(t_i, T) \right\} (1 - u_{x_2}) \prod_{i=3}^{N} (1 - u_{x_i})$$

where $P(t_i, T)$ is the probability that a single lineage at time t_i has some descendant at later time T and $(1 - u_{x_i})$ is the probability of single progeny (i.e. no further speciation) after an amount of time x_i. The birth events at time t_i has probability proportional to $(i-1)\lambda P(t_i, T)$, where $\lambda P(t, T)$ is the birth rate in the generalised birth-death process. This likelihood can be used to estimate λ (speciation rate) and μ (extinction rate) by maximization of the likelihood function.

2.3 Using clade size and clade age

Magallón and Sanderson (2001) proposed and estimation of the net diversification rate $r = \lambda - \mu$ that uses the present day species diverisity (clade size) and its age. To this end, they make a distinction between the age of the clade's stem lineage and the age of its crown group (Figure 2 explains the difference between these two ages of a clade). The maximum likelihood estimators of diversification rates at time t they obtain are

$$\hat{r}_s = \hat{\lambda}_s = \frac{\log(n)}{t}$$

for stem group age and

$$\hat{r}_c = \hat{\lambda}_c = \frac{\log(n) - \log 2}{t}$$

for a crown group age, assuming that the extinction rate is negligible. But stochastic extinction in a birth and death process causes these estimators to be biased due to presence of extinction. They used the mean clade size over time, conditioned on the survival of the clade, computed by Raup (1985) as

$$\bar{N}(t) = \frac{\epsilon e^{rt}}{1 - \alpha^\epsilon}$$

where ϵ is the number of lineages present at the start of the birth and death process and α is the probability of observing zero descendant at any time t for a process that started before t.

Magallón and Sanderson (2001) proposed method-of-moments estimator as an alternative method to maximum-likelihood estimators given above to address the issue of presence of extinction in the birth-and-death process. Using the work done by Rohatgi (1976), they equated the mean clade size over time with an observaton on diverisity, n,. Their improved estimation of r is then given by

$$\hat{r}_{as} = \frac{1}{t} \log \left\{ n(1 - a) + a \right\}$$
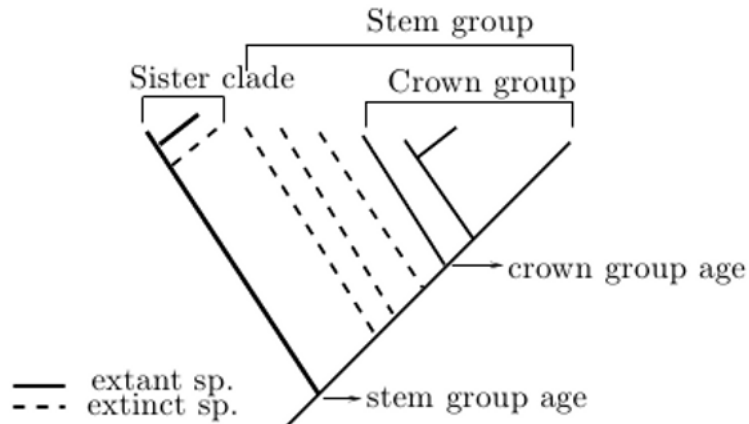
for stem group age, and

Figure 2: Schematic figure illustrating the distinction between the clade stem group age and crown group age. The age of the stem is the time of divergence of the clade from its sister taxon. The age of the crown group is the time of the deepest branching within the crown group. Bold lines denotes extant species; dotted lines represent extinct species.

$$\hat{r}_{ac} = \frac{1}{t} \left\{ \log \left[\frac{1}{2} n (1 - a^2) + 2a + \frac{1}{2} (1 - a) \sqrt{n(na^2 - 8a + 2na + n)} \right] - \log 2 \right\}$$

for crown group age. a is the relative extinction rate ($a = \frac{\mu}{\lambda}$). They applied their approach to the estimation of the rates of diversification for the angiosperms as whole, and for selected clades within the angiosperms.

2.4 Joint likelihood of the branch times and clade size

We begin by deriving the joint likelihood of the clade size n and the branching times, $\{t_2, t_3, \ldots, t_k\}$ of a sample of $k \leqslant n$ of these extant species. Let $\mathcal{L}(t_1, \ldots, t_k, n)$ denote the joint likelihood function we wish to derive. We measure time from the present to the past so that $t = 0$ denotes the present and $t = T$ denotes the origin time. t_1 denotes the first time at which the ancestral lineage came into existence, and $\{t_2, t_3, \ldots, t_k\}$ the respective times when the second, third, $\ldots k$-th lineage first appears. We assume that this clade has evolved according to a birth-death process, with speciation rate λ and extinction rate μ. We assume that each of the species in the clade is observed with probability $\rho \in]0, 1]$, independently of others. Hence, the number k of observed species is a binomial random variable with parameters n and ρ: $\mathbb{P}(k|n, \rho) = \binom{n}{k} \rho^k (1 - \rho)^{n-k}$. Here we know k, and want to reconstruct n. See Figure 3.

