ANNUAL REVIEWS

Annual Review of Ecology, Evolution, and Systematics

Integrating Fossil Observations Into Phylogenetics Using the Fossilized Birth–Death Model

April M. Wright,[1] David W. Bapst,[2]
Joëlle Barido-Sottani,[3] and Rachel C.M. Warnock[4]

[1] Department of Biological Sciences, Southeastern Louisiana University, Hammond, Louisiana, USA; email: april.wright@selu.edu

[2] Department of Geology and Geophysics, Texas A&M University, College Station, Texas, USA

[3] Institut de Biologie de l'École Normale Supérieure (IBENS), ENS, CNRS, INSERM, Université PSL (Paris Sciences & Lettres), Paris, France

[4] GeoZentrum Nordbayern, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany

Annu. Rev. Ecol. Evol. Syst. 2022. 53:12.1–12.23

The *Annual Review of Ecology, Evolution, and Systematics* is online at ecolsys.annualreviews.org

https://doi.org/10.1146/annurev-ecolsys-102220-030855

Keywords

paleontology, Bayesian, fossils, phylogeny, diversification, sampling

Abstract

Over the past decade, a new set of methods for estimating dated trees has emerged. Originally referred to as the fossilized birth–death (FBD) process, this single model has expanded to a family of models that allows researchers to coestimate evolutionary parameters (e.g., diversification, sampling) and patterns alongside divergence times for a variety of applications from paleobiology to real-time epidemiology. We provide an overview of this family of models. We explore the ways in which these models correspond to methods in quantitative paleobiology, as the FBD process provides a framework through which neontological and paleontological approaches to phylogenetics and macroevolution can be unified. We also provide an overview of challenges associated with applying FBD models, particularly with an eye toward the fossil record. We conclude this piece by discussing several exciting avenues for the inclusion of fossil data in phylogenetic analyses.

1. INTRODUCTION: THE BIRTH OF A JOINT MODEL OF DIVERSIFICATION AND PHYLOGENY

For many years, dating of phylogenetic trees was primarily accomplished via the process of constraining the age of nodes using fossil and other temporal information. However, over the past two decades, new methods have been proposed that allow fossils to be more completely integrated into phylogenetic dating analyses. These methods rely on a series of models, collectively called fossilized birth–death (FBD) processes (Stadler 2010, Heath et al. 2014), that enable a rich interplay between morphology, molecules, and stratigraphy. The derivation of the original FBD model is one of the most important steps forward for the synthesis of neontological and paleobiological perspectives in phylogenetic methodology in recent memory. As we lay out in this manuscript, inferences using FBD processes fulfill many of the goals of quantitative paleontologists, while offering neontologists new ways to ask an array of questions about topology, time, and evolutionary tempo.

Birth–death processes are widely used in phylogenetics and phylodynamics to model the birth and death of lineages in a phylogeny.An overview of this family of models can be found in MacPherson et al. (2021). The FBD model in particular was originally proposed as a sampling model to describe how diversification and sampling lead to an observed set of samples on a phylogeny (Stadler 2010). The model consists of four main parameters: λ, the rate at which lineages are added to the tree (often called the speciation rate); μ, the rate at which lineages are removed from the tree, also known as the extinction rate; ψ , the rate at which lineages are sampled in the past, typically called the fossil sampling rate; and ρ, the sampling probability in the present. The FBD model is different from other birth–death-sampling models in representing present and fossil sampling as different quantities. For the model to be identifiable, at least one parameter has to be fixed, usually ρ in macroevolutionary analyses. λ, μ, and ψ are typically treated as unknown parameters in the analysis (**Figure 1**), and values for these parameters are sampled via Markov chain Monte Carlo (MCMC) methods in a Bayesian analysis.

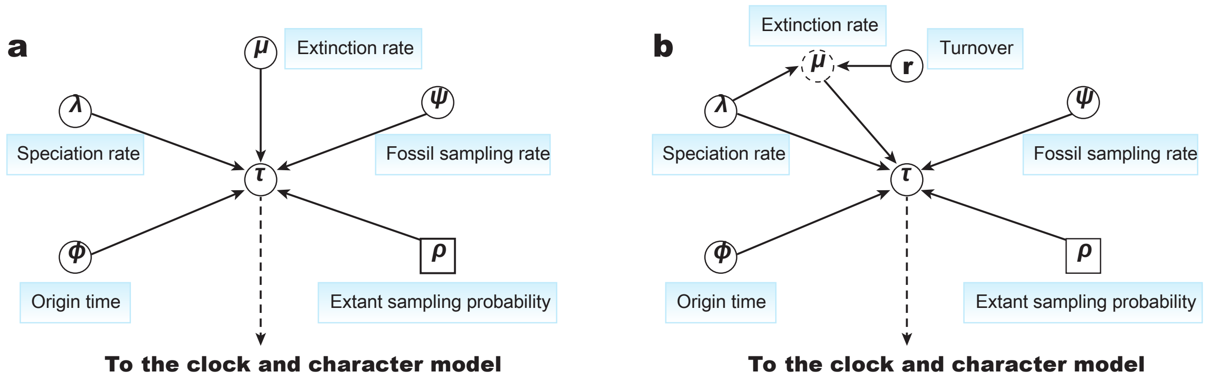

Figure 1

A graphical representation of the fossilized birth–death (FBD) model under two different parameterizations. In both cases, the dated phylogeny (indicated by the T in the middle of the graph) is the product of a set of parameters. (*a*) The tree is drawn from a distribution parameterized by a time of origination and rates of speciation, extinction, and fossil sampling, all of which are parameters estimated in the inference. The proportion of extant lineages sampled is typically assumed to be known without error (indicated by a *black box*). (*b*) A reparameterization of the model in which extinction is treated as a deterministic variable, transformed from the turnover rate. While these models use the same calculations, the one in panel *b* indicates a line of thinking in which speciation and extinction are not fully independent. The flexibility of implementing the FBD model in a Bayesian context allows researchers to constrain priors or reparametrize models to better reflect their own expectations about the process of evolution. In this figure, extinction rate is in a dotted circle, indicating that it is a stochastic node, that is, a node transformed from two model variables.

