



Inferring the Timescale of Phylogenetic Trees from Fossil Data

Walker Pett, Tracy A. Heath

► To cite this version:

Walker Pett, Tracy A. Heath. Inferring the Timescale of Phylogenetic Trees from Fossil Data. Scornavacca, Celine; Delsuc, Frédéric; Galtier, Nicolas. Phylogenetics in the Genomic Era, No commercial publisher | Authors open access book, pp.5.1:1–5.1:18, 2020. hal-02536361

HAL Id: hal-02536361

<https://hal.science/hal-02536361>

Submitted on 10 Apr 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.




Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

Chapter 5.1 Inferring the Timescale of Phylogenetic Trees from Fossil Data


Walker Pett

Department of Ecology, Evolution, and Organismal Biology
Iowa State University, Ames, Iowa, 50011 USA
willpett@iastate.edu

 <https://orcid.org/0000-0003-3733-0815>

Tracy A. Heath

Department of Ecology, Evolution, and Organismal Biology
Iowa State University, Ames, Iowa, 50011 USA
phylo@iastate.edu

 <https://orcid.org/0000-0002-0087-2541>

Abstract

Time-stamped historical observations are required for scaling phylogenetic estimates to absolute time and, as a consequence, genomic data alone are not sufficient for dating the tree of life. The fossil record is the primary source of dated evidence of lineages over time and several statistical models for integrating paleontological and neontological data have been introduced. This chapter provides an overview of how fossil data are recovered from the rock record. We then describe two approaches to dating phylogenetic trees: (1) node dating where fossils are treated as calibrations for speciation times in an extant phylogeny and (2) the fossilized birth-death process as a mechanistic model that accounts for lineage diversification and fossil sampling. We conclude by discussing promising extensions of diversification models that can account for the structure of the fossil record and enable a more complete treatment of extinct and modern taxa in macroevolutionary analyses.

How to cite: Walker Pett and Tracy A. Heath (2020). Inferring the Timescale of Phylogenetic Trees from Fossil Data. In Scornavacca, C., Delsuc, F., and Galtier, N., editors, *Phylogenetics in the Genomic Era*, chapter No. 5.1, pp. 5.1:1–5.1:18. No commercial publisher | Authors open access book. The book is freely available at <https://hal.inria.fr/PGE>.

Funding This work was supported by the National Science Foundation (USA) grants DEB-1556615, DEB-1556853, and DBI-1759909.

1 Introduction

Reconstructing the timescale of the tree of life is fundamental to understanding the pattern and process of species diversification. Inferring the topology of evolutionary relationships among species has been greatly facilitated by the advancements of the genomic era. These innovations include the accumulation of vast quantities of genomic character data, the development of high-dimensional statistical models of molecular evolution (Chapters 1.1 and 1.4 [Pupko and Mayrose 2020; Lartillot 2020a]), and increasingly robust computational tools (e.g. Chapters 1.3 and 1.5 [Kozlov and Stamatakis 2020; Lartillot 2020b]). However, estimating the timing of species divergences remains notoriously difficult, in part because good estimates can usually only be obtained by considering multiple sources of information simultaneously.



© Walker Pett and Tracy A. Heath.

Licensed under Creative Commons License CC-BY-NC-ND 4.0.

Phylogenetics in the genomic era.

Editors: Celine Scornavacca, Frédéric Delsuc and Nicolas Galtier; chapter No. 5.1; pp. 5.1:1–5.1:18

A book completely handled by researchers.



No publisher has been paid.

5.1:2 Inferring the Timescale of Phylogenetic Trees from Fossil Data

This is because phylogenetic character data—for example from molecular sequences or species morphology—can only tell us about the total amount of evolutionary change, or distance, among lineages. Evolutionary distances are traced out at a certain rate over a certain amount of time, and consequently information about the timing of species divergence is inherently confounded with information about the rate of evolutionary change (Chapter 3.2 [Boussau and Scornavacca 2020]). For this reason, independent estimates of rates and times cannot be obtained from phylogenetic character data alone. Thus, while the increasing sizes of molecular datasets may provide precise distance estimates, other sources of information about rates and times are ultimately required to improve the precision of divergence time estimates (dos Reis and Yang, 2013).

The fossil record provides the richest source of information about the absolute timing of lineage diversification in the tree of life. A fossil specimen is direct evidence for the existence of an ancient lineage and allows us to observe its age and preserved characteristics. These observations can then directly inform the phylogenetic placement of fossil specimens as well as the timing of their divergence from related species. By combining paleontological (fossil) and neontological (extant) data in a joint phylogenetic analysis, we can obtain independent estimates of molecular evolutionary rates and divergence times.

In order to include timing information from the fossil record in a phylogenetic analysis of molecular sequence data, a model is needed to describe the process of collecting and observing fossil data through time (Section 2). This model can range in complexity from a phenomenological description of the data (Section 3), to a richer mechanistic model of species diversification and fossil recovery (Section 4). By integrating these models of fossil occurrence time data with other types of data in a hierarchical Bayesian framework, the task of statistically disentangling rate and time is greatly simplified (Section 5).

2 Formalizing our Knowledge of the Fossil Record

Our understanding of the history of life on Earth begins with the geology of Earth's crust. Organisms bury evidence of their existence in the soils and sediments of their local environment, and over time these remains are compacted and preserved in rock layers. Eventually, geological or meteorological activity can expose these rock layers at the surface, where they can be observed and their ages determined, using radiometry or stratigraphic methods. When a fossil is discovered, its age can often only be determined as falling somewhere within the minimum and maximum extent of the rock layer containing it. If a fossil species spans multiple layers, its stratigraphic age range is determined from the ages of those layers. We will summarize the information contained in these fossil age observations using the symbol \mathcal{F} . Other observations not related to a fossil's age can also offer a wealth of phylogenetically informative characters, such as data from morphological and life history traits, biogeography, or even ancient DNA sequences (which have been recovered from specimens as old as 700,000 years; Orlando et al., 2013). These latter character observations are collectively denoted by the symbol \mathcal{D} , which may also include observations from extant species.

As mentioned already, evolutionary rates and times are inherently confounded as evolutionary distances. Inferences about divergence times from character data \mathcal{D} alone will therefore be informed entirely by our *a priori* assumptions about the joint process of character evolution and species diversification. To get independent estimates of time, we must summarize our knowledge of the process of fossil preservation and collection generating not only the observed character data \mathcal{D} but also the age data \mathcal{F} . Specifically, it will be convenient to formalize

our assumptions using a probabilistic mathematical model. In Bayesian parlance, the model is specified quantitatively as a joint *prior distribution* over the evolutionary parameters of interest, summarizing our knowledge of the model parameters prior to any data collection. Then, after collecting observations from the fossil record, we use Bayesian statistics to update our knowledge by estimating the *posterior distribution* of the model parameters.

2.1 Prior distribution on divergence times

We begin by specifying a prior distribution over our model parameters. In the context of divergence time estimation, our model can be thought of as broadly consisting of two sets of parameters. One set, labeled \mathcal{T} , includes those related to the diversification and species sampling process, such as the tree topology, divergence times, fossil sampling rates, etc. The other set, which we'll call θ , includes those related to the process of character evolution, such as rates of morphological evolution, or molecular substitution rates. We specify our prior distribution as the product of independent densities $f(\mathcal{T})$ and $f(\theta)$, such that

$$f(\mathcal{T}, \theta) = f(\mathcal{T})f(\theta). \quad (1)$$

The density $f(\mathcal{T})$ can be defined using a stochastic branching process like the Yule (Yule, 1924) or birth-death processes (Kendall, 1948). The density $f(\theta)$ can be defined in various ways (e.g. using a relaxed or other clock model, see Chapter 4.4 [Bromham 2020]) to describe the processes generating observed characters.