The joint likelihood of observing such a phylogeny conditioned on at least k extant species being sampled can be expressed as

$$\mathcal{L}(t, n|\lambda, \mu, k) = \mathrm{g}(t|n, \lambda, \mu, k)\mathbb{P}(n|\lambda, \mu, k), \tag{1}$$
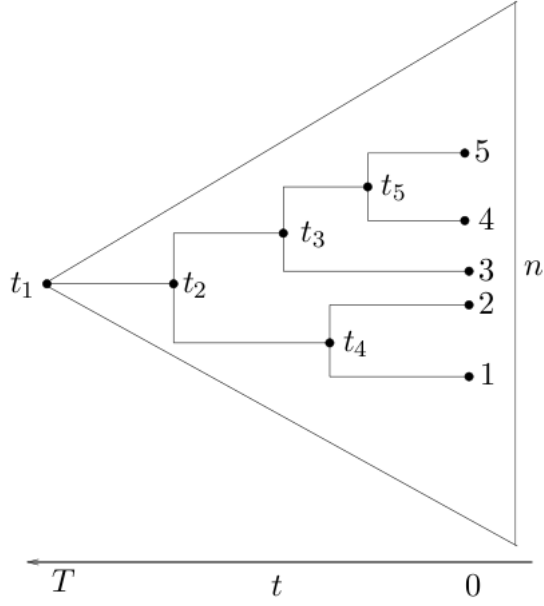
Figure 3: Schematic figure illustrating the phylogeny. The tree with branching times $\{t_2, t_3, t_4, t_5\}$ is embedded in a larger clade of size n.

where **g** denotes the probability of $t = (t_1, t_2, \ldots, t_k)$ given that there are n species in the whole clade at present and $\mathbb{P}$ denotes the probability of the clade size being n conditionally on sampling k extant species from this clade. A detailed derivation of the joint likelihood in equation (1) is given in appendix A.

Because by definition

$$\mathcal{L}(t, n | \lambda, \mu, k) = \mathbb{P}(T = t, N = n | \lambda, \mu, k),$$

it is easy to show that

$$\sum_{n=k}^{\infty} \mathcal{L}(t, n | \lambda, \mu, k) = \mathbb{P}(T = t | \lambda, \mu, k).$$

If we write $\mathcal{L}(t | \lambda, \mu, k) = \mathbb{P}(T = t | \lambda, \mu, k)$, the joint likelihood in (1) can thus, be expressed as

$$\sum_{n=k}^{\infty} \mathbf{g}(t | n, \lambda, \mu, k) \mathbb{P}(n | \lambda, \mu, k) = \sum_{n=k}^{\infty} \mathbb{P}^{\lambda, \mu, k}(T = t, N = n) \tag{2}$$
$$= \mathcal{L}(t | \lambda, \mu, k).$$

The right hand side of equation (2) is the same as Eq. (1) in Morlon et al., (2011). Thus, equation (2) becomes

$$\sum_{n=k}^{\infty} \mathsf{g}(t|n,\lambda,\mu,k)\mathbb{P}(n|\lambda,\mu,k) = \frac{\rho^k \Psi(t_2,t_1)\lambda^{k-2}\prod_{i=2}^{k}\Psi(s_{i,1},t_i)\Psi(s_{i,2},t_i)}{1-\Phi(t_1)} \tag{3}$$

$$= \mathcal{L}(t_1,\ldots,t_k).$$

where $\Psi(s,t)$ denotes the probability that a lineage alive at time t leaves exactly one surviving descendant at time $s < t$ in the reconstructed phylogeny, and $\Phi(t)$ denotes the probability that a lineage alive at time t has no descendant in the sample. $s_{i,1}$ and $s_{i,2}$ denote the times at which the daughter lineages introduced at time t_i.

In the situation where speciation rate λ, and extinction rate μ, are assumed to be constant through time making the process homogeneous, the functions $\Phi(t)$ and $\Psi(s,t)$ according to Morlon et al., (2011) are given by

$$\Phi(t) = 1 - \frac{e^{(\lambda-\mu)t}}{\frac{1}{\rho} + \frac{\lambda}{\lambda-\mu}\left(e^{(\lambda-\mu)t}-1\right)} \tag{4}$$

and

$$\Psi(s,t) = e^{(\lambda-\mu)(t-s)}\left[1 + \frac{\frac{\lambda}{\lambda-\mu}(e^{(\lambda-\mu)t}-e^{(\lambda-\mu)s})}{\frac{1}{\rho}+\frac{\lambda}{\lambda-\mu}(e^{(\lambda-\mu)s}-1)}\right]^{-2}. \tag{5}$$

In the birth-death process model, the probability that the clade size is equal to n given that k extant species are sampled with sampling fraction ρ, $\mathbb{P}(n|\lambda,\mu,k)$, can be computed as

$$\begin{aligned}
\mathbb{P}(n|\lambda,\mu,k) &= \frac{\binom{n}{k}\rho^k(1-\rho)^{n-k}\alpha^{n-1}(1-\alpha)}{\sum_{j=k}^{\infty}\binom{j}{k}\rho^k(1-\rho)^{j-k}\alpha^{j-1}(1-\alpha)}\\
&= \frac{\binom{n}{k}(1-\rho)^{n-k}\alpha^{n-k}}{\sum_{j=k}^{\infty}\binom{j}{k}(1-\rho)^{j-k}\alpha^{j-k}}
\end{aligned} \tag{6}$$

where α is the probability of observing zero descendant species. Nee et al., (1994) and Rabosky et al., (2007) provide the analytical expression of α (see appendix B for the explicit formula for α). Let

$$\phi(k,\lambda,\mu,\alpha) = \sum_{j=k}^{\infty}\binom{j}{k}(1-\rho)^{j-k}\alpha^{j-k}. \tag{7}$$

Then equation (3) becomes

$$\sum_{n=k}^{\infty}\mathsf{g}(t|n,\lambda,\mu,k)\binom{n}{k}(1-\rho)^{n-k}\alpha^{n-k} = \phi(k,\lambda,\mu,\alpha)\mathcal{L}(t|\lambda,\mu,k). \tag{8}$$