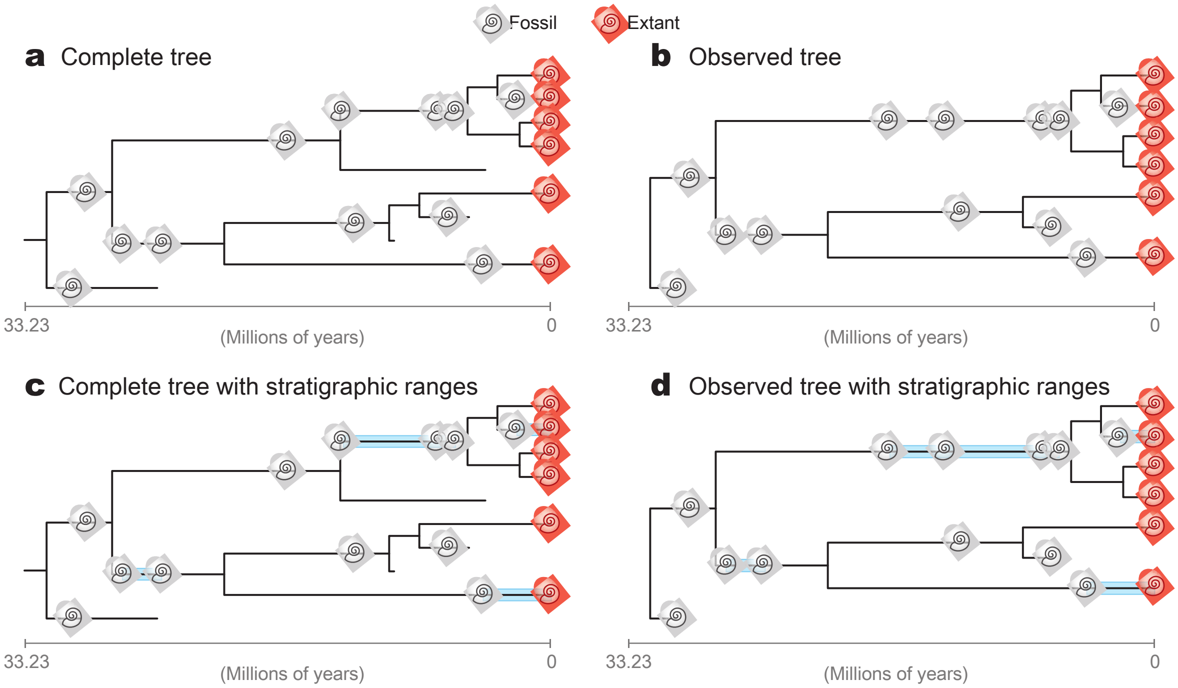

Figure 2

Alternative representations of a tree simulated under the fossilized birth–death (FBD) model. (*a*) Under the FBD process, there is considered to be a complete tree. This is a tree that is the total outcome of the FBD process, including fossil sampling (*gray shells*) and extant sampling (*red shells*). (*b*) The observed tree is the tree that can be estimated from the data actually observed by researchers. Lineages sampled in the present are indicated in red. In the complete process, all lineages appear. In the observed trees, this is not the case, as sampling may be patchy, and thus not every lineage is observed. Another factor in the FBD model is that lineages may have descendants on a tree whose ancestor is also sampled. These are often referred to as sampled ancestors. (*c*,*d*) The blue areas indicate stratigraphic ranges for a morphotaxon, in which there may be many observations of that taxon.

This model was applied to divergence time estimation in a phylogenetic context several years later (Gavryushkina et al. 2014, Heath et al. 2014) and has since been extended to fit a variety of sampling and biological conditions. The FBD model is a different way of thinking about a phylogeny. In many neontological studies, a tree is estimated from molecular or morphological data, and trees are often drawn from a uniform distribution in which no tree is a priori more likely than another. As discussed above, the FBD model is a diversification and sampling model. It describes the series of speciation and extinction events that lead to the true tree. Because sampling is inherently patchy, in both the present and the past, the FBD model also describes the process of sampling that leads to the tree we actually observe (**Figure 2**). We can further inform the inference of the topology by including fossil ages, as the FBD's sampling model assigns some trees higher likelihoods than others and thus ensures that we are more likely to recover trees that agree with our historical record of fossil samples. The age of a fossil tip on the tree can be used to inform the topology and the age of the speciation events that generate the clade to which the fossil belongs. The FBD process has been referred to as both a tree model (because it describes the process that generates the tree with fossils and extant samples) and a tree prior (because certain trees are not considered as likely, a priori, due to the tree model).

In a Bayesian phylogenetic analysis using the FBD model as a tree prior, character data are used to estimate a tree. Both extinct and extant tips of the tree are considered to be part of the same diversification process, and hence, both molecular and morphological data can be included in the analysis via a model of character evolution (Gavryushkina et al. 2016, Zhang et al. 2016). For morphological data, this is often accomplished via the Markov *k*-state (Mk) model (Lewis 2001), which has been demonstrated to be reasonably effective under a variety of conditions (Wright & Hillis 2014). Molecular data are incorporated via any number of models of sequence evolution (Jukes et al. 1969, Hasegawa et al. 1985, Tavaré 1986). Each data type is usually placed in a separate partition, and all partitions are considered to inform the overall model likelihood. It should be noted that each data type can be subpartitioned, for example, by gene or anatomical subregion [for an example, see Clarke & Middleton (2008)]. Other data types, such as continuous traits (Parins-Fukuchi 2018), can also be used in these analyses, though this is less common at the present.

For fossils that do not have character information, the exact phylogenetic placement of the taxon is not estimated. In this case, the fossil is constrained to a clade, which allows its age information to inform the most recent common ancestor (MRCA) age estimate of that clade. The exact phylogenetic placement is marginalized out of the calculation (Gavryushkina et al. 2014, Heath et al. 2014). In this way, all available fossils can be used, even if there are multiple per clade or not all fossils have character information.

The distribution of rates of evolution on a tree is described via the clock model. These models can come in a number of variations. The simplest is the strict molecular clock, in which all branches have a single rate of evolution (Zuckerkandl & Pauling 1965). While this model is the simplest, it is also the most restrictive. On the other end of the spectrum are uncorrelated clocks, which imply that an ancestor's rate of evolution may be very different than that of its descendants. These are some of the more flexible clock models. For thorough comparisons between these types of clock models, see Lepage et al. (2006, 2007).

Together, the tree, clock, and character models have been termed the tripartite model for inference with the FBD model (Warnock & Wright 2020). While these concepts—character models, clock models, distribution of evolutionary rates—are couched in a molecular lingo, the goals of this analysis are shared by life scientists of all stripes. A character model allows us to test questions about the evolution of characters, such as anatomical partitioning, ordering, and character rate asymmetry (Clarke & Middleton 2008, Bapst et al. 2018). The clock model allows researchers to test hypotheses about how evolutionary rates should be distributed over time. For example, early burst models versus constant rates of evolution over time can both be tested through clock models (Wright et al. 2021). Quantitative paleontologists have spent decades assessing diversification dynamics from counts of species (or of other taxonomic levels) through time, but analyzing diversification from phylogenies allows for a more integrated approach for accounting for sampling, rate heterogeneity, and of course, relatedness. Furthermore, the FBD model is the spiritual successor of various approaches to integrating the timing of fossil occurrences with phylogenetics.

2. FOSSILS IN PHYLOGENETIC PALEONTOLOGY: DIFFERING PERSPECTIVES

Paleontologists have had a lengthy and complicated history with phylogenetics. Before modern quantitative cladistics was developed in the middle of the twentieth century, paleontologists often claimed the role of final arbiters in many questions of systematics, as fossils were the only direct source of information about the evolutionary past of modern species. Qualitative assessments of morphological similarity and the stratigraphic order of appearance were given a high degree of weight in determining support for relationships, and reading the rocks was taken very literally, with the evolution of groups often depicted in scholarly papers and textbooks as chains of fossil ancestors and descendants.

New developments in molecular biology and computational resources in the 1960s energized the role of neontologists in systematics, such that by the end of the twentieth century, phylogenetics and systematics had become dominated by those willing and able to collect character data, particularly molecular sequence data, and analyze it with reproducible algorithms. Both paleontologists and neontologists developed new procedures and philosophical approaches to codifying phenotypic differences as matrices of morphological characters, but the use of morphological data in phylogenetics has remained a minority approach, especially as genome-level data has become more available. And yet, despite all this, there remain in 2022, systematists in both neontology and paleontology who still rely heavily on qualitative interpretations of morphology, and there are even paleontologists who lean very heavily on qualitative interpretations based on the stratigraphic order of appearance of taxa.

2.1. Fossil Sampling Times and the Fossilized Birth–Death Process

When we incorporate fossil sampling times into phylogenetic analyses within a mechanistic model-based framework, it becomes extremely important to consider what these ages represent (Hopkins et al. 2018). Under the FBD model, samples are recovered continuously along each lineage (or branch) according to a Poisson sampling process with rate ψ (Stadler 2010).

Within a paleontological context, these samples can be interpreted as fossil occurrence data collected from a sedimentary rock unit, deposited at some unknown point in the geologic past. In reality, we do not know the precise age of practically any fossil. Each occurrence (or coeval occurrences sampled from the same locality or horizon) is associated with a range of ages, reflecting the imprecise temporal resolution of the stratigraphic record – the relative order and dating of sedimentary rock layers and the traces of organisms recorded as fossil within those rocks. We can account for fossil age uncertainty within a Bayesian framework with the use of priors (see text for further discussion) (Drummond & Stadler 2016, Barido-Sottani et al. 2019).