2.2 Posterior distribution on divergence times

To construct the posterior distribution over \mathcal{T} and θ , we collect observations in the form of character data \mathcal{D} and timing data from the fossil record \mathcal{F} . Then, using Bayes' theorem, the posterior distribution is proportional to the product of the prior distribution over \mathcal{T}, θ and the likelihood of \mathcal{D} and \mathcal{F}

$$f(\mathcal{T}, \theta \mid \mathcal{D}, \mathcal{F}) \propto f(\mathcal{D} \mid \mathcal{T}, \theta)f(\mathcal{F} \mid \mathcal{T})f(\mathcal{T})f(\theta), \quad (2)$$

where the term $f(\mathcal{D} \mid \mathcal{T}, \theta)$ is the likelihood of the observed character data and the term $f(\mathcal{F} \mid \mathcal{T})$ is the likelihood of the observed fossil age data.

Importantly, we assume that the likelihood of the character data depends on \mathcal{T} and θ , while the likelihood of the fossil age data \mathcal{F} depends only on \mathcal{T} . The development of methods for calibrating trees to absolute time scales is primarily concerned with definitions of $f(\mathcal{T})$ and $f(\mathcal{F} \mid \mathcal{T})$. In the following sections, we will discuss two main approaches for defining these densities.

3 Node Calibration Densities

Historically, the most common statistical approach to inferring divergence times on a phylogenetic tree has been through the use of node calibrations. In this approach, information from the fossil record about the age of a particular clade is used to directly constrain the age of a node in the tree (typically the most-recent-common ancestor of the clade) during a phylogenetic analysis. This can be framed in a probabilistic approach by associating each calibrated node age with a probability density function, or node calibration density. In Bayesian inference, these densities are then used to compute the posterior distribution over \mathcal{T} , the tree topology and node ages. The way in which these node calibration densities are specified and applied has been the subject of a wide array of empirical and methodological studies (e.g., Ho and Phillips, 2009; Warnock et al., 2012).

3.1 Conceptual formulation of node calibrations

Before exploring the ways in which node calibration densities can be applied in a phylogenetic analysis, we must first ask: *What, exactly, do node calibration densities represent?* There is considerable discussion of the representational and conceptual meaning of node calibration densities, but there are essentially two basic interpretations.

One interpretation formulates node calibrations directly as prior densities on the node ages (Yang and Rannala, 2005; Heled and Drummond, 2012). That is, the prior density $f(\mathcal{T})$ is constructed in such a way as to simultaneously account for uncertainty in the ages of both calibrated and uncalibrated nodes. Methods using this interpretation derive a conditional density on the uncalibrated node ages with fixed ages for the calibrated nodes, and then define the marginal prior on the calibrated node ages using a calibration density. It has been shown that for multiple node calibrations, this type of conditional prior leads to counterintuitive topologically inconsistent realized priors (Rannala, 2016), and may be computationally intractable (dos Reis, 2016).

For this reason, it is perhaps conceptually simpler to take the second point of view, which interprets node calibrations instead as representing the likelihood of the fossil data $f(\mathcal{F} | \mathcal{T})$. In this interpretation, the fossil likelihood is typically formulated as a product of marginal densities for each node in the tree for which there is relevant fossil data. In other words, for each calibrated node i with divergence time t_i , it is assumed that the likelihood of the fossil observations for that node is proportional to some density g_i . The full fossil likelihood is then a product of these node densities

$$f(\mathcal{F} | \mathcal{T}) \propto \prod_i g_i(t_i | \alpha),$$

where the parameterization α and functional form of the density g_i is specified by the researcher. In this approach, the prior density on the node ages $f(\mathcal{T})$ can then be specified simply using a familiar uncalibrated tree-generating process, such as the Yule or birth-death process.

Despite the relatively straightforward distinction between the two approaches, there is still considerable misunderstanding surrounding the interpretation of node calibration densities, and the two are commonly confused. In particular, a common misconception arises when the fossil data likelihood for a particular node is interpreted as a prior density on the age of that node. This interpretation as a prior is sometimes described as “incoherent” because it leads to the specification of two independent prior densities on the node ages, one from the node calibration g_i , and one from the tree prior $f(\mathcal{T})$ (Heled and Drummond, 2012; Heath et al., 2014).

Fortunately, the consequences of this misinterpretation are purely conceptual and should not have any quantitative impact on the resulting inferences. As an example, consider the distribution on the waiting time until recovery of the first fossil sample after the divergence of a clade. Assuming the waiting time is exponentially distributed, the density is the same whether we define the distribution with respect to the clade divergence time or the fossil age (Figure 1). Thus, whether we think of the density as likelihood or prior, both will lead to identical posterior distributions. Nevertheless, it is useful to follow the more principled likelihood interpretation as this will lead to more coherent and consistent application of node calibration methods generally.

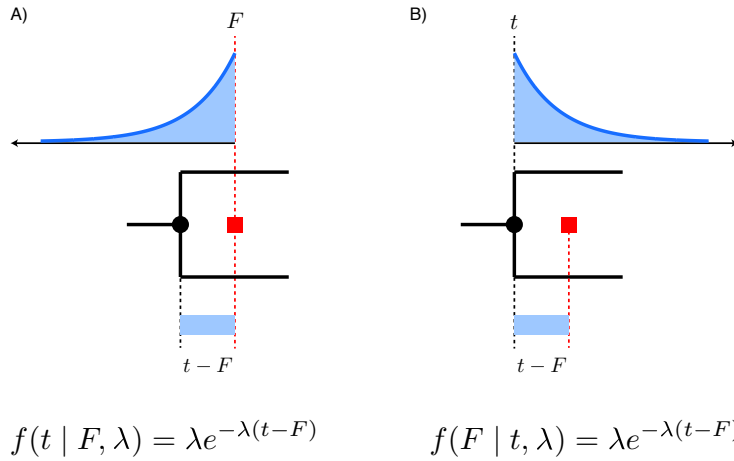


Figure 1 Alternative interpretations of a node calibration density as an (A) incoherent prior or a (B) likelihood. Given the age F of the first fossil specimen recovered after cladogenesis at time t , the waiting time $t - F$ is assumed to be exponentially distributed with rate λ . Both densities yield the same posterior distribution. A) The density is interpreted as a prior on the divergence time t . This results in the incoherent specification of two independent prior densities on t , coming from the node calibration as well as the tree prior. B) The density is interpreted as the likelihood fossil recovery age F .

3.2 Node calibrations with qualitative fossil data

In the simplest case, the form (*e.g.*, log-normal, exponential) and parameterization (*e.g.*, the mean, variance, and upper or lower bounds) of a node calibration density is chosen to qualitatively reflect the researcher's belief about the age of a clade based on their interpretation of the fossil data (Yang and Yoder, 2003; Yang and Rannala, 2005). In other words, the process of fossil preservation and observation is treated qualitatively, whereby the researcher's interpretation of these phenomena is implicit in the shape of $f(\mathcal{F} \mid \mathcal{T})$.