If we let $m = n-k$ and $\beta = 1-\rho$, and then differentiate equation (8) M times with respect to β we have

$$\frac{\partial^M}{\partial \beta^M} \left(\phi(k, \lambda, \mu, \beta, \alpha) \mathcal{L}(t|\lambda, \mu, k) \right) = \frac{\partial^M}{\partial \beta^M} \sum_{m=0}^{\infty} \mathbf{g}(t|m+k, \lambda, \mu, k) \binom{m+k}{k} \times$$

$$\beta^m \alpha^m \tag{9}$$

$$= \sum_{m=M}^{\infty} \alpha^m \mathbf{g}(t|m+k, \lambda, \mu, k) \binom{m+k}{k} \times$$

$$m(m-1) \ldots (m-M+1) \beta^{m-M}.$$

It is easily seen from equation (9) that for $\beta = 0$, if $m - M > 0$ then the term in the sum disappears. On the other hand if $m = M$ then for all β, $\beta^{m-M} = 1$. Thus, equation (9) becomes

$$\frac{\partial^M}{\partial \beta^M} (\phi(k, \lambda, \mu, \beta, \alpha) \mathcal{L}(t|\lambda, \mu, k)) \mid_{\beta=0} = \kappa \mathbf{g}(t|M+k, \lambda, \mu, k), \tag{10}$$

with constant of proportionality, κ, given by

$$\kappa = \frac{\alpha^M (M+k)!}{k!}.$$

Thus, the joint likelihood we are looking for from equation (1) is

$$\mathcal{L}(t, M+k|\lambda, \mu, k) = \frac{k!}{\alpha^M (M+k)!} \frac{\partial^M}{\partial \beta^M} (\phi(k, \lambda, \mu, \beta, \alpha) \mathcal{L}(t|\lambda, \mu, k)) \mid_{\beta=0} \times$$

$$\mathbb{P}(M+k|\lambda, \mu, k). \tag{11}$$

From equation (7), with the sum indexed by $N = j - k$ and with $\beta := 1 - \rho$ we have

$$\phi(k, \lambda, \mu, \alpha, \beta) = \sum_{N=0}^{\infty} \binom{k+N}{k} \alpha^N \beta^N. \tag{12}$$

Let us write $\phi(k, \lambda, \mu, \alpha, \beta) = \phi(\beta)$ to simplify the notation. Similarly, let $\Theta(\beta) = \mathcal{L}(t|\lambda, \mu, \beta, k)$:

$$\Theta(\beta) = \frac{\rho^k \Psi(t_2, t_1, \beta) \lambda^{k-2} \prod_{i=2}^{k} \Psi(s_{i,1}, t_i, \beta) \Psi(s_{i,2}, t_i, \beta)}{1 - \Phi(t_1, \beta)}. \tag{13}$$

Thus, the joint likelihood becomes

$$\mathcal{L}(t, M+k|\lambda, \mu, k) = \frac{k!}{\alpha^M (M+k)!} \left(\frac{\partial^M (\phi \Theta)}{\partial \beta^M} \mid_{\beta=0} \right) \mathbb{P}(M+k|\lambda, \mu, k) \tag{14}$$

Using the general formular for the M-th derivative of the product of two functions, we have

$$\frac{\partial^M (\phi \Theta)}{\partial \beta^M} \mid_{\beta=0} = \sum_{j=0}^{M} \binom{M}{j} \frac{\partial^j \phi}{\partial \beta^j}(0) \frac{\partial^{M-j} \Theta}{\partial \beta^{M-j}}(0). \tag{15}$$

Using the expression of the function $\phi(\beta)$ given in equation (12), for any integer j, the j-th derivative of ϕ with respect to β is equal to

$$\frac{\partial^j \phi}{\partial \beta^j}(\beta) = \sum_{N=j}^{\infty} \binom{k+N}{k} \alpha^N N(N-1)\ldots(N-j+1)\beta^{N-j}. \tag{16}$$

Taking $\beta = 0$ in the above formula, we get

$$\frac{\partial^j \phi}{\partial \beta^j}(0) = \binom{k+j}{j} \alpha^j j(j-1)\ldots 1 \tag{17}$$

$$= \frac{(k+j)!}{k!}\alpha^j. \tag{18}$$

Coming back to (15), we arrive at

$$\frac{\partial^M (\phi\Theta)}{\partial \beta^M}\big|_{\beta=0} = \sum_{j=0}^{M} \binom{M}{j} \frac{(k+j)!}{k!}\alpha^j \frac{\partial^{M-j}\Theta}{\partial \beta^{M-j}}(0). \tag{19}$$

Substituting equation (19) into the joint likelihood function in equation (14) gives

$$\mathcal{L}(t, M+k|\lambda,\mu,k) = \frac{k!}{\alpha^M (M+k)!} \left(\sum_{j=0}^{M} \binom{M}{j} \frac{(k+j)!}{k!}\alpha^j \frac{\partial^{M-j}\Theta}{\partial \beta^{M-j}}(0) \right) \mathbb{P}(M+k|\lambda,\mu,k) \tag{20}$$