The processes that control the preservation and sampling of a given fossil occurrence, even in the simplest form of the FBD model, are the product of more than simple diversification. A diversifying clade has more lineages and thus is much more likely to have a rich fossil record, according to the FBD's model of fossil sampling, than a group with a depressed speciation rate and a heightened extinction rate. However, the presence of a lineage is a necessary but not sufficient condition for fossilization, and a lineage might exist for millions of years before leaving traces we recover from the fossil record, a fact which is clear given lengthy gaps in time between occurrences of the same taxon in the fossil record. Such gaps in a lineage can extend to an entire clade composed of multiple taxa, such as the apparently long gaps in the Mesozoic record of eutherian ancestors (e.g., Luo et al. 2011). Thus, the age uncertainty of an entire lineage might differ considerably from the age uncertainty that we have for a single specimen. This has important implications for our use of fossil occurrences as node calibrations or within the FBD process.

The original FBD model assumes that fossil taxa are known from discrete occurrences reflecting an event in which traces of individuals of that taxon were preserved and later sampled by paleontologists. However, taxa in the fossil record are recognized based on their morphology, and so in well-preserved groups, particularly the fossils of marine invertebrates, the morphological characteristics designating a single species or even subspecies might be observed from many occurrences, across millions or tens of millions of years. Paleontologists often focus their work on identifying the first and last appearance times of these persistent morphotaxa (**Figure 2**), and the intervening time between is often referred to as a taxon's stratigraphic range. In some cases, the first and last appearance times of taxa (or the imprecise stratigraphic intervals to which these first and last appearances can be placed) might be much better established than information about the age of each individual occurrence, or even the number and distribution of occurrences in between. When we have only stratigraphic range data, it is more appropriate to use the taxon-level FBD process or the FBD range model described in Section 3.1, both of which explicitly incorporate stratigraphic ranges (Stadler et al. 2018). Although it should be noted that a researcher who has both ranges and counts of occurrences can make use of either model.

2.2. Stratigraphy Wars: The Order and Timing of Appearances in the Fossil Record and Their Use in Phylogenetics

The role of stratigraphy has been a fraught issue in systematics. Although some scientists criticize Darwin for perhaps overemphasizing the incompleteness of the fossil record in the Origin of Species (Darwin 1859,Raup 1979,Gould 1980), the truth remains that the fossil record is very much an incomplete record and cannot be read literally. We should expect gaps and missing lineages even in the best fossil records, and thus the order of stratigraphic appearance is at least occasionally misleading. Simultaneously though, unless the fossil record is the product of bizarre and nonuniformitarian processes that we wholly do not understand, it is very difficult to imagine how the fossil record could be completely at odds with the broad order and timing of actual evolutionary events. Of course, this creates a prime situation for misunderstandings: In general, it is easier for the human mind to work with absolutes rather than probabilistic expectations (in other words, it is not that the fossil record is always accurate or always inaccurate but rather that the fossil record is often accurate but is also often imperfect and sometimes very misleading).

The initial use of stratigraphy in inferring or testing phylogenetic relationships was purely qualitative, depending greatly on expert opinion of whether a group's stratigraphic record was trustworthy. Sometimes, paleontologists argued that certain morphological features evolved multiple times, given gaps in the stratigraphic history implying that the lineage holding that feature had gone extinct. This view was often countered by those who preferred to interpret any apparent morphological similarity as the result of an imperfect fossil record, or relict lineages that had survived in some province not captured within the paleontological record. These illusionary gaps in evolutionary history, when certain species apparently had originated but were not yet preserved as recognizable remains in the fossil record, were known as ghost lineages [sometimes also synonymous or similar to ghost branches, ghost taxa, and/or ghost ranges (Norell 1992)]. There is truth to both of these perspectives: While convergence and iterative evolution are quite real phenomena, their true extent was both exaggerated and understated without quantitatively assessing the weight of stratigraphic evidence against evidence from phenotypic characters.

With the rise of quantitative phylogenetics, some paleontologists developed methods of evaluating alternative phylogenetic hypotheses based on their agreement with the stratigraphic order of appearance of taxa on those topologies (Benton & Storrs 1994, Huelsenbeck 1994, Siddall 1998, Wills 1999). However, such stratigraphic congruence measures were sometimes unsatisfying as a way of choosing optimal phylogenetic hypotheses because we would expect some incongruence with the order of fossil appearance in real empirical data sets. The greater issue is that we did not yet have models that could tell us how much incongruence to expect. Foote et al. (1999) tried to tackle such models for a single specific evolutionary event: the origin of crown-group mammals. Foote et al. restated this issue as a question of how much lower the preservation probability of Mesozoic eutherian mammals would have to be to support a substantial duration of missing evolutionary history relative to their well-sampled Cenozoic descendants. Interestingly, the models of Foote et al. (1999) were a precursor to the FBD process and are based on independent processes of speciation, extinction, and sampling.

Daniel Fisher (1980) proposed an approach called stratocladistics to quantify the mismatch between stratigraphy and phylogeny as stratigraphic debt, similar to the character-based debt used in maximum-parsimony analyses (a review of this concept can be found in Fisher 2008). In application (e.g., Bodenbender & Fisher 2001), a stratocladistics analysis involved combining a topology returned by a quantitative tree search with information about taxon ages, such that stratigraphically incongruent relationships and possible ancestor–descendant pairs were identified and then minimized by manual adjustment. Marcot & Fox (2008) developed stratocladistic software to simultaneously assess both character debt and stratigraphic debt against each other and identify the best solution without manual tuning. Substantial objections were raised, both on methodological grounds (how to balance stratigraphic debt with character debt) and philosophical grounds, particularly that character information should always take precedence over stratigraphic information. This led some workers to adopt a hard stance that relationships supported by morphological cladistics should never be overruled by incongruence with the fossil record. Paleontologists of all approaches and perspectives can be found in modern paleontology, and this mixture has sometimes resulted in particularly acrimonious disagreement. There is no better single example of this than a series of responses and comments published as a Nature Debate in 1998, edited and moderated by Andrew Smith (1998). In hindsight, these arguments possibly had the effect of diminishing the development and application of stratocladistic methods.

Some systematicists, in response to similar concerns, developed qualitative protocols for resolving ambiguous areas of topology with respect to stratigraphy, such as treating early-fossil species that lacked derived characteristics as sampled ancestors (Smith 1994). Alternatively, others developed methods for quantifying the probability of a taxon's existence before its first appearance in the fossil record and use confidence intervals to test the probability of long unsampled ghost lineages (Marshall 1995, Wagner 1995). Wagner (1998) combined these confidence-interval methods for testing ghost lineages with the procedures for comparing morphological and stratigraphic debt from stratocladistics, creating a method known as stratolikelihood phylogenetics, which Wagner applied to the fossil record of hyaenids. While in some ways this method was very close in philosophy to modern FBD approaches, it could use iterative approaches to calculate only pseudolikelihoods for a tree, given observed morphological data and stratigraphic appearances, and borrowed much from parsimony-based approaches.

Further progress was stymied by two obstacles: (*a*) the lack of a likelihood function adjusted for observed variation in morphology (later provided by Lewis 2001) and (*b*) the lack of a tractable likelihood function for a branching diagram that encapsulates the processes of birth, death, and sampling across time (ultimately developed as the FBD process, by Stadler 2010, Gavryushkina et al. 2014, Heath et al. 2014). Today, using the FBD model, we now have statistically valid methods for inferring dated phylogenies with sampled ancestors from the fossil record.