A variety of mathematical distributions have been proposed to represent these subjective calibrations with implicit fossil data, including normal, lognormal, exponential and uniform distributions (see Hedges and Kumar, 2004; Drummond et al., 2006; Donoghue and Benton, 2007; Ho, 2007; Ho and Phillips, 2009). These can be chosen such that the upper and/or lower bounds of the calibration density are “hard” or “soft” indicating whether there is assumed to be a non-zero probability of a fossil occurring outside the calibration bounds (Yang and Rannala, 2005; Sanders and Lee, 2007; Inoue et al., 2009). Soft bounds can be implemented for example by assuming the variance of the calibration density is such that 5% of the density falls beyond the maximum age constraint. While soft minimum bounds typically represent uncertainty in the age of the youngest calibration fossil (Benton and Donoghue, 2007), soft maximum bounds are typically justified either on the basis of models of diversification and preservation probability below the oldest known fossil in a clade (Foote et al., 1999; Tavaré et al., 2002), or using phylogenetic bracketing (Reisz and Müller, 2004; Müller and Reisz, 2005). Divergence time estimates can be extremely sensitive to the parameterization of the calibration density, but the impact on divergence time estimates of different prior densities is minimized when both minimum and maximum constraints are used (Warnock et al., 2012).

3.3 Node calibrations with quantitative fossil data

Other node calibration methods have been developed to make the representation of fossil data more quantitative and reproducible. For example, drawing on paleontological methods for estimating the stratigraphic ranges of fossil species (Strauss and Sadler, 1989), some node calibration methods make the explicit assumption that fossil recovery for a particular clade follows a constant-rate Poisson process through time, which implies that the ages of fossil specimens will be uniformly distributed over the clade's lifespan (Marshall, 2008; Dornburg et al., 2011; Wilkinson et al., 2011; Claramunt and Cracraft, 2015). Then, from the order statistics of a uniform distribution, it can be shown that the likelihood for the age of the oldest fossil F in a clade is equal to $f(F | t, n) = \frac{1}{t^n} n F^{n-1}$ (Strauss and Sadler, 1989), where t is the age of the clade and n is the number of fossil specimens. The calibration density $g(t | F, n)$ is proportional to the likelihood, and thus depends only on the number of fossil specimens

$$g(t | F, n) \propto f(F | t, n) \propto \frac{1}{t^n}, \quad t > F.$$

For example, Claramunt and Cracraft (2015) used this approach to calibrate the origin of modern birds.

Other methods construct node calibrations by modeling the process of fossil preservation as an exponential waiting time between clade divergence and fossil deposition (Wilkinson et al., 2011; Heath, 2012). Together, these methods take a step toward better formalizing the process of fossil data collection and interpretation, which ultimately makes their conclusions more testable and extensible.

Despite efforts to formalize the interpretation and characterization of node calibration densities, this approach to dating phylogenies still suffers from some limitations. Most notably, because calibration densities are only informed by the oldest node descended from a given calibrated node, these methods ignore much of the information present in the fossil record. Furthermore, fossil sampling times are observations of the underlying diversification process that gave rise to the phylogeny uniting the fossils and their extant relatives. In a statistical inference framework, these data can inform the parameters of the diversification model (*i.e.*, speciation and extinction), leading to more accurate and precise estimates.

3.4 Secondary calibrations

The posterior distribution summarizes our knowledge of the model parameters after taking some observations into account. If subsequent observations are made, the posterior can continue to be updated by considering it as a prior in relation to new data. This behavior can be leveraged to use posterior divergence time estimates from past studies as node calibrations in new analyses.

Specifically, consider the marginal posterior distribution $f(\mathcal{T} | \mathcal{F})$ obtained from a study using fossil data \mathcal{F} . This posterior is proportional to the fossil likelihood for \mathcal{F} and the prior over \mathcal{T}

$$f(\mathcal{T} | \mathcal{F}) \propto f(\mathcal{F} | \mathcal{T})f(\mathcal{T}). \quad (3)$$

Now imagine that new fossil observations \mathcal{F}' are collected. We compute the joint posterior conditioned on both \mathcal{F} and \mathcal{F}' as

$$f(\mathcal{T} | \mathcal{F}, \mathcal{F}') \propto f(\mathcal{F}, \mathcal{F}' | \mathcal{T})f(\mathcal{T}),$$

where $f(\mathcal{F}, \mathcal{F}' | \mathcal{T})$ is the joint likelihood of both sets of observations. If we assume that \mathcal{F} and \mathcal{F}' are sampled independently of each other, we may factor the likelihood such that

$$\begin{aligned} f(\mathcal{T} | \mathcal{F}, \mathcal{F}') &\propto f(\mathcal{F}' | \mathcal{T})f(\mathcal{F} | \mathcal{T})f(\mathcal{T}) \\ &\propto f(\mathcal{F}' | \mathcal{T})f(\mathcal{T} | \mathcal{F}), \end{aligned}$$

where we have made use of Equation 3 to substitute the likelihood and prior terms involving \mathcal{F} for the previously obtained marginal posterior $f(\mathcal{T} | \mathcal{F})$. The updated posterior can be estimated in a straightforward manner using MCMC and a previously obtained sample from $f(\mathcal{T} | \mathcal{F})$. In other words, we simply use the previously obtained posterior distribution as our new prior distribution on \mathcal{T} .

Importantly, in this approach we must consider the secondary node calibrations as part of the prior distribution over \mathcal{T} , and not as a fossil data likelihood. If we treated it as a likelihood and reweighed it according to an unconditioned prior, this would lead to the incoherent specification of two prior distribution terms for \mathcal{T} : one through the secondary calibration and one through our own specification of the prior. In order to avoid such incoherence in the prior, we must therefore either (1) specify the prior distribution as consisting entirely of the previously obtained sample, or (2) use an approach like those described at the beginning of Section 3.1 to condition the prior on the previous estimate (Yang and Rannala, 2005; Heled and Drummond, 2012). Despite this conceptual limitation of secondary calibrations, they are almost exclusively misapplied as fossil data likelihood terms, resulting in overly precise divergence time estimates (for review, see Schenk, 2016).

4 The Fossilized Birth-Death Process

As a result of the limitations of node calibration densities, it may be more satisfying to take an explicit, mechanistic approach by specifying a joint process $f(F, \mathcal{T})$ that simultaneously describes the generation of both the tree and the fossil data. Such an approach allows us to assume that fossil specimens are observations of lineages generated by the same diversification process that gave rise to the sampled living taxa. This allows for the inference of biologically meaningful parameters governing the diversification of both extant and extinct taxa—such as rates of speciation, extinction, and sampling—and leads to a more precise, quantitative representation of the process by which fossil specimens are sampled along lineages. Such a model was first described by Stadler (2010), who extended the birth-death process (Kendall, 1948; Nee et al., 1994; Gernhard, 2008; Stadler, 2009; Thompson, 1975) to account for lineages sampled back in time (see also Didier et al., 2012). By integrating fossil occurrence times into the branching model, this serially sampled birth-death process allows for estimation of macroevolutionary parameters under complex, mechanistic models of lineage diversification and fossil sampling.

4.1 Models for serially sampled data

Stadler (2010) introduced a serially sampled birth-death process that is well-suited to applications in macroevolution and in the study of infectious diseases (see Chapter 5.3 [Zhukova et al. 2020]). When applied to macroevolutionary analyses of species-level data, this process requires samples from the fossil record and was thus coined the *fossilized birth-death process* (FBD) in Heath et al. (2014). Cladogenesis under a birth-death model begins with a single lineage that starts at time ϕ , this is the origin time of the process. Over the course of diversification, lineages speciate at rate λ and go extinct at rate μ . In the present (*i.e.*, $t = 0$),

5.1:8 Inferring the Timescale of Phylogenetic Trees from Fossil Data

every living lineage has a probability ρ of being sampled. In the absence of sampled fossils, all birth-death processes are governed by parameters λ , μ , and ρ . The FBD model includes a parameter for the rate of fossil recovery, denoted ψ , to account for observations from the fossil record. This parameter acts as a Poisson rate of sampling lineages over time.