3 Implementation and application

3.1 Computation of $(M-j)$-th derivative of function $\Theta(\beta)$

The joint likelihood in Section 2 equation (20) can be implemented only when the $(M-j)$-th derivative of the function $\Theta(\beta)$ is computed. The function, $\Theta(\beta)$ (see Appendix C for exact formula), is a function of product of $3k+2$ terms of β. We use a trick to compute this derivative by finding the derivative of each of the terms in the product. We differentiate each of the terms in the product and substitute $\beta = 0$. These computation can be generated into a matrix of the form

$$
\mathbf{M} = \begin{bmatrix}
(\rho e^{rt_2} - a)^2 & 2(\rho e^{rt_2} - a) & 2 & 0 & \cdots & 0 & 0 \\
(\rho e^{rs_{2,1}} - a)^2 & 2(\rho e^{rs_{2,1}} - a) & 2 & 0 & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
(\rho e^{rs_{k,1}} - a)^2 & 2(\rho e^{rs_{k,1}} - a) & 2 & 0 & \cdots & 0 & 0 \\
(\rho e^{rs_{2,2}} - a)^2 & 2(\rho e^{rs_{2,2}} - a) & 2 & 0 & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
(\rho e^{rs_{k,2}} - a)^2 & 2(\rho e^{rs_{k,2}} - a) & 2 & 0 & \cdots & 0 & 0 \\
\frac{1}{\rho e^{rt_1} - a} & \frac{-1}{(\rho e^{rt_1} - a)^2} & \frac{2}{(\rho e^{rt_1} - a)^3} & \cdots & \cdots & \cdots & \frac{(-1)^{M-j}(M-j)!}{(\rho e^{rt_1} - a)^{1+M-j}} \\
\frac{1}{(\rho e^{rt_2} - a)^4} & \frac{-4}{(\rho e^{rt_2} - a)^5} & \frac{20}{(\rho e^{rt_2} - a)^6} & \cdots & \cdots & \cdots & \frac{(-1)^{M-j}(3+M-j)!}{3!(\rho e^{rt_1} - a)^{4+M-j}} \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
\frac{1}{(\rho e^{rt_k} - a)^4} & \frac{-4}{(\rho e^{rt_k} - a)^5} & \frac{20}{(\rho e^{rt_k} - a)^6} & \cdots & \cdots & \cdots & \frac{(-1)^{M-j}(3+M-j)!}{3!(\rho e^{rt_k} - a)^{4+M-j}}
\end{bmatrix}.
$$

The rows in the matrix $\mathbf{M}$, describe each function and then its derivatives with respect to β at $\beta = 0$. The columns are the order of the derivatives of each function in the product. The first column is the $3k+2$ terms of the function $\Theta(\beta = 0)$. The second, third and up to $(M-j)$-th columns are the first, second,...$(M-j)$-th derivative of each of the terms in the product of functions.

In the next section, we develop an algorithm to compute the $(M-j)$-th derivative of Θ by using the appropriate entries of this matrix. To this end, we need an algorithm able to find the set of all possible vectors of size $3k+2$ whose entries sum to $M-j$.

3.2 The possible vectors generator

We generate this vectors by sorting in terms of the first coordinate. For example, if we let

$$D_{3k+2, M-j} = \text{set of all possible vectors of size } 3k+2 \text{ whose entries sum to } M-j,$$

then once the first coordinate is fixed to say $(M-j) - i$, then we find how to distribute k orders of derivative in a vector of size $3k+1$. Thus using the idea of integer partitions, we partition the order of the derivative $((M-j) - i)$ into vector of length $3k+2$ such that the sum of the entries is equal to $M-j$. Obtaining all these possible vectors and using the matrix $\mathbf{M}$ we can write the $(M-j)$-th derivative of the function $\Theta(\beta)$ in the form

$$\frac{\partial^{M-j}\Theta(\beta=0)}{\partial\beta^{M-j}} = \sum_{\sum_{i=1}^{3k+2} n_i = M-j} \prod_{i=1}^{3k+2} \mathbf{M}_{i,n_{i+1}}, \tag{21}$$

where n_i gives the order of derivation to each of the term in the product of functions. Thus, these possible vectors determine the combination of the entries of the matrix $\mathbf{M}$ to be multiplied together in order to obtain all the expressions in the $(M-j)$-th derivative of function $\Theta(\beta)$ at $\beta = 0$.

By lack of time to develop a full algorithm, we consider an approximated version of equation (21) in which the quotient terms in the function $\Theta(\beta)$ are differentiated $M-j$ times with respect to β and the rest of the terms being constant. This approximated version is given by

$$\begin{aligned}
\frac{\partial^{M-j}}{\partial\beta^{M-j}}(\Theta(\beta=0)) \approx & \frac{(-1)^{M-j}(M-j)!}{(\rho e^{rt_1}-a)^{1+M-j}}(\rho e^{rt_2}-a)^2 \times \\
& \left(\prod_{i=2}^{k} \frac{[(\rho e^{rs_{i,1}}-a)(\rho e^{rs_{i,2}}-a)]^2}{(\rho e^{rt_i}-a)^4}\right) + \\
& \frac{(\rho e^{rt_2}-a)^2}{\rho e^{rt_1}-a} \prod_{i=2}^{k}[(\rho e^{rs_{i,1}}-a)(\rho e^{rs_{i,2}}-a)]^2 \times \\
& \left\{\sum_{i=2}^{k} \frac{(-1)^{M-j}(3+M-j)!}{3!(\rho e^{rt_k}-a)^{4+M-j}}\right\}.
\end{aligned} \tag{22}$$