2.3. Post-Hoc Time Scaling and Other Diversions Before the Rise of FBD Models

In the late 2000s and early 2010s, the use of phylogenetic comparative methods for making macroevolutionary inferences flourished widely in evolutionary biology. Neontologists, and many paleontologists, developed and applied phylogenetic comparative methods, particularly some paleontologists who had not previously focused on systematics or stratigraphy. However, the vast majority of such methods required accurately dated trees, regardless of the difficulty at the time of acquiring a dated phylogeny—no quantitative method for dating phylogenetic relationships existed in paleontology. Even stratocladistic methods did not ultimately create a dated tree.

The initial reaction was to treat this as an afterthought and to date a topology (usually from a maximum-parsimony analysis or even something more abstract, such as a supertree) by referring directly to the timing of first appearances in the fossil record and assigning the age of the earliest appearing taxon in a clade as the node age for that clade (roughly described in Smith

1994). While such a literal dating of divergence dates certainly overlooks the potential for gaps in the fossil record, this was considered an acceptable approach. However, many workers encountered difficulties applying their desirable comparative methods, due to computational issues with the zero-length intranode edges in such dated phylogenies. Such zero-length edges existed between each node and its earliest descendant, and if there was any incongruence at all, such as an early-appearing member of a very derived group, the resulting phylogeny might have many such zero-length edges (Bapst 2013). To deal with this obstacle to applying comparative methods, new approaches were developed and implemented by Laurin (2004), Ruta et al. (2006), Brusatte et al. (2008), Hunt & Carrano (2010), and Lloyd et al. (2012). These methods adjusted the nodes of a dated paleontological phylogeny to be later in the past, thus stretching out the unwanted zero-length edges. The vast majority of these approaches worked independently of character information, except for the method of Ruta et al. (2006), which tried to assign older divergence dates to branches with more apparent character changes under parsimony.

These simple post hoc methods gave a false sense of certainty to dated paleontological phylogenies, rather than fully accounting for the uncertainties in estimating a time-calibrated phylogeny linking fossil taxa, which must not only be reconstructed from the incomplete fossil record but also account for the uncertainty surrounding potential ancestor–descendant relationships among the considered fossil occurrences (Bapst 2013). Using probability distributions drawn from the FBD model, Bapst (2013, 2014) introduced a post hoc method which would time scale an undated topology, given estimates of the rates of sampling, extinction, and speciation. This set of three rates gave the method its name (cal3 for three-rate calibrated). However, the independent estimation of rates was often considered difficult for many smaller data sets, or perhaps even impossible, and the method was not widely adopted (Bapst & Hopkins 2017).

Instead, Bayesian phylogenetics software soon fully integrated the FBD model as a tree prior for dated phylogenies of fossil taxa, allowing for approaches that simultaneously inferred the topology, the divergence dates, and the rates of major processes. Now, paleontologists seeking dated phylogenies are often pointed toward using fully simultaneous Bayesian inference rather than stepwise post hoc methods like cal3 [e.g., the advice in Bapst & Hopkins (2017)]. Importantly, even if character information isn’t available, such as for supertrees, the topological hypothesis can be set as constraints in most Bayesian phylogenetic software, and the dating of the tree can be optimized under the FBD process alone, without considering character information (Lloyd & Slater 2021).

2.4. Node Calibration: Approaches from Neontology

Node calibration approaches have been one of the most popular ways to time scale a phylogenetic tree, particularly for biologists working with molecular data. Under the node calibration framework, a fossil is assigned (not inferred as part of the analysis) to a subclade of the tree. The extant tree is typically estimated from molecular characters using a model of sequence evolution. Since the fossil is assumed to be a descendant of the same MRCA of the extant taxa in that subclade, it is assumed that this subclade must be at least as old as the fossil. On the surface, this is similar to the minimum-age dating described in Section 2.2. However, how long is the waiting time between the MRCA and the fossilization event (**Figure 3**)? In a Bayesian framework, this waiting time is typically parameterized by a node calibration density. The calibration density is a probability distribution that describes the time between the MRCA and the fossil. For example, in **Figure 4**, multiple different probability distributions can be seen, displaying different relationships between the MRCA and fossil. But what is the fossil calibration specifying? On the surface, this is a distribution on the amount of time we expect between a cladogenic speciation event and the sampling of the oldest fossil that subtends it. In reality, parameterizing the calibration density is more complicated, and it may not be possible to know if a fossil used for calibration has been sampled early

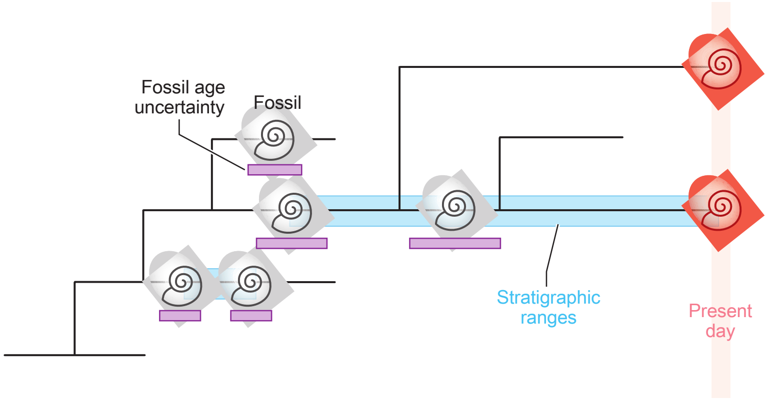

Figure 3

A complete tree highlighting the distinction between fossil age uncertainty and stratigraphic ranges. Here, a budding speciation process is assumed, and each horizontal lineage or branch represents a unique morphotaxon. Shells represent individual (extant or fossil) occurrences. The vertical pink bar represents the present. Fossils are recovered along lineages and each of these is associated with a range of fossil age uncertainty, indicated by the horizontal purple lines. A taxon can be represented by one or more sampled occurrences. Other taxa may be unsampled in the fossil record but survive to the present or be both unsampled as a fossil and go extinct before the present day. The interval between the first and last appearance is known as the stratigraphic range (*blue areas*).

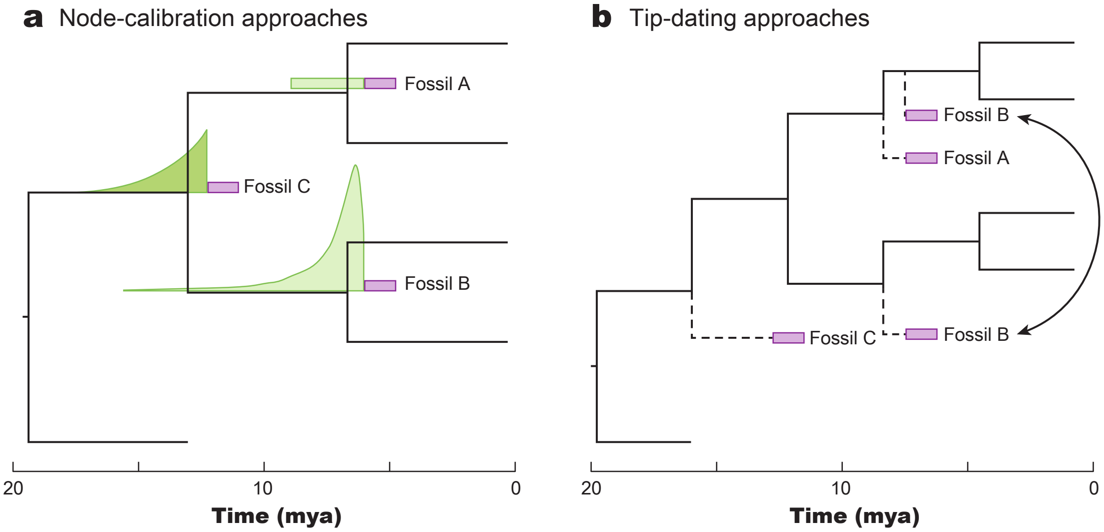

Figure 4

A schematic showing node calibration versus tip-dating methods. Fossils are represented with an age uncertainty (indicated by the *purple bar*). (*a*) The node-calibration densities specify the relationship between the fossil taxon and the split it subtends. On this graphic, the green distributions indicate the range of plausible ages for the divergence time between these two clades. As can be seen with Fossil B and Fossil C, it is possible for priors to conflict. (*b*) With tip-dating approaches, the position of fossils on the tree is estimated, as indicated by the two positions for Fossil B. Dashed lines on this phylogeny indicate that the placement is not fixed but estimated. No calibrations are needed, as a branch is estimated, connecting the fossil to the tree. It should be noted that fossils without character information can be used with the fossilized birth–death process but must be constrained to a clade. Abbreviation: mya, million years ago.

in the lineage's existence or late in its existence. The waiting time, therefore, is conflating many different processes: uncertainty on the ages of fossils, sampling, preservation, and speciation.