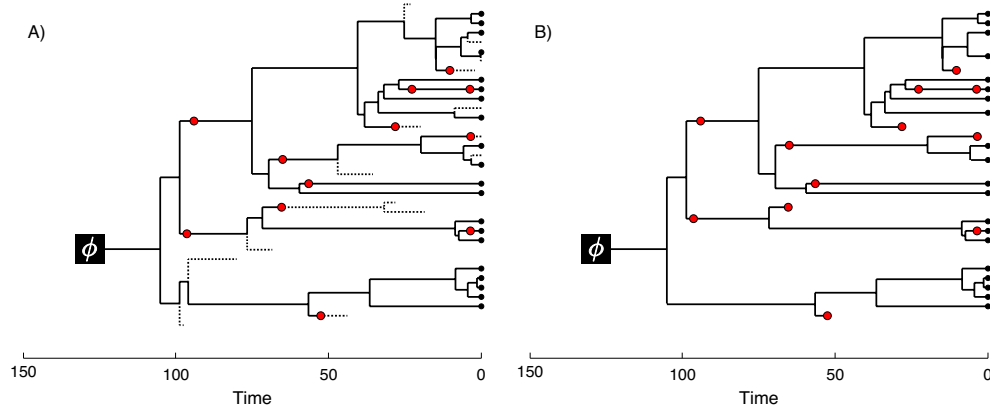


Figure 2 The fossilized birth-death process generates a (A) complete tree and a (B) sampled tree. (Figure modified from Figure 1 of Stadler, 2010)

The FBD model generates a *complete tree* and set of fossils Figure 2A. The *sampled tree* (also called the reconstructed tree) is the phylogeny after sampling, with all unobserved lineages pruned away Figure 2B. The probability density defined in Stadler (2010) allows us to compute the probability of any sampled tree while accounting for unobserved lineages in the complete tree. Figure 3 depicts the FBD process using a graphical model (for more on probabilistic graphical models for phylogenetics, see Höhna et al., 2014). This figure illustrates that the probability density of the sampled tree \mathcal{T} —which includes the tree topology, divergence times, and observed fossil occurrences \mathcal{F} —is dependent on the origin time ϕ , speciation rate λ , extinction rate μ , fossilization rate ψ , and the extant species sampling probability ρ .

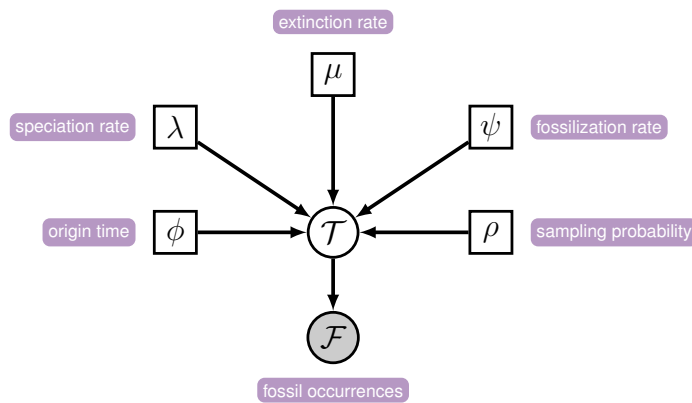


Figure 3 A graphical model depicting the structure of the fossilized birth-death process. The probability of the tree topology and divergence times \mathcal{T} depend on the parameters of the FBD process: the time of origin ϕ , speciation rate λ , extinction rate μ , the rate of fossil recovery ψ , and the probability of sampling ρ . The observed fossil occurrences \mathcal{F} are, in turn, dependent on \mathcal{T} and the upstream parameters.

4.2 Sampled ancestors and the taxonomic assignment of fossil specimens

Under the FBD process, each fossil specimen is assumed to represent an independent sample from a continuously evolving lineage. Thus, there is a non-zero probability of obtaining a fossil sample that also has sampled descendant lineages Figure 2B. Indeed, Foote (1996) estimated that the probability of sampled ancestor-descendant pairs in the fossil record is non-negligible under a variety of cladogenetic models. Thus, it is important for diversification models to correctly account for sampled ancestors in order to accurately estimate speciation and extinction rates.

Under the FBD model, the proportion of fossil samples that also have sampled descendants is correlated with the probability of sampling extant lineages (ρ), the fossil sampling rate (ψ), and turnover ($r = \frac{\mu}{\lambda}$). We demonstrate this using simulations under the FBD model in Figure 4 for four different values of turnover and two different values of ρ , all over a range of values for ψ . There is a clear interplay between the parameters of the FBD model, which interact to yield different samples. Notably, even when extinction is relatively high ($r = 0.9$), extant sampling is low ($\rho = 0.1$), and the fossil recovery rate is low ($\psi = 0.01$), there is still a substantial proportion of fossils that also have sampled descendants (which may be fossil or extant samples), as seen in Figure 4.

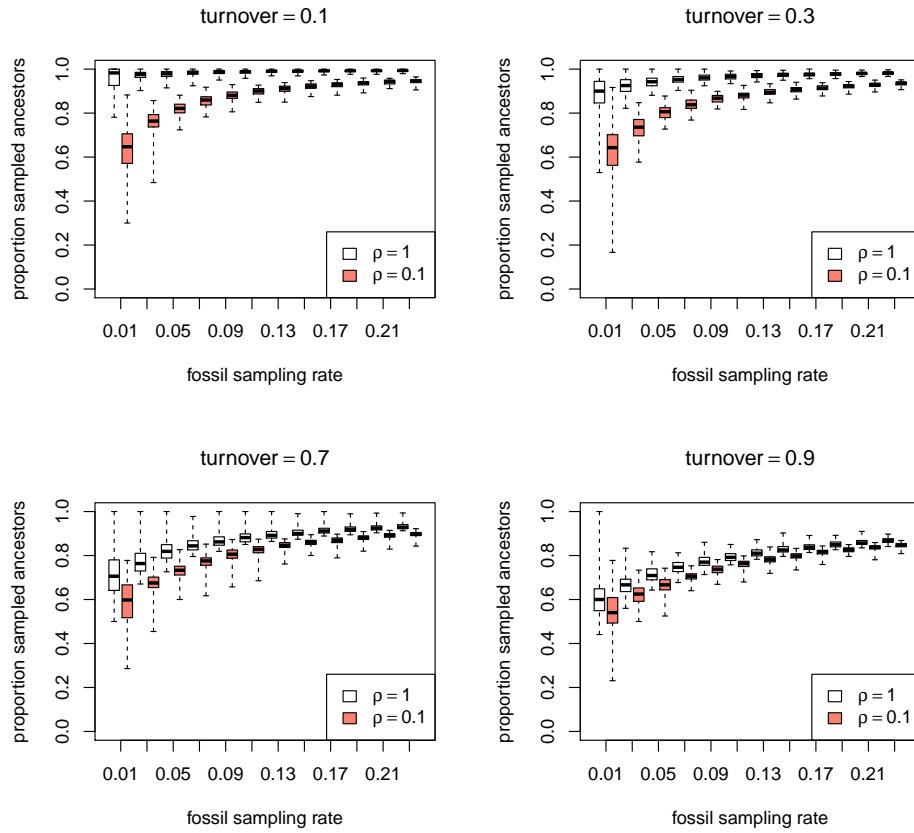
Sampled ancestors present a curious taxonomic challenge. In the most straightforward formulation, no assumptions are made regarding the taxonomic assignment of sampled specimens, so all speciation events are assumed to correspond with branching events, and the speciation rate is therefore equal to the rate of cladogenesis. However, if two fossil samples are taken along a single evolving lineage, but are then assigned to different taxonomic species, then this implies an intervening anagenetic speciation event. Real speciation rates are therefore larger than the rate of cladogenesis, since some taxonomic species arise through anagenesis. In addition, the fossil record contains far more specimens than described taxa, and so many fossils are identified as belonging to the same taxonomic species. Consequently, in order to apply the FBD model to real fossil data, some assumptions must be made about the process of assigning fossil specimens to taxonomic species. Stadler et al. (2018) described a model for assigning fossil specimens to the same taxonomic species stratigraphic range.