Using this approximated version of the $M-j$-th derivative and the exact joint likelihood formula in Appendix C equation (35), the joint likelihood can be written as

$$\begin{aligned}
\mathcal{L}(t, M+k|\lambda,\mu,k) = & \frac{\beta^M}{M! \sum_{N=0}^{\infty}\binom{N+k}{k}\beta^N\alpha^N} \sum_{j=0}^{M}\binom{M}{j}\frac{(k+j)!}{k!}\alpha^j \times \\
& \left\{\frac{\rho^{k-1}\lambda^{k-2}}{(1-a)}\prod_{i=2}^{k} e^{r[2t_i-(s_{i,1}+s_{i,2})]}\right\} \times \\
& \left[\frac{(-1)^{M-j}(M-j)!}{(\rho e^{rt_1}-a)^{M-j}}(\rho e^{rt_2}-a)^2\right] \times \\
& \left[\prod_{i=2}^{k}\frac{[(\rho e^{rs_{i,1}}-a)(\rho e^{rs_{i,2}}-a)]^2}{(\rho e^{rt_i}-a)^4}\right] + \\
& \left(\sum_{i=2}^{k}\frac{(-1)^{M-j}(3+M-j)!}{3!(\rho e^{rt_i}-a)^{4+M-j}}\right) \times \\
& \left(\frac{(\rho e^{rt_2}-a)^2}{\rho e^{rt_1}-a}\prod_{i=2}^{k}[(\rho e^{rs_{i,1}}-a)(\rho e^{rs_{i,2}}-a)]^2\right)
\end{aligned} \tag{23}$$

where the size of the clade, $n = M + k$. We coded this approximated version of the joint likelihood in R.

3.3 Robustness of the joint likelihood approach

Using simulations, we tested the ability of the joint likelihood approach to estimate the true parameters of birth-and-death process (speciation and extinction rates). We found that the approach performed well with either increasing or decreasing sampling fraction (see Figure 4). Figure 5 illustrates the distribution of parameter estimates across phylogenies.

3.4 Empirical phylogeny results: the cetaceans

We applied the joint likelihood to the cetacean pylogeny which is much studied in cladogenesis models. This molecular phylogeny contains 87 out of 89 extant cetacean species. Under the assumption of constant birth and death rates across the phylogeny, we found no support for the presence of extinction (Table 1).

λ	μ	LogL	AIC_c
0.275	0.20	-291.208	4.071

Table 1: Approximated joint likelihood model fitted to the cetacean phylogeny. LogL stands for the maximum log likelihood; AIC_c stands for the second order Akaike's information criterion.

Comparing our results to the work done by Morlon et al., (2011), where only branching times were used, our estimates showed much improvement in the estimation of speciation and extinction rates even after using an approximated version of the joint likelihood. On the basis of AIC_c values, we obtain a much smaller value 4.071 against 130.770 in Morlon et al., (2011) paper which used only the informations on braching times separating nodes in a phylogeny.
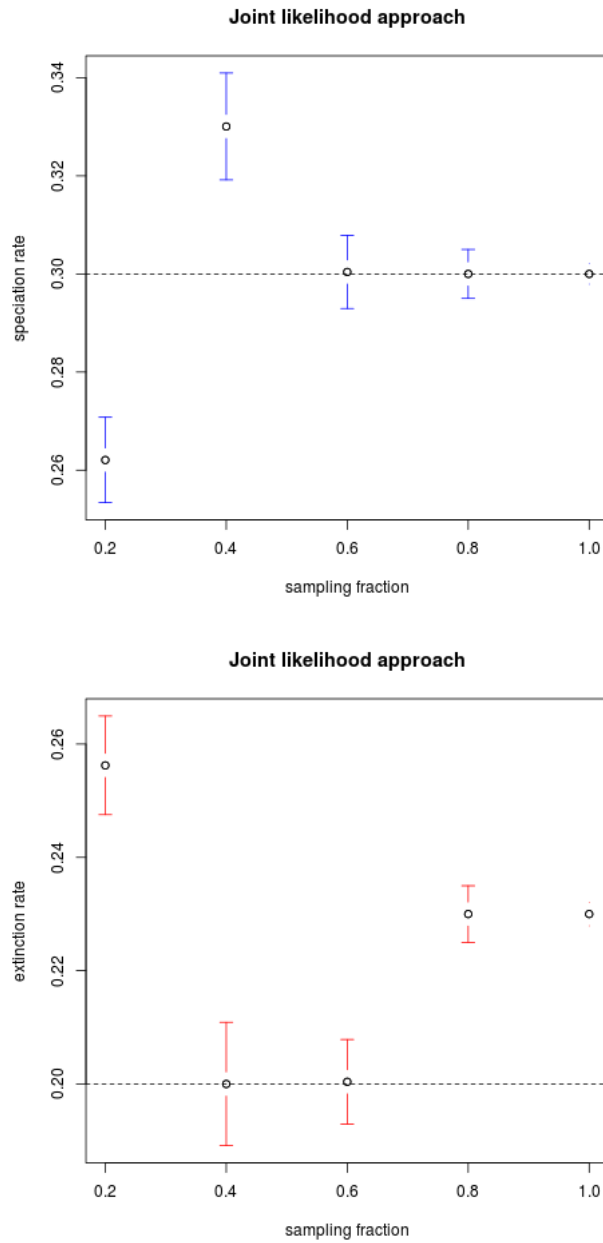
Figure 4: The joint likelihood method provides robust estimates of speciation and extinction rates. The figure shows maximum likelihood parameter estimates for phylogenies simulated under homogenous birth-and-death process. Points and error bars indicate the median and 95% quantile range of the maximum likelihood parameter estimates, across 100 simulated phylogenies for each parameter. Before estimating the parameters, species were randomly sampled from the simulated phylogenies. The sampling fraction ρ ranges from 20% (poorly sampled) to 100% (fully sampled).