In practice, divergence time analyses tend to be extremely sensitive to their calibration priors. Warnock et al. (2012) demonstrated differences of tens or even hundreds of millions of years for the age of nodes, depending on the prior chosen. However, the selection of calibration priors is often arbitrary. As shown in **Figure 4**, distributions may appear similar for most parameter values but still have important differences in the tail of the distribution. And yet, choosing which distribution to use at all has persistently been a vexing task for biologists, in part because of the conflation of different quantities (uncertainty, waiting time) but also because choosing an appropriate prior relies on knowledge about the likelihood of the data under different probability distributions, which researchers may or may not have.

As mentioned in Section 2.3, under the node-calibration framework, fossils are placed by the researcher. It is assumed that the location of fossils on the tree is known without error, i.e., that the fossil can be placed on the tree to calibrate the age of a specific node. This means that collected phylogenetic characters for these specimens are not really used in the analysis. It also means that, in a Bayesian analysis, the relative support for different placements of the fossil is not evaluated. In this way, calibrations use only a subset of the total data that may be attached to a fossil.

It should be noted that over the years, there have been many efforts to standardize the ways in which fossil calibrations are used. The open-source Fossil Calibration Database (Ksepka et al. 2015), for example, implements the best practices for the choice and justification of fossils for calibrations laid out by Parham et al. (2011). In particular, these efforts have been made by paleontologists to attempt to assist molecular biologists in choosing appropriate fossils for calibrations. Other authors have described models to more adequately describe fossil uncertainty using decay-rate estimation (Terry & Novak 2015), incorporation of speciation rates and fossil sampling (Matschiner et al. 2017), and incorporation of confidence intervals on fossil age ranges [a useful summary of this literature can be found in Marshall (2010)]. Many of the approaches previously taken are similar to, if not subsumed by, the FBD process.

2.5. Toward a Useful Synthesis of Perspectives

As described in this section, both paleontologists and neontologists have struggled with how best to make use of the wealth of information in the fossil record. While it has been relatively straightforward to manage character data for estimation, how best to incorporate ages and stratigraphy have been more problematic. The FBD process provides an avenue to estimate the phylogeny while using a sampling process to incorporate information about fossil ages. In many ways, this reflects a traditional paleontological line of thinking about how fossils are sampled while being compatible with molecular data and techniques. It is also worth emphasizing that many of the lessons learned through the application of previous approaches remain relevant in the context of the FBD process, such as the best practices developed for reporting age and the taxonomic uncertainty associated with fossils used for node dating.

3. THE FOSSILIZED BIRTH–DEATH MODEL IN PRACTICE

3.1. Model Extensions to the Fossilized Birth–Death Process

Multiple extensions have been developed for the FBD process. To account for more realistic variations in birth, death, and sampling rates, the episodic (also called skyline) FBD process allows for phylogeny-wide changes at specific points in time, while the multistates (or multitypes) FBD process makes rates dependent on a lineage-specific character. Both of these models are implemented in the BDMM package for the Bayesian evolutionary analysis by sampling trees 2 (BEAST2) program (Kühnert et al. 2016) and thus can be combined further into a general FBD process incorporating both time- and lineage-dependent rate variation. More recently, the FBD-multispecies coalescent (MSC) model (Ogilvie et al. 2021) integrates the FBD process with the MSC model, allowing multiloci analyses in which the species tree and the morphological character matrix evolve under the FBD and Mk models, while the molecular gene trees evolve under the MSC model.

Other extensions have been developed to handle empirical data sets more accurately. Support for fully extinct trees, with no extant samples, was added to the BEAST2 implementation of the FBD model by Barido-Sottani et al. (2020). While the FBD process was originally designed to analyze specimen-level data (Heath et al. 2014), many available data sets contain only stratigraphic range information for fossil taxa, giving only the first and last occurrence of each morphospecies (or other morphotaxon). The FBD model was extended to account for these stratigraphic ranges in Stadler et al. (2018), and we refer to this model elsewhere as the taxon-level FBD process.

The FBD process has also been extended to better account for data sets in which some or all taxa might be missing character information. For most organismal groups, only a small proportion of the overall number of sampled fossil specimens have been identified with sufficient taxonomic resolution to place them in a phylogenetic context. Some fossilized traces may lack any usable morphological data at all, except for representing an unknown member of a broad taxon, such as trackways, burrows, coprolites, etc. One extension of the FBD process, the occurrence birth–death process (OBDP) (Andréoletti et al. 2020, Manceau et al. 2021), integrates a special class of events referred to as occurrences to deal with these phylogenetically ambiguous observations when the geologic age of specific collections is known.Occurrences in this context are fossilspecimens that lack any information on their position in the phylogeny (i.e., no morphological or taxonomic information is available for them), thus they cannot be integrated directly into the phylogeny. However, the ages of occurrences can still provide information on the number of lineages through time, as well as the birth and death rates. Under the OBDP, we can assign a different sampling rate for the occurrence subset of indistinct fossils. A new model, the FBD range process, was designed to estimate diversification and sampling rates from data sets that contain only stratigraphic range information, i.e., where no morphological or molecular sequences are available for reconstructing a phylogeny (Warnock et al. 2020) (for more details, see Section 4.2). Both the FBD range models and the OBDP are currently implemented in RevBayes.

3.2. Model-Based Challenges

While the FBD process and its extensions provide valuable tools to better account for empirical data sets and evolutionary processes, they also raise additional challenges. The first set of issues is tied to the models themselves and their implementation.

3.2.1. Identifiability issues. Recent work by Louca & Pennell (2020) has shown that the general birth–death model suffers from a major identifiability issue on extant trees, meaning that an infinity of functions for the birth rate and death rate would result in the same tree likelihood. Louca et al. (2021) then extended this result to trees with sampling through time, potentially including trees generated under the FBD process.

Most existing empirical results are not affected by this issue, as constraining the rates to be constant (as in the original FBD process) or piecewise constant (as in the episodic FBD process) restricts the parameter space to only one most likely configuration (Legried & Terhorst 2021, Louca et al. 2021). However, the implications for more complex models, such as models integrating lineage-specific variations in rate, have not been explored yet. In addition, this issue highlights that assumptions made by the user on the exact shape of the rate functions (e.g., constant, piecewise constant, exponentially decaying) can have a drastic impact on the result of the inference and thus should be carefully chosen.

Even models that are identifiable in theory may not remain so based on the limited amount of data present in actual empirical data sets. For instance, Černý et al. (2021) found that rate estimates obtained on fully extinct clades by models integrating lineage-specific rate variations were only partially reliable, a fact the authors attribute in part to the small size of the available trees.

3.2.2. Model selection issues. Another issue raised by more complex models concerns model fit, specifically whether we can adequately distinguish which model is the most appropriate for a given data set. Recent work suggests that stepping-stone analyses, which are commonly used to perform model selection, are inadequate for analyses involving fossil specimens, as these analyses fail to account properly for the evidence present in the fossil samples (May & Rothfels 2021). This can lead to a simpler model being chosen despite strong evidence of rate variation in the fossil record (May et al. 2021).