4.3 Fossil placement under the fossilized birth-death process

Much like with node calibration approaches, divergence time estimates under the FBD may be sensitive to the phylogenetic placement of fossils. Unlike node calibrations, however, under the FBD model, all of the fossils—not just the oldest—that can be assigned to a node are valid observations. Nevertheless, it is important to consider the best practices outlined by Parham et al. (2012) when choosing fossils and justifying their placement in the phylogeny.

In many cases, quantitative character data have not been coded or are otherwise not available for a particular fossil taxon. In these instances, qualitative information about the topological placement of a fossil may be derived from its observed occurrence time and/or the taxonomic literature (Heath et al., 2014). For many fossils, it may be ambiguous whether the fossil lineage falls on the stem lineage or within the crown of an extant clade (Benton and Donoghue, 2007). In many cases, it is only possible to define the fossil and its relatives as a monophyletic “total group”, within which there are many possible placements of the fossil as a crown or stem fossil. Bayesian inference methods using the FBD process can account for this uncertainty by integrating over the different possible fossil placements using Markov chain Monte Carlo (MCMC).

5.1:10 Inferring the Timescale of Phylogenetic Trees from Fossil Data



■ **Figure 4** The proportion of sampled ancestors simulated under different FBD parameters. Here, *turnover* (r) is defined as $r = \frac{\mu}{\lambda}$, where λ and μ denote *speciation* and *extinction*, respectively. For each of the four values for turnover, we show the proportion of sampled ancestors for 100 simulated replicates as we varied the fossil sampling rate (ψ) and for two different values of the probability of sampling extant taxa (ρ).

When character data are available for fossil taxa, an integrative modeling approach is needed to combine observations from both extant and extinct species. The model and methods described in Stadler (2010), Ronquist et al. (2012), Zhang et al. (2016), and others provide a framework for using the FBD model in more fully integrative Bayesian analysis of fossil and extant samples (see section 5).

4.3.1 Empirical studies applying the fossilized birth-death model

Analysis under the FBD process enables researchers to use more of the data from the fossil record, which, in turn, can lead to more robust estimates and a more comprehensive understanding of lineage diversification. Using simulated trees and data, Heath et al. (2014) demonstrated that when using fossil occurrences to date extant phylogenies under the FBD model, node age estimates are more accurate than conventional calibration density approaches. Importantly, this study also showed that the precision of FBD node age estimates increases as the number of fossil occurrences increases, providing a better representation of statistical uncertainty in these parameters. Didier et al. (2017) also developed a maximum likelihood

approach to estimate parameters of the diversification model when fossil occurrences are observed. Their analyses of simulated datasets demonstrate that estimates of speciation and extinction rates are more accurate when fossil ages are included compared to estimates based on trees of extant taxa with the node ages fixed to their true values.

The FBD model has gained traction in both neontological and paleontological studies because its assumptions are more justifiable than node-calibration density approaches (as described in Section 3). As a result, empirical studies are emerging that provide new insights into the macroevolution of numerous clades in the tree of life. For example, the FBD process has been used for the calibration of extant phylogenies of royal ferns (Grimm et al., 2014), tetraodontiform fishes (Arcila et al., 2015), and pines (Saladin et al., 2017); in combined evidence analyses of hymenopterans (Zhang et al., 2016), lemurs (Herrera and Dávalos, 2016), myriapods (Fernández et al., 2016), sloths (Slater et al., 2016), penguins (Gavryushkina et al., 2017), baleen whales (Slater et al., 2017), and sponges (Schuster et al., 2018); or to study extinct clades using morphological characters and occurrence times of theropods (Bapst et al., 2016), and crinoids (Wright et al., 2017).

5 Integrative Hierarchical Models for Calibrating Time Trees

One important advantage of the fossilized birth-death modeling approach is that it makes our evolutionary analysis more integrative. By directly modeling fossil sampling jointly with cladogenesis and extinction, we can connect information from disparate evolutionary processes and synthesize all of it in a single *hierarchical* analysis. In a hierarchical model, relationships among collections of model parameters are structured in a directional, tree-like manner, such that information from a number of empirical observations of different datatypes can be considered jointly through their shared dependence on a smaller number of upstream model parameters. For example, the FBD model allows us to unify information from both the fossil record and the molecular record, by connecting models of fossil sampling and molecular evolution indirectly through a time tree model of speciation and extinction (as shown in the hierarchical model in Figure 5). In other words, the FBD model provides the foundation on which to construct much larger and more elaborate probabilistic models that link a wide range of information sources in a Bayesian hierarchical inference framework.

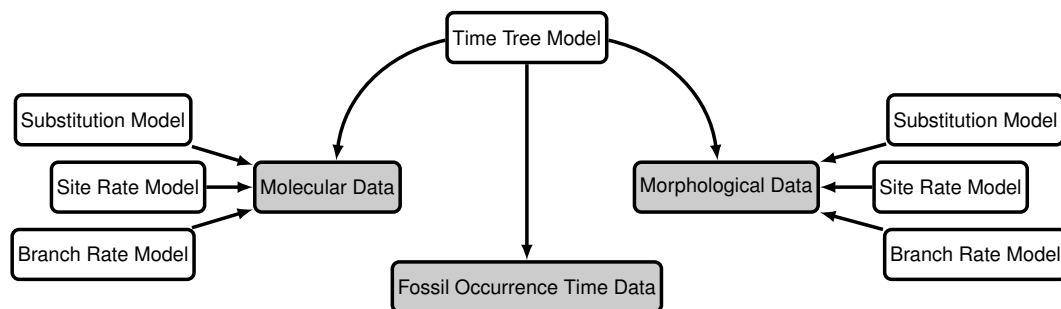


Figure 5 An example of a hierarchical model using molecular, morphological and fossil occurrence time data. Nodes represent collections of random variables and model parameters. Shaded nodes represent variables that are associated with empirical observations.

5.1 Integrating molecules and morphology

One immediate consequence of including fossil and extant taxa in a joint hierarchical analysis is that character data from both extinct and extant species can be combined. In particular, morphological data from fossil specimens can be directly combined with data from extant species. These *combined-evidence* (also called “total-evidence” or “tip-dating”) approaches allow for inference of fossil species relationships with extant taxa (Ronquist et al., 2012). Thus, this integrative statistical approach leads to more reproducible placement of fossil species using explicit methodology, instead of reliance on previously published studies or taxonomy. The placement of calibration fossils using morphological data also results in more robust estimates of divergence times and diversification dynamics among extant species (Heath et al., 2014; Gavryushkina et al., 2014). Furthermore, any uncertainty in the placement of fossils can be accommodated by implementing this approach in a Bayesian framework, resulting in parameter estimates that better reflect the information in the data.

The first combined-evidence analyses described in the literature either did not use an appropriate diversification model (Pyron, 2011, used a pure-birth model that did not allow extinction or fossil sampling) or used a generic, phenomenological model (Ronquist et al., 2012, assumed a uniform model for node ages, fossil sampling times, and topologies). Nevertheless, these studies outlined a framework for a more fully integrative approach to inferring dated phylogenies of living and fossil taxa. With the introduction of new MCMC proposal mechanisms for FBD trees (Heath et al., 2014; Gavryushkina et al., 2014; Zhang et al., 2016), combined-evidence analyses can now include much more appropriate and realistic models of lineage diversification and sampling. Using a combined-evidence approach, the fossilized birth-death process has been applied to combined datasets from Hymenoptera (Zhang et al., 2016), penguins (Gavryushkina et al., 2017), mammals (Upham et al., 2019) and squamates (Pyron, 2016), to name a few. Luo et al. (2020) used simulations to explore the accuracy of combined-evidence dating approaches under the fossilized birth death process, and found that including morphological data led to more accurate estimation of divergence times, but that most dating information was contained in fossil occurrence age data.