Parameter estimates for simulated phylogenies

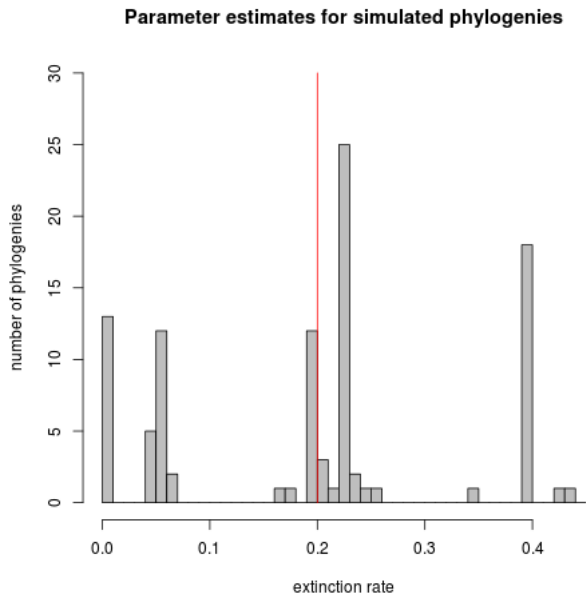

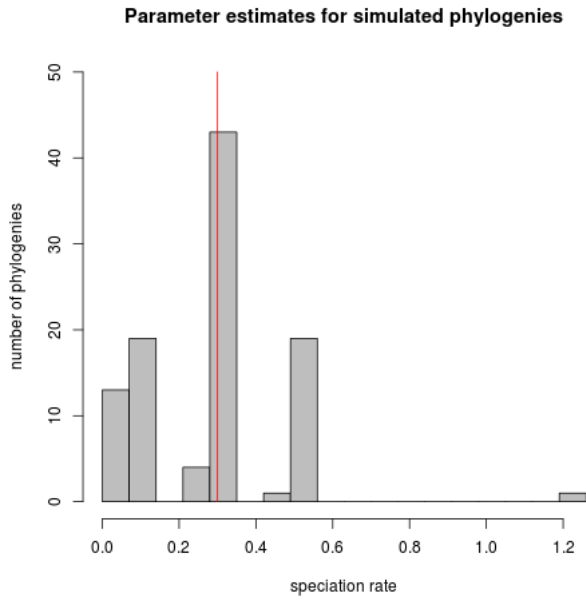**Parameter estimates for simulated phylogenies**

Figure 5: The histograms represent the distribution of parameter estimates for the 100 simulated phylogenies. The red line indicates the true simulated parameters of diversification.

4 Discussion

Estimating speciation and extinction rates opens the possibility to study the dynamics of biodiversity over geological time scales. The phylogenies of extant species can be used to estimate speciation and extinction rates. This evolutionary tree joining species that are alive today contains no information on the extinct species. Previous authors have showed that branching times seperating nodes in the phylogeny can be used to estimate rates of speciation and extinction. Other authors showed how to use the clade size and age obtained from molecular data to estimate net diversification rates. We have combined these two informations (branching times and size/age of clade) to infer rates of speciation and extinction.

The joint likelihood probabilistic model developed here is particularly well suited to the study of incomplete phylogenies. This is very useful, because fully sampled phylogenies are rarely available. Our analysis also suggests that estimation of rates of diversification can be improved if both informations on branching times and size/age of clade are used in the birth-death models of cladogenesis.

We have shown how to develop and implement the joint likelihood probabilistic model to infer the rates of speciation and extinction using both informations on branching times and age/size of clade. We have tested the performance of our approach on simulated phylogenetic trees and demonstrate the robustness of our approach. We applied our approach to the well-studied phylogeny of the cetaceans and confirmed the results in Morlon et al., (2011) that considering the cetacean phylogeny as a whole give no support for the presence of extinction. This may be due to the implicit assumption that diverisfication rates are homogenous across lineages.

There are several potential extensions and applications of the joint likelihood approach in macro-evolution. First, the joint likelihood approach should allow us to incorporate information from fossil data. Obviously, incorporation of fossil data to phylogenetic inference will improve our ability to understand long term diversity dynamics of biodiversity. Second, the assumption that speciation rate is always greater than or equal to the extinction rate could be relaxed. This is biologically relevant and might influence our conclusions.

There are nevertheless limitations to our approach. We used an approximated version of the joint likelihood and thus, do not rule out the possibility that the full version of the joint likelihood would provide even much better estimate of speciation and extinction rates. Another major limitation of our approach is that we did not account for rate variation across lineages and in time. All these limitations remain topics for future research. Empirical phylogenies are more imbalanced than predicted by models with homogenous rates (Morlon et al., 2010), and inferences based on models with homogenous rates might be biased.

A   Derivation of the joint likelihood

Here we explain the decomposition of the likelihood expression in Section 2.4, equation (1). By the definition of the likelihood we have

$$\mathcal{L}(t, n | \lambda, \mu, k) = \mathbb{P}(T = t, N = n, | \lambda, \mu, k)$$
$$= \mathbb{P}^{\lambda, \mu, k}(T = t, N = n)$$
$$= \mathbb{P}^{\lambda, \mu, k}(T = t | N = n) \mathbb{P}^{\lambda, \mu, k}(N = n)$$

where the symbol $\mathbb{P}^{\lambda, \mu, k}$ denote conditional probability.