One possible solution is to use reversible-jump MCMC (rjMCMC), which allows us to explore different models as part of the same inference rather than comparing several inferences after the fact. However, rjMCMC requires significant work to integrate various models together (Green 1995). This is compounded by the current fragmentation of the implementation landscape, with different versions and extensions of the FBD process implemented in BEAST2, RevBayes, or standalone software.

3.2.3. Computational issues. One key challenge for Bayesian inference is the computational cost and time involved in getting the inference to converge. More complex models have a direct effect on the cost of analyses by adding new parameters to be estimated, and by increasing the cost of calculating the likelihood. They can also have an indirect effect by increasing the amount of data and the number of samples required to make the model identifiable and obtain accurate estimates, as the number of tips in the tree is a key factor in the computational cost of phylogenetic inferences. While a targeted effort on a specific model or class of models can lead to considerable speed-ups, as has been the case for the MSC implementation in BEAST2 (Ogilvie et al. 2017), such an effort has so far not happened for the FBD process.

In many empirical analyses, subsampling the fossil specimens remains necessary for obtaining results in a reasonable amount of time (Andréoletti et al. 2020, O'Reilly & Donoghue 2020). However, the choice of subsampling scheme can impact the accuracy of the inference and requires good understanding of the structure of the fossil record, as well as the position of fossils in the tree (O'Reilly & Donoghue 2020). While random subsampling leads to unbiased estimates of diversification rates, it runs the risks of removing the oldest fossils from certain clades, which can lead to clade age estimates that are incompatible with the fossil record. By contrast, selective sampling of the oldest fossils for each clade leads to overestimates of clade ages (Matschiner et al. 2017, Matschiner 2019).

Analyses fitting complex models integrating lineage-specific rate heterogeneity often try to limit the number of tip taxa used to reduce the computation time needed, but previous work has shown that such analyses are sensitive to excluding sampled ancestors from the fossil record (Beaulieu & O'Meara 2021). To work around some of these issues, several extensions to the FBD process, such as the taxon-level FBD model or the OBDP, are designed to add information to the inference without increasing the number of tips of the tree.

3.3. Data-Based Challenges

A second set of issues is tied to the empirical data to which the FBD process is applied, in particular the fossil record.

3.3.1. Accounting for fossilization and sampling rate variations. The fossil record is biased by many different factors that influence the process of fossilization but also the process of recovering and classifying fossil samples. For instance, environmental conditions and morphological characteristics have a large impact on the probability of a given species to be preserved in the fossil record (Benson et al. 2021). One major issue is the oscillatory structure of sedimentary records, often referred to as sequence stratigraphy, which results from compounding both short and long time-scale cycles in climate and sea-level change. As a result, any set of geologic beds contains a sequence of high-deposition layers alternating with low-deposition layers; some of the previously deposited rock record may even be erased via weathering (Patzkowsky & Holland 2012). In the fossil record, this has the additional complication of creating sequences of oscillating change corresponding to changes in water depth or elevation shifts at any single location, resulting in scenarios where taxa with long stratigraphic ranges may have many occurrences in a short time, as their preferred environment is retained by the burial of rock, and then a long gap with no occurrences before being briefly found again in abundance. This structuring of the fossil record violates the rate-homogenous sampling assumptions of the basic FBD model. Similarly, the presence of groups that are preserved only through exceptional fossilization processes under rare conditions, such as the organisms found in any given Lagerstätte, may be controlled by the timing of an available environmental window during which such preservation is possible. The process of data gathering itself can lead to biases: For example, geographical differences in resources available for data collection can result in large discrepancies in sampling between regions of the world (Monarrez et al. 2021, Raja et al. 2022). While some of these geographic biases can be corrected by improved sampling, many regions and habitats simply have no unaltered sedimentary records to investigate, leading to extreme scenarios, such as detailed records with a high number of occurrences being known from one continent, while the lineages on another continent during the same interval are completely unknowable. This spatial heterogeneity of the fossil record is one of its ubiquitous and frustrating patterns, reflecting that everything that comes to us from paleontology is at the fickle whims of long-ago changes in sediment deposition, erosion, and burial (Smith 2001, Patzkowsky & Holland 2012, Benson et al. 2013, Vilhena & Smith 2013).

Some of these biases, such as temporal or character-driven variations in fossilization rates, can be accounted for by extensions of the FBD process. However, this generally requires strong assumptions from the user, such as the specific timing of rate changes or the specific character or set of characters driving underlying rate variation. Previous work on birth–death models to integrate character-driven rate variations has shown that they are prone to false positive results when the driving character is misidentified (Rabosky & Goldberg 2015). In addition, accounting for all sources of biases requires many different characters to be integrated in the model. This can lead to an unsustainable number of additional parameters, as each character combination present in the data set is associated with its own rate regime.

An alternative to handling these biases in the analysis is to filter and normalize the data before inputting it into the inference, allowing the use of a simpler model. For instance, family or time-specific subsampling can be applied to the fossil record to reduce the discrepancies between fossil densities for each clade or time interval (Andréoletti et al. 2020). This approach is similar to those commonly included in nonphylogenetic paleontological methods (see Section 4.2). However, this makes the downstream analysis completely dependent on the accuracy of the filtering and means that the uncertainty present in the data is not adequately represented in the inference output. Additionally, there is no way to check that the filtering is indeed removing bias rather than simply replacing it with another.

Extant species sampling can also be biased. For example, diversified sampling of extant species, where only one representative of each genus or subfamily is included in the data set, is a common method of reducing the complexity of large phylogenetic inference. Accounting properly for this sampling scheme is critical to obtain accurate estimates of divergence times under the FBD process (Zhang et al. 2016, Matschiner 2019).

3.3.2. Handling uncertainty. Due to the uncertainty involved in dating fossils, fossil specimens are dated not to a specific age but rather to a range of possible ages. This uncertainty can either be resolved through the user assigning a fixed age to each specimen (generally the midpoint of the range or a randomly chosen point within the range), or integrated into the inference (Drummond & Stadler 2016). Comparing these two approaches shows that fixing the fossil ages leads to inaccurate estimates of divergence times, tree topology, and diversification rates in FBD analyses (Barido-Sottani et al. 2019, 2020).

3.3.3. Placing fossils in the phylogeny. As detailed in Section 1, fossil specimens can be placed in the phylogeny either by using information from the taxonomy, in the form of topological constraints on the tree, or through a total-evidence approach using morphological characters. Although accurate topological constraints have been shown to perform better in terms of tree reconstruction accuracy (Luo et al. 2020), total-evidence approaches allow for including more uncertainty in the inference and may be the only option for data sets where taxonomy is unclear, such as corals (Ardila et al. 2012).

The use of morphological character matrices raises several issues. First, the morphological matrix needs to be of sufficient size to obtain accurate estimates. Previous research has shown that 30-character matrices are not sufficient to obtain accurate tree topologies (Barido-Sottani et al. 2020) and that the accuracy of divergence time estimates drops markedly for matrices with fewer than 100 characters (Parins-Fukuchi & Brown 2017). If taxonomy is well-known, combining morphological matrices with some topological constraints can compensate for lower numbers of characters. However, the accuracy of fossil placement is always dependent on the amount of data provided to the inference.

A second issue with morphological character data is the ability of the substitution model to accurately represent the processes involved in the evolution of morphological processes, which are generally more complex than for molecular characters and may vary significantly between individual characters. Failing to account for correlations and dependencies between morphological characters has been shown to lead to inaccurate topology and divergence time estimates (Ronquist et al. 2016). However, more recent work suggests that morphological model mismatches are mostly unproblematic for root-age estimates if good molecular data are available (Klopfstein et al. 2019, Luo et al. 2020). Other research indicates that dating of other nodes for fossil-only data sets certainly performs better when using the FBD model, as opposed to previous approaches of inferring topologies separately and using post hoc dating methods, as the latter can artificially inflate the precision of age estimates (Bapst et al. 2016).