5.2 Sampling bias

As our integrative modeling approach grows in complexity, and encompasses more and different types of data, it becomes extremely important to carefully consider the sampling strategies used to acquire those data, and ask whether any sampling biases may be impacting the analysis. If so, we must look for ways of explicitly accounting for the data sampling strategy when specifying the model. Here we discuss several important sources of sampling biases as they pertain to divergence time estimation and the fossilized birth-death process.

5.2.1 Taxonomic - diversified sampling

One important type of sampling bias may arise as a result of a commonly employed diversified taxon sampling strategy. In this approach, only representatives from major lineages are included in an analysis in an attempt to achieve broad taxonomic representation while at the same time minimizing the size of the dataset and redundancy in the analysis (*e.g.*, Jarvis et al., 2014, sampled a single representative from every avian order to estimate the relationships and divergence times of birds). This type of sampling will induce longer terminal branches than under a random sampling scheme (Heath, 2008). Höhna et al. (2011) described an approach using a birth-death process that accounts for diversified sampling by assuming that lineages are sampled such that the total phylogenetic diversity (tree length) is maximized.

Analytic expressions are available for the probability density of a birth-death tree under this type of diversified sampling (Höhna et al., 2011). Zhang et al. (2016) applied this approach to the FBD model by assuming that exactly one representative extant species per clade descending from some cutoff time is selected. Accounting for diversified sampling seems to be important in some empirical datasets, resulting in much younger age estimates for some clades (Vea and Grimaldi, 2016; Ronquist et al., 2016).

5.2.2 Macroevolutionary - conditioning on survival

Some clades are never sampled simply because they did not survive long enough to be observed in the fossil record. This leads to a systematic undersampling of clades undergoing relatively high rates of extinction, which can lead to a bias in the estimation of background speciation and extinction rates. This effect can be accounted for by explicitly computing the conditional probability of a clade, given that at least one sample was recovered. Guindon (2018) devised an MCMC algorithm for computing this conditional probability under the fossilized birth-death process. Usually, however, the background macroevolutionary regime of speciation and extinction is not of interest when conducting inference on a single clade, and indeed may be essentially impossible to estimate with any accuracy (see for example the Lartillot, 2014, blog post). Therefore, in most cases it is probably unnecessary to condition on survival.

5.2.3 Stratigraphic - fossil sampling through time

It is well known that the fossil record is incomplete and unevenly sampled, and a wide range of factors impact what organisms are preserved and how. Obviously, organisms like mammals and snails with hard parts like a skeleton or shell will be better preserved in the fossil record, and will therefore be systematically oversampled. Similarly, fossils deposited in the more recent past will also be better preserved, and are therefore more likely to be recovered. Many paleontological studies have attempted to account for these biases when reconstructing estimates of species diversity in ancient clades (Sepkoski et al., 1981; Foote and Sepkoski, 1999; Raup, 1972, 1976). These studies have often relied on estimates obtained by combining sampling effort or proxy data with fossil abundance data in a multivariate model, a method known as “residual diversity estimation” (Smith and McGowan, 2007; Sakamoto et al., 2017).

While no studies have applied a similar approach jointly with the fossilized birth-death process, the mathematics enabling the estimation of time-heterogeneous fossil sampling and diversification have been described for the FBD (Gavryushkina et al., 2014). Thus, implementing a fossilized birth-death variation of the residual diversity estimation approach is feasible within a hierarchical modeling framework, and will be a valuable goal of future studies.

6 Prospectus

As the field of statistical phylogenetics has matured in the genomic era, vast quantities of molecular data have become widely available for studying a variety of extant (and some recently extinct) species. This has made it possible to obtain very good estimates of the evolutionary relationships among many clades whose relationships were previously unknown due to a lack of phylogenetically informative characters coming from other sources, particularly species with no representation in the fossil record. Ultimately, however, genomic data alone are insufficient for resolving the absolute ages of species divergences. Thus,

although advancements in sequencing technologies have yielded data that provide great evolutionary insights, reconstruction of the macroevolutionary timeline is limited to classical—and often laborious—methods in paleontology for collecting and dating fossil specimens. Consequently, technological innovations in collecting, organizing, and curating paleontological data will be critical if we are to make major progress in elucidating the absolute divergence times of the tree of life. In clades with poor fossil records, the development of models that account for additional sources of dating information, such as biogeographic patterns (Landis, 2017), or patterns in the conservation of horizontal gene transfer events (Davín et al., 2018) will be crucial for inferring divergence times. However, the timescale of evolutionary events in some clades without good representation in the fossil or other type of historic record may never be fully understood.

7 Acknowledgements

We wish to thank the editors F. Delsuc, N. Galtier, and C. Scornavacca for the opportunity to contribute our paper to *Phylogenetics in the Genomic Era*. Our colleagues J. Barido-Sottani, J. Buckner, W. Dismukes, J. Justison, K. Quinteros, J. Satler, and D. Żyła provided helpful comments that greatly improved this chapter.

References

- Arcila, D., Pyron, R. A., Tyler, J. C., Ortí, G., and Betancur-R, R. (2015). An evaluation of fossil tip-dating versus node-age calibrations in tetraodontiform fishes (Teleostei: Percomorphaceae). *Molecular Phylogenetics and Evolution*, 82:131–145.
- Bapst, D., Wright, A., Matzke, N., and Lloyd, G. (2016). Topology, divergence dates, and macroevolutionary inferences vary between different tip-dating approaches applied to fossil theropods (Dinosauria). *Biology Letters*, 12(7):20160237.
- Benton, M. J. and Donoghue, P. C. (2007). Paleontological evidence to date the tree of life. *Molecular Biology and Evolution*, 24(1):26–53.
- Boussau, B. and Scornavacca, C. (2020). Reconciling gene trees with species trees. In Scornavacca, C., Delsuc, F., and Galtier, N., editors, *Phylogenetics in the Genomic Era*, chapter 3.2, pages 3.2:1–3.2:23. No commercial publisher | Authors open access book.
- Bromham, L. (2020). Substitution rate analysis and molecular evolution. In Scornavacca, C., Delsuc, F., and Galtier, N., editors, *Phylogenetics in the Genomic Era*, chapter 4.4, pages 4.4:1–4.4:21. No commercial publisher | Authors open access book.
- Claramunt, S. and Cracraft, J. (2015). A new time tree reveals Earth history’s imprint on the evolution of modern birds. *Science Advances*, 1(11):e1501005.
- Davín, A. A., Tannier, E., Williams, T. A., Boussau, B., Daubin, V., and Szöllősi, G. J. (2018). Gene transfers can date the tree of life. *Nature Ecology & Evolution*, 2(5):904.
- Didier, G., Fau, M., and Laurin, M. (2017). Likelihood of tree topologies with fossils and diversification rate estimation. *Systematic Biology*, 66(6):964–987.
- Didier, G., Royer-Carenzi, M., and Laurin, M. (2012). The reconstructed evolutionary process with the fossil record. *Journal of Theoretical Biology*, 315:26–37.
- Donoghue, P. C. and Benton, M. J. (2007). Rocks and clocks: calibrating the tree of life using fossils and molecules. *Trends in Ecology & Evolution*, 22(8):424–431.
- Dornburg, A., Beaulieu, J. M., Oliver, J. C., and Near, T. J. (2011). Integrating fossil preservation biases in the selection of calibrations for molecular divergence time estimation. *Systematic Biology*, 60(4):519–527.