This implies

$$\mathcal{L}(t, n | \lambda, \mu, k) = \mathbb{P}^{\lambda, \mu, k}(T = t | N = n) \mathbb{P}(n | \lambda, \mu, k) \tag{24}$$

Then,

$$\mathbb{P}^{\lambda, \mu, k}(T = t | N = n) = \frac{\mathbb{P}^{\lambda, \mu, k}(T = t, N = n)}{\mathbb{P}^{\lambda, \mu, k}(N = n)}$$
$$= \frac{\mathbb{P}(T = t, N = n, \lambda, \mu, k)}{\mathbb{P}(\lambda, \mu, k)} \cdot \frac{1}{\frac{\mathbb{P}(N = n, \lambda, \mu, k)}{\mathbb{P}(\lambda, \mu, k)}}$$
$$= \frac{\mathbb{P}(T = t, N = n, \lambda, \mu, k)}{\mathbb{P}(N = n, \lambda, \mu, k)}$$
$$= \mathbb{P}(T = t | N = n, \lambda, \mu, k).$$

Denoting the expression $\mathbb{P}(T = t | N = n, \lambda, \mu, k)$ by $\mathsf{g}(t | n, \lambda, \mu, k)$, we obtain the expression in equation (1)

$$\mathcal{L}(t, n | \lambda, \mu, k) = \mathsf{g}(t | n, \lambda, \mu, k) \mathbb{P}(n | \lambda, \mu, k). \tag{25}$$

B   Explicit expression for α

To compute α (the probability of observing zero descendant species), we first need the probability $\mathbb{P}(t, T)$, that a single lineage alive at time t has at least one surviving descendant at time T, assuming that it evolves according to a birth and death process between the times t and T. This probability is given in Nee et al., (1994):

$$\mathbb{P}(t, T) = \frac{\lambda - \mu}{\lambda - \mu e^{-(\lambda - \mu)(T - t)}}. \tag{26}$$

Note that, $\mathbb{P}(0, T)$ is the probability that a birth-death process which starts at time 0 with a single lineage is not extinct at time T. The constant α is then defined by

$$\alpha = 1 - \mathbb{P}(0, T)$$

$$= 1 - \frac{\lambda - \mu}{\lambda - \mu e^{-(\lambda - \mu)T}}$$

$$= \frac{\mu - \mu e^{-(\lambda - \mu)T}}{\lambda - \mu e^{-(\lambda - \mu)T}} \tag{27}$$

$$= \frac{\mu(e^{(\lambda - \mu)T} - 1)}{\lambda(e^{(\lambda - \mu)T} - \frac{\mu}{\lambda})}$$

$$\alpha = a \left(\frac{e^{rT} - 1}{e^{rT} - a} \right),$$

where $a = \frac{\mu}{\lambda}$ and $r = \lambda - \mu$.

C   The exact expression of the joint likelihood

From equation (20) in Section 2, the joint likelihood is given by

$$\mathcal{L}(t, M + k | \lambda, \mu, k) = \frac{k!}{\alpha^M (M + k)!} \left(\sum_{j=0}^{M} \binom{M}{j} \frac{(k+j)!}{k!} \alpha^j \frac{\partial^{M-j} \Theta}{\partial \beta^{M-j}}(0) \right) \times$$
$$\mathbb{P}(M + k | \lambda, \mu, \rho, k), \tag{28}$$

where

$$\Theta(\beta) = \frac{\rho^k \Psi(t_2, t_1, \beta) \lambda^{k-2} \prod_{i=2}^{k} \Psi(s_{i,1}, t_i, \beta) \Psi(s_{i,2}, t_i, \beta)}{1 - \Phi(t_1, \beta)} \tag{29}$$

and

$$\mathbb{P}(M + k | \lambda, \mu, \rho, k) = \frac{\binom{M+k}{k} \beta^M \alpha^M}{\sum_{N=0}^{\infty} \binom{N+k}{k} \beta^N \alpha^N}. \tag{30}$$

We can thus give an exact expression of the function $\Theta(\beta)$ by simplifying the functions Φ and Ψ from equations (5) and (7) in Section 2.

$$\Phi(t, \beta) = 1 - \frac{e^{(\lambda - \mu)t}}{\frac{1}{\rho} + \frac{\lambda}{\lambda - \mu} \left(e^{(\lambda - \mu)t} - 1 \right)}$$

$$= 1 - \frac{\rho(\lambda - \mu)e^{(\lambda - \mu)t}}{\lambda - \mu + \rho\lambda(e^{(\lambda - \mu)t} - 1)} \tag{31}$$

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

$$= \frac{\beta + a(\rho e^{rt} - 1)}{\beta + \rho e^{rt} - a},$$

where

$$\beta = 1 - \rho, \qquad a = \frac{\mu}{\lambda}, \qquad \text{and} \qquad r = \lambda - \mu.$$

Here we emphasize the dependence on β through the notation $\Phi(t, \beta)$. Then $1 - \Phi(t, \beta)$ can be computed as

$$1 - \Phi(t, \beta) = \frac{\rho e^{rt}(1 - a)}{\beta + \rho e^{rt} - a}. \tag{32}$$