3.4. Human-based Challenges

The last set of issues concerns the use of the FBD process by empirical researchers and the potential misunderstandings and issues of representation which can arise in empirical studies.

3.4.1. Understanding complex models. As the range of available models is extended to include more and more complex variations, one key challenge is how to improve the communication between users and developers. Understanding the different models and their assumptions and limitations is critical for users to be able to decide which is the most appropriate for their data sets. In return, understanding the features of empirical data sets and where they deviate from common model assumptions ensures that developers' efforts can be focused in the directions that are the most useful to users. As an example, the original FBD process was designed to handle individual fossil occurrences as the operative units, with a single morphospecies possibly represented by multiple such occurrence units, potentially with uncertainty in the age of each occurrence. However, paleontologists often publish taxon-level age ranges, describing not the age uncertainty of single occurrences but rather providing the first and last appearance dates of an entire morphospecies or genus. In part due to this breakdown in communication, many empirical analyses have used the specimen-level FBD model with taxon-level data (Heath et al. 2014, Gavryushkina et al. 2016). While the specimen-level FBD model is adequate for groups with sparse fossil records (e.g., non-avian dinosaurs, soft-bodied invertebrate groups) for which taxa are frequently represented by single specimens, the sampling assumptions of the specimen-level model are badly violated by fossil-rich clades for which ages are often denoted as stratigraphic ranges, and previous applications of the specimen-based model to such data may have led to less accurate results.

Several avenues are available to improve communication about and understanding of the models. The use of graphicalrepresentations that outline clearly the different components of specific models helps users understand how models are structured and how they relate to each other (Höhna et al. 2016). In addition, the implementation of many models within larger frameworks such as BEAST2 or RevBayes allows for tutorials and teaching resources to be collected together for easier access and discoverability (Barido-Sottani et al. 2018).

3.4.2. Output representation. The representation of the output of FBD inferences is another challenge to consider. As this section has shown, there is considerable uncertainty both in the fossil record and in the selection of analysis setup. One strength of Bayesian analysis is that this uncertainty can be integrated in the inference and can in turn influence the level of uncertainty in the posterior distributions. However, this is helpful only if the uncertainty is then appropriately represented in the final results. This issue is particularly important for FBD inferences, as they typically involve higher levels of uncertainty in the topology of the final tree. Although tools such as DensiTree (Bouckaert 2010) can show partial or full posterior distributions of trees, the representation that is usually chosen is a summary tree, for instance the maximum clade credibility (MCC) tree or maximum a posteriori (MAP) tree. O'Reilly & Donoghue (2018) have shown that both the MCC and the MAP summary methods include large numbers of incorrect clades when summarizing diffuse posterior tree distributions, such as trees inferred from morphological data. They recommend using the majority rule consensus (MRC) tree instead, as it better represents uncertainty. However, MAP and MCC trees remain commonly used in the literature (Azevedo et al. 2021, Chamorro et al. 2021).

4. NEW HORIZONS IN FOSSILIZED BIRTH–DEATH ANALYSES

4.1. Testability

As highlighted in Section 3.2.2, questions remain about how best to assess model fit in such complex models. While this challenge is certainly large, having a mathematically coherent model for divergence time estimation opens the door to simulation-based study of performance (see the sidebar titled Species and the Fossilized Birth–Death Process). In the absence of a single model, node-calibration approaches to dating are difficult to simulate. By contrast, when one can simulate under the analytical model, one can ask sophisticated questions about absolute fit. This also allows approaches such as posterior predictive simulation (Höhna et al. 2018), in which a researcher simulates new data sets seeded from values in the posterior sample. When these data sets are similar to the empirical data, according to the chosen summary statistics, that is an indication of good

SPECIES AND THE FOSSILIZED BIRTH–DEATH PROCESS

Under the original FBD model, as with other birth–death models, lineages give rise to new lineages under a continuous Poisson process with rate λ. Each birth or branching event gives rise to a single new lineage, usually figured in phylogenetics as a single branch splitting or bifurcating into two. This is because the original FBD model treats each occurrence independently, and a phylogenetic analysis using this model equates each operational taxonomic unit to a single occurrence, such that each specimen of a species is treated separately [for a great example of a specimen-based FBD analysis, see Cau (2017)]. Such a model is thus agnostic about the manner in which different instances of a persistent morphological identity can be mapped onto the tree (Silvestro et al. 2018). Each branching event could equally reflect an ancestral morphotaxa splitting into two descendants or a single population of an ancestral taxon becoming its own species and subsequently diverging morphologically, with the descendant coming to coexist with its own morphological ancestor. Paleontologists refer to these two modes as bifurcating speciation and budding speciation respectively (Wagner & Erwin 1995), Foote (1996). To integrate an explicit concept of budding speciation into the original FBD model, we would need to assign one child branch as the ancestor and one child branch as the descendant following each branching event.

The taxon-level FBD process incorporates information about morphospecies through time, and thus must make explicit assumptions about the mode of speciation represented at each branching event. The simplest version of this model assumes a budding process (**Figure 3**), but model extensions also allow for bifurcating and anagenetic speciation modes (Stadler et al. 2018).

model fit. When they are not, this is an indication of poor fit. These methods provide a fairly intuitive way to understand model adequacy and can be used to identify model assumptions that are problematic. For all applications listed below, evaluation via simulation is a necessary step to understand model performance. By providing a coherent mathematical framework (see the sidebar titled Consistency and Coherence), the FBD process ensures that new hypotheses and model extensions can be rigorously tested in the future.

4.2. Beyond Topology and Divergence Times

As described in Section 3.3.1, the fossil record is temporally and spatially incomplete and is determined by a wide range of biological, geological, and socioeconomic factors (Benson et al. 2021, Raja et al. 2022). This causes issues for practically any method that uses temporal evidence from

CONSISTENCY AND COHERENCE

Statistical consistency and statistical coherence are terms that sound similar but mean very different things. Statistical consistency refers to the probability of a model converging to the true value as more data are added. This is a desirable property in that it allows researchers to have confidence that the addition of data leads to correct results. For example, the Mk model for estimating phylogenetic trees from discrete data is statistically consistent (Lewis 2001).

Statistical coherence refers to the quality of information within a data set or across similar data sets.Node calibration is often referred to as incoherent (for example, Heath et al. (2014). What this use of the phrase denotes is not necessarily true statistical incoherence, as defined here. When describing the FBD process, researchers use the term incoherent to denote that there is no single analytical model that describes a node-dating analysis and all the associated data. In consequence, a researcher cannot simulate under the model and test the effect of model violations on the result of the analysis.

the fossil record. Correcting for nonuniform sampling has been a major challenge from the onset of quantitative paleobiology, since the earliest demonstrations of the apparent correlation between global taxonomic diversity and sedimentary rock volume through time (Raup 1972). This affects not only efforts to recover the origin time of clades but also estimates of origination and extinction rates or species richness from fossil occurrence data.

Almost all approaches to mitigating the impact of incomplete or nonuniform sampling in paleobiology involve subsetting or otherwise ignoring certain classes of data. For example, paleontologists often estimate origination and extinction rates from stratigraphic ranges, and several methods have been developed to minimize the impact of sampling variation between intervals, such as per-capita rates (Foote 2000). Per-capita rates are calculated by examining the change in taxon numbers only at interval boundaries and thus consider only those taxa found in more than one interval, effectively ignoring singleton taxa (which are sampled only once). To further remove the effect of sampling heterogeneity across time and space, more recently developed approaches for estimating diversification rates require the subsampling of fossil occurrence data (Alroy 2008, Close et al. 2018, Kocsis et al. 2019). This strategy leads to a reduction in the amount of data actually used to calculate rates or richness.