- dos Reis, M. (2016). Notes on the birth–death prior with fossil calibrations for Bayesian estimation of species divergence times. *Philosophical Transactions of the Royal Society B*, 371(1699):20150128.
- dos Reis, M. and Yang, Z. (2013). The unbearable uncertainty of Bayesian divergence time estimation. *Journal of Systematics and Evolution*, 51(1):30–43.
- Drummond, A. J., Ho, S. Y. W., Phillips, M. J., and Rambaut, A. (2006). Relaxed phylogenetics and dating with confidence. *PLoS Biology*, 4(5):e88.
- Fernández, R., Edgecombe, G. D., and Giribet, G. (2016). Exploring phylogenetic relationships within Myriapoda and the effects of matrix composition and occupancy on phylogenomic reconstruction. *Systematic Biology*, 65(5):871–889.
- Foote, M. (1996). On the probability of ancestors in the fossil record. *Paleobiology*, 22:141–151.
- Foote, M., Hunter, J. P., Janis, C. M., and Sepkoski, J. J. (1999). Evolutionary and preservational constraints on origins of biologic groups: divergence times of eutherian mammals. *Science*, 283(5406):1310–1314.
- Foote, M. and Sepkoski, J. J. (1999). Absolute measures of the completeness of the fossil record. *Nature*, 398(6726):415.
- Gavryushkina, A., Heath, T. A., Ksepka, D. T., Stadler, T., Welch, D., and Drummond, A. J. (2017). Bayesian total-evidence dating reveals the recent crown radiation of penguins. *Systematic Biology*, 66(1):57–73.
- Gavryushkina, A., Welch, D., Stadler, T., and Drummond, A. J. (2014). Bayesian inference of sampled ancestor trees for epidemiology and fossil calibration. *PLoS Computational Biology*, 10(12):e1003919.
- Gernhard, T. (2008). The conditioned reconstructed process. *Journal of Theoretical Biology*, 253:769–778.
- Grimm, G. W., Kapli, P., Bomfleur, B., McLoughlin, S., and Renner, S. S. (2014). Using more than the oldest fossils: dating Osmundaceae with three Bayesian clock approaches. *Systematic Biology*, 64(3):396–405.
- Guindon, S. (2018). Accounting for calibration uncertainty: Bayesian molecular dating as a “doubly intractable” problem. *Systematic Biology*, 67(4):651–661.
- Heath, T. A. (2008). *Understanding the importance of taxonomic sampling for large-scale phylogenetic analyses by simulating evolutionary processes under complex models*. PhD thesis, University of Texas at Austin, <https://repositories.lib.utexas.edu/handle/2152/18347>.
- Heath, T. A. (2012). A hierarchical Bayesian model for calibrating estimates of species divergence times. *Systematic Biology*, 61(5):793–809.
- Heath, T. A., Huelsenbeck, J. P., and Stadler, T. (2014). The fossilized birth-death process for coherent calibration of divergence-time estimates. *Proceedings of the National Academy of Sciences, USA*, 111(29):E2957–E2966.
- Hedges, S. B. and Kumar, S. (2004). Precision of molecular time estimates. *Trends in Genetics*, 20(5):242–247.
- Heled, J. and Drummond, A. J. (2012). Calibrated tree priors for relaxed phylogenetics and divergence time estimation. *Systematic Biology*, 61(1):138–149.
- Herrera, J. P. and Dávalos, L. M. (2016). Phylogeny and divergence times of lemurs inferred with recent and ancient fossils in the tree. *Systematic Biology*, 65(5):772–791.
- Ho, S. Y. (2007). Calibrating molecular estimates of substitution rates and divergence times in birds. *Journal of Avian Biology*, 38(4):409–414.
- Ho, S. Y. W. and Phillips, M. J. (2009). Accounting for calibration uncertainty in phylogenetic estimation of evolutionary divergence times. *Systematic Biology*, 58:367–380.

- Höhna, S., Heath, T. A., Boussau, B., Landis, M. J., Ronquist, F., and Huelsenbeck, J. P. (2014). Probabilistic graphical model representation in phylogenetics. *Systematic Biology*, 63(5):753–771.
- Höhna, S., Stadler, T., Ronquist, F., and Britton, T. (2011). Inferring speciation and extinction rates under different sampling schemes. *Molecular Biology and Evolution*, 28(9):2577–2589.
- Inoue, J., Donoghue, P. C., and Yang, Z. (2009). The impact of the representation of fossil calibrations on Bayesian estimation of species divergence times. *Systematic Biology*, 59(1):74–89.
- Jarvis, E. D., Mirarab, S., Aberer, A. J., Li, B., Houde, P., Li, C., Ho, S. Y., Faircloth, B. C., Nabholz, B., Howard, J. T., et al. (2014). Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science*, 346(6215):1320–1331.
- Kendall, D. G. (1948). On the generalized “birth-and-death” process. *The Annals of Mathematical Statistics*, 19(1):1–15.
- Kozlov, A. M. and Stamatakis, A. (2020). Using raxml-ng in practice. In Scornavacca, C., Delsuc, F., and Galtier, N., editors, *Phylogenetics in the Genomic Era*, chapter 1.3, pages 1.3:1–1.3:25. No commercial publisher | Authors open access book.
- Landis, M. J. (2017). Biogeographic dating of speciation times using paleogeographically informed processes. *Systematic Biology*, 66(2):128–144.
- Lartillot, N. (2014). Should we condition on non-extinction? Blog: The Bayesian Kitchen (<http://bayesiancook.blogspot.com/2014/11/should-we-condition-on-non-extinction.html>).
- Lartillot, N. (2020a). The bayesian approach to molecular phylogeny. In Scornavacca, C., Delsuc, F., and Galtier, N., editors, *Phylogenetics in the Genomic Era*, chapter 1.4, pages 1.4:1–1.4:17. No commercial publisher | Authors open access book.
- Lartillot, N. (2020b). Phylobayes: Bayesian phylogenetics using site-heterogeneous models. In Scornavacca, C., Delsuc, F., and Galtier, N., editors, *Phylogenetics in the Genomic Era*, chapter 1.5, pages 1.5:1–1.5:16. No commercial publisher | Authors open access book.
- Luo, A., Duchêne, D. A., Zhang, C., Zhu, C.-D., and Ho, S. Y. W. (2020). A simulation-based evaluation of tip-dating under the fossilized birth–death process. *Systematic Biology*, 69(2):325–344.
- Marshall, C. R. (2008). A simple method for bracketing absolute divergence times on molecular phylogenies using multiple fossil calibration points. *The American Naturalist*, 171(6):726–742.
- Müller, J. and Reisz, R. R. (2005). Four well-constrained calibration points from the vertebrate fossil record for molecular clock estimates. *BioEssays*, 27(10):1069–1075.
- Nee, S., May, R. M., and Harvey, P. H. (1994). The reconstructed evolutionary process. *Philosophical Transactions of the Royal Society B*, 344:305–311.
- Orlando, L., Ginolhac, A., Zhang, G., Froese, D., Albrechtsen, A., Stiller, M., Schubert, M., Cappellini, E., Petersen, B., Moltke, I., et al. (2013). Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. *Nature*, 499(7456):74.
- Parham, J. F., Donoghue, P. C. J., Bell, C. J., Calway, T. D., Head, J. J., Holroyd, P. A., Inoue, J. G., Irmis, R. B., Joyce, W. G., Ksepka, D. T., Patané, J. S. L., Smith, N. D., Tarver, J. E., van Tuinen, M., Yang, Z., Angielczyk, K. D., Greenwood, J. M., Hipsley, C. A., Jacobs, L., Makovicky, P. J., Müller, J., Smith, K. T., Theodor, J. M., Warnock, R. C. M., and Benton, M. J. (2012). Best practices for justifying fossil calibrations. *Systematic Biology*, 61(2):346–359.