We also simplify the function Ψ as follows

$$
\begin{aligned}
\Psi(s, t, \beta) &= e^{(\lambda-\mu)(t-s)} \left[1 + \frac{\frac{\lambda}{\lambda-\mu}(e^{(\lambda-\mu)t} - e^{(\lambda-\mu)s})}{\frac{1}{\rho} + \frac{\lambda}{\lambda-\mu}(e^{(\lambda-\mu)s} - 1)} \right]^{-2} \\
&= e^{(\lambda-\mu)(t-s)} \left[1 + \frac{\rho\lambda(e^{(\lambda-\mu)t} - e^{(\lambda-\mu)s})}{\lambda - \mu + \rho\lambda(e^{(\lambda-\mu)s} - 1)} \right]^{-2} \\
&= e^{(\lambda-\mu)(t-s)} \left[\frac{(1 - \rho) + \rho e^{(\lambda-\mu)t} - \frac{\lambda}{\mu}}{(1 - \rho) + \rho e^{(\lambda-\mu)s} - \frac{\lambda}{\mu}} \right]^{-2} \\
&= e^{r(t-s)} \left[\frac{\beta + \rho e^{rt} - a}{\beta + \rho e^{rs} - a} \right]^{-2} \\
&= e^{r(t-s)} \left[\frac{\beta + \rho e^{rs} - a}{\beta + \rho e^{rt} - a} \right]^{2}
\end{aligned} \tag{33}
$$

Hence from equation (32) and (33), the likelihood function $\Theta(\beta)$ in equation (29) is equal to

$$
\begin{aligned}
\Theta(\beta) &= \frac{\rho^k e^{r(t_1-t_2)} \left[\frac{\beta + \rho e^{rt_2} - a}{\beta + \rho e^{rt_1} - a} \right]^2 \lambda^{k-2}}{\frac{\rho e^{rt_1}(1-a)}{\beta + \rho e^{rt_1} - a}} \times \\
&\qquad \prod_{i=2}^{k} e^{r(t_i - s_{i,1})} \left[\frac{\beta + \rho e^{rs_{i,1}} - a}{\beta + \rho e^{rt_i} - a} \right]^2 e^{r(t_i - s_{i,2})} \left[\frac{\beta + \rho e^{rs_{i,2}} - a}{\beta + \rho e^{rt_i} - a} \right]^2 \\
&= \frac{\rho^{k-1} e^{-rt_2} \left[\beta + \rho e^{rt_2} - a \right]^2 \lambda^{k-2}}{(1 - a) \left[\beta + \rho e^{rt_1} - a \right]} \times \\
&\qquad \prod_{i=2}^{k} \frac{e^{r[2t_i - (s_{i,1} + s_{i,2})]} \left[(\beta + \rho e^{rs_{i,1}} - a)(\beta + \rho e^{rs_{i,2}} - a) \right]^2}{\left[\beta + \rho e^{rt_i} - a \right]^4}.
\end{aligned} \tag{34}
$$

Thus, after differentiating the function $\Theta(\beta)$ $(M - j)$-th times with respect to β and substituting $\beta = 0$, we obtain the exact joint likelihood function. It is very tideous doing this differentiating analytically and thus we resort to numerical approximations. We can thus write the analytical form of the joint likelihood as

$$\mathcal{L}(t, M+k|\lambda, \mu, k) = \frac{k!}{\alpha^M (M+k)!} \times$$

$$\sum_{j=0}^{M} \binom{M}{j} \frac{(k+j)!}{k!} \alpha^j \frac{\rho^{k-1}\lambda^{k-2}}{(1-a)} \prod_{i=2}^{k} e^{r[2t_i - (s_{i,1}+s_{i,2})]} \times \frac{\partial^{M-j}}{\partial \beta^{M-j}}$$

$$\left(\frac{(\beta + \rho e^{rt_2} - a)^2}{\beta + \rho e^{rt_1} - a} \prod_{i=2}^{k} \frac{[(\beta + \rho e^{rs_{i,1}} - a)(\beta + \rho e^{rs_{i,2}} - a)]^2}{[\beta + \rho e^{rt_i} - a]^4} \right) (0)$$

$$\times \frac{\binom{M+k}{k} \beta^M \alpha^M}{\sum_{N=0}^{\infty} \binom{N+k}{k} \beta^N \alpha^N}.$$

$$(35)$$

References

[1] W. Feller, *An introduction to probability theory and its applications.* volume 1, John Wiley & Sons Inc., New York, third edition, 1968

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

[3] Susana Magallón, Michael J. Sanderson, Absolute diversification rates in angiosperm clades, *Mol. Biol. Evol.*, 55(9):1762–1780, 2001

[4] Hélène Morlon, Todd L. Parsons, Joshua B. Plotkin, Reconcilling molecular phylogenies with the fossil record, *Proc. Natl. Acad. Sci.* USA, 108(39):16327–16332, 2011

[5] Morlon H, Potts MD, Plotkin JB (2010) Inferring the Dynamics of Diversification: A Coalescent Approach, *PLoS Biol*, 8(9): e1000493. doi:10.1371/journal.pbio.1000493

[6] Daniel L. Rabosky, Stephen C. Donnellan, Amanda L. Talaba, Irby J. Lovette, Exeptional among-lineage variation in diversification rates during the radiations of Australia's most diverse vertebrate clade, *Proc. R. Soc. B*, 274(6):2915–2923, 2007

[7] David M. Raup, Mathematical models of cladogenesis, *Paleobiology.*, 11:42–52, 1985

[8] V. K. Rohatgi, *An introduction to probability theory and mathematical statistics.* Wiley, New York, 1976

[9] Nee S., Robert M. May, Paul H. Harvey The reconstructed evolutionary process, *Phil. Trans. R. Soc. Lond. B*, 344:305–311, 1994

[10] Tanja Stadler, On complete sampling under birth-death models and connections to the sampling-based coalescent, *Journal of Theoritical Biology*, 261(1):58–66, 2009