Better estimates of speciation and extinction rates can be obtained under the FBD range model, in part because this approach can include much more information (Warnock et al. 2020). One key advantage of modeling the sampling process is that incomplete and uneven sampling can be accounted for explicitly. This means that singletons and unsampled lineages are an expected outcome of the process. Another advantage is that the model assumptions are transparent and flexible. Using variants of the FBD process, we can account for rate variation across time and lineages (see Section 3.1). In addition, the FBD process can be applied to problems traditionally seen as the purview of quantitative paleobiology, including estimating richness.

Since origination, extinction, and sampling are core parameters of the FBD model, these parameters can be coestimated along with the topology and divergence times. Information about the model parameters largely comes from the distribution of sampling times (i.e., fossil occurrence data or stratigraphic ranges). In fact, as noted already in Section 3.1, these parameters can be estimated, even without phylogenetic character data, based on the temporal distribution of fossil occurrences or stratigraphic ranges (Stadler et al. 2018, Warnock et al. 2020). Per-interval estimates can be obtained using FBD model extensions that allow for rate variation through time.

The application of the FBD process to fossil age data only is equivalent to how the software PyRate estimates origination and extinction rates (Silvestro et al. 2014). Although the original model implemented in PyRate uses a birth–death process that assumes each species is sampled at least once, the FBD range process has recently been added to this software. Capture-mark-recapture (CMR) is another model-based approach that can be used to obtain per-interval estimates of origination and extinction from fossil occurrence data (Liow & Nichols 2010). CMR methods are based on population ecology models used for monitoring, and numerous extensions allow for nonuniform sampling. The CMR approach used currently in paleobiology does not assume that each taxa is sampled at least once but also does not make use of any additional information about sampling frequency within a given interval (although in theory this could be possible), while both the PyRate and the FBD model approaches allow us to take advantage of information about the number of samples recovered for each interval.

More recently, several approaches have been introduced that allow us to estimate richness. Previous implementations of the FBD model integrated analytically over the number of individuals in the past, i.e., the total number of species or richness through time. Vaughan et al. (2019) introduced a particle filter approach to infer richness under the OBDP model, which uses a phylogeny combined with occurrence data (see Section 3.1). Recently Gupta et al. (2020) and Manceau et al. (2021) introduced a framework to calculate the probability density for the past number of individuals under the OBPD model, which increases the computational efficiency of this approach. This approach has been extended to allow for piecewise constant variation in model parameters and was used recently to infer the past diversity of cetaceans (Andréoletti et al. 2020).

Beyond new and improved methodological approaches, the FBD model has brought the issue of sampled ancestors in the fossil record to wider attention. While theoretical models predict that direct and indirect ancestors are regularly observed in the fossil record (Foote 1996), there was little discussion of considering ancestors for actual fossil taxa outside of nonquantitative systematic treatments (but see Wagner & Erwin 1995). The FBD opens the door toward identifying or at least considering the potential of sampled ancestors in phylogenetic analyses (Bapst et al. 2016, Gavryushkina et al. 2016), especially for densely sampled fossil records, as well as considering how the persistent morphotaxa recognized by paleontologists behave with respect to divergence events (Bapst & Hopkins 2017, Silvestro et al. 2018, Parins-Fukuchi et al. 2019, Parins-Fukuchi 2021, Wright et al. 2021). This question of whether new morphologically distinguishable species arise via bifurcation or budding strikes at how we implicitly perceive and conceptualize evolutionary change in phenotype and carries the potential to change fundamental ideas in macroevolution (Pennell et al. 2014). The long-term evolutionary implications of ancestor–descendant dynamics, such as whether coexisting with your ancestral morphotaxon makes you more likely to go extinct (Pearson 1998), can now be reexamined in a new light.

An additional advantage of applying the FBD model within a hierarchical Bayesian framework is that we can constrain model parameters using existing paleontological or geological knowledge. For example, we can inform our priors using previous estimates of origination, extinction, or sampling rates (O'Reilly & Donoghue 2020, Wright et al. 2021). We could also use sampling proxy data to inform relative changes in rates across time intervals or trait data to inform changes in preservation potential across clades or geographic areas. Since we are often dealing with large and complex uncertainties, we want to include as much available information as possible. Having the explicit model-based framework provided by the FBD process and related models provides many extendable and flexible options that can be tailored to suit different problems not only in phylogenetics but also in quantitative paleobiology and evolutionary biology.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

A.M.W. was funded by National Science Foundation grants DEB-2113425 and NSF-DEB-2045842. The authors would like to thank Dr. Gene Hunt for the invitation to contribute this paper, as well as Emilia Jarachowska and Basanta Khakurel for helpful feedback.

LITERATURE CITED

Alroy J. 2008. Dynamics of origination and extinction in the marine fossil record. *PNAS* 105(Suppl. 1):11536–42

Andréoletti J, Zwaans A, Warnock RCM, Aguirre-Fernández G, Barido-Sottani J, et al. 2020. A skyline birth-death process for inferring the population size from a reconstructed tree with occurrences. bioRxiv 2020.10.27.356758. **https://doi.org/10.1101/2020.10.27.356758**

Ardila NE, Giribet G, Sánchez JA. 2012. A time-calibrated molecular phylogeny of the precious corals: reconciling discrepancies in the taxonomic classification and insights into their evolutionary history. *BMC Evol. Biol.* 12(1):246

Azevedo GHF, Parreiras JS, Bougie T, Michalik P, Wunderlich J, Ramírez MJ. 2021. Fossils constrain biogeographical history in a clade of flattened spiders with transcontinental distribution. *J. Biogeogr.* 48(12):3032–46

Bapst DW. 2013. A stochastic rate-calibrated method for time-scaling phylogenies of fossil taxa. *Methods Ecol. Evol.* 4(8):724–33

Bapst DW. 2014. Assessing the effect of time-scaling methods on phylogeny-based analyses in the fossil record. *Paleobiology* 40(3):331–51

Bapst DW, Hopkins MJ. 2017. Comparing *cal3* and other a posteriori time-scaling approaches in a case study with the pterocephaliid trilobites. *Paleobiology* 43(1):49–67

Bapst DW, Schreiber HA, Carlson SJ. 2018. Combined analysis of extant rhynchonellida (brachiopoda) using morphological and molecular data. *Syst. Biol.* 67(1):32–48

Bapst DW, Wright AM, Matzke NJ, Lloyd GT. 2016. Topology, divergence dates, and macroevolutionary inferences vary between different tip-dating approaches applied to fossil theropods (dinosauria). *Biol. Lett.* 12(7):20160237

Barido-Sottani J, Aguirre-Fernández G, Hopkins MJ, Stadler T, Warnock R. 2019. Ignoring stratigraphic age uncertainty leads to erroneous estimates of species divergence times under the fossilized birth–death process. *Proc. R. Soc. B* 286(1902):20190685

Barido-Sottani J, Bošková V, Plessis LD, Kühnert D, Magnus C, et al. 2018. Taming the BEAST—a community teaching material resource for BEAST 2. *Syst. Biol.* 67(1):170–74

Barido-Sottani J, van Tiel NM, Hopkins MJ, Wright DF, Stadler T, Warnock RC. 2020. Ignoring fossil age uncertainty leads to inaccurate topology and divergence time estimates in time calibrated tree inference. *Front. Ecol. Evol.* 8:183

Beaulieu JM, O'Meara BC. 2021. Fossils do not substantially improve, and may even harm, estimates of diversification rate heterogeneity. bioRxiv 2021.11.06.467550. **https://doi.org/10.1101/2021.11.06.467550**

Benson RB, Butler R, Close RA, Saupe E, Rabosky DL. 2021. Biodiversity across space and time in the fossil record. *Curr. Biol.* 31(19):R1225–36

Benson RB, Mannion PD, Butler RJ, Upchurch P, Goswami A, Evans