- Pupko, T. and Mayrose, I. (2020). A gentle introduction to probabilistic evolutionary models. In Scornavacca, C., Delsuc, F., and Galtier, N., editors, *Phylogenetics in the Genomic Era*, chapter 1.1, pages 1.1:1–1.1:21. No commercial publisher | Authors open access book.
- Pyron, R. A. (2011). Divergence time estimation using fossils as terminal taxa and the origins of Lissamphibia. *Systematic Biology*, 60:466–481.
- Pyron, R. A. (2016). Novel approaches for phylogenetic inference from morphological data and total-evidence dating in squamate reptiles (lizards, snakes, and amphisbaenians). *Systematic Biology*, 66(1):38–56.
- Rannala, B. (2016). Conceptual issues in Bayesian divergence time estimation. *Philosophical Transactions of the Royal Society B*, 371(1699):20150134.
- Raup, D. M. (1972). Taxonomic diversity during the Phanerozoic. *Science*, 177(4054):1065–1071.
- Raup, D. M. (1976). Species diversity in the Phanerozoic: an interpretation. *Paleobiology*, 2(4):289–297.
- Reisz, R. R. and Müller, J. (2004). Molecular timescales and the fossil record: a paleontological perspective. *Trends in Genetics*, 20(5):237–241.
- Ronquist, F., Klopstein, S., Vilhelmsen, L., Schulmeister, S., Murray, D. L., and Rasnitsyn, A. P. (2012). A total-evidence approach to dating with fossils, applied to the early radiation of the Hymenoptera. *Systematic Biology*, 61(6):973–999.
- Ronquist, F., Lartillot, N., and Phillips, M. J. (2016). Closing the gap between rocks and clocks using total-evidence dating. *Philosophical Transactions of the Royal Society B*, 371(1699):20150136.
- Sakamoto, M., Venditti, C., and Benton, M. J. (2017). ‘Residual diversity estimates’ do not correct for sampling bias in palaeodiversity data. *Methods in Ecology and Evolution*, 8(4):453–459.
- Saladin, B., Leslie, A. B., Wüest, R. O., Litsios, G., Conti, E., Salamin, N., and Zimmermann, N. E. (2017). Fossils matter: improved estimates of divergence times in *Pinus* reveal older diversification. *BMC Evolutionary Biology*, 17(1):95.
- Sanders, K. L. and Lee, M. S. (2007). Evaluating molecular clock calibrations using Bayesian analyses with soft and hard bounds. *Biology Letters*, 3(3):275–279.
- Schenk, J. J. (2016). Consequences of secondary calibrations on divergence time estimates. *PLoS One*, 11(1):e0148228.
- Schuster, A., Vargas, S., Knapp, I. S., Pomponi, S. A., Toonen, R. J., Erpenbeck, D., and Wörheide, G. (2018). Divergence times in demosponges (Porifera): first insights from new mitogenomes and the inclusion of fossils in a birth-death clock model. *BMC Evolutionary Biology*, 18(1):114.
- Sepkoski, J. J., Bambach, R. K., Raup, D. M., and Valentine, J. W. (1981). Phanerozoic marine diversity and the fossil record. *Nature*, 293(5832):435.
- Slater, G. J., Cui, P., Forasiepi, A. M., Lenz, D., Tsangaras, K., Voirin, B., de Moraes-Barros, N., MacPhee, R. D., and Greenwood, A. D. (2016). Evolutionary relationships among extinct and extant sloths: the evidence of mitogenomes and retroviruses. *Genome Biology and Evolution*, 8(3):607–621.
- Slater, G. J., Goldbogen, J. A., and Pyenson, N. D. (2017). Independent evolution of baleen whale gigantism linked to Plio-Pleistocene ocean dynamics. *Royal Society B: Biological Sciences*, 284(1855):20170546.
- Smith, A. B. and McGowan, A. J. (2007). The shape of the Phanerozoic marine palaeodiversity curve: how much can be predicted from the sedimentary rock record of Western Europe? *Palaeontology*, 50(4):765–774.

- Stadler, T. (2009). On incomplete sampling under birth-death models and connections to the sampling-based coalescent. *Journal of Theoretical Biology*, 261(1):58–66.
- Stadler, T. (2010). Sampling-through-time in birth-death trees. *Journal of Theoretical Biology*, 267(3):396–404.
- Stadler, T., Gavryushkina, A., Warnock, R. C., Drummond, A. J., and Heath, T. A. (2018). The fossilized birth-death model for the analysis of stratigraphic range data under different speciation modes. *Journal of Theoretical Biology*, 447:41–55.
- Strauss, D. and Sadler, P. M. (1989). Classical confidence intervals and Bayesian probability estimates for ends of local taxon ranges. *Mathematical Geology*, 21(4):411–427.
- Tavaré, S., Marshall, C. R., Will, O., Soligo, C., and Martin, R. D. (2002). Using the fossil record to estimate the age of the last common ancestor of extant primates. *Nature*, 416(6882):726.
- Thompson, E. A. (1975). *Human Evolutionary Trees*. Cambridge University Press, Cambridge, UK.
- Upham, N. S., Esselstyn, J. A., and Jetz, W. (2019). Inferring the mammal tree: Species-level sets of phylogenies for questions in ecology, evolution, and conservation. *PLoS Biology*, 17(12):e3000494.
- Vea, I. M. and Grimaldi, D. A. (2016). Putting scales into evolutionary time: the divergence of major scale insect lineages (Hemiptera) predates the radiation of modern angiosperm hosts. *Scientific Reports*, 6:23487.
- Warnock, R. C., Yang, Z., and Donoghue, P. C. (2012). Exploring uncertainty in the calibration of the molecular clock. *Biology Letters*, 8(1):156–159.
- Wilkinson, R. D., Steiper, M. E., Soligo, C., Martin, R. D., Yang, Z., and Tavaré, S. (2011). Dating primate divergences through an integrated analysis of palaeontological and molecular data. *Systematic Biology*, 60:16–31.
- Wright, D. F., Zamora, S., and Rahman, I. A. (2017). Bayesian estimation of fossil phylogenies and the evolution of early to middle Paleozoic crinoids (Echinodermata). *Journal of Paleontology*, 91(4):799–814.
- Yang, Z. and Rannala, B. (2005). Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. *Molecular Biology and Evolution*, 23(1):212–226.
- Yang, Z. and Yoder, A. D. (2003). Comparison of likelihood and Bayesian methods for estimating divergence times using multiple gene loci and calibration points, with application to a radiation of cute-looking mouse lemur species. *Systematic Biology*, 52(5):705–716.
- Yule, G. U. (1924). A mathematical theory of evolution, based on the conclusions of Dr. J. C. Wills, F. R. S. *Philosophical Transactions of the Royal Society of London, Biology*, 213:21–87.
- Zhang, C., Stadler, T., Klopstein, S., Heath, T. A., and Ronquist, F. (2016). Total-evidence dating under the fossilized birth-death process. *Systematic Biology*, 65(2):228–249.
- Zhukova, A., Gascuel, O., Duchêne, S., Ayres, D. L., Lemey, P., and Baele, G. (2020). Efficiently analysing large viral data sets in computational phylogenomics. In Scornavacca, C., Delsuc, F., and Galtier, N., editors, *Phylogenetics in the Genomic Era*, chapter 5.3, pages 5.3:1–5.3:43. No commercial publisher | Authors open access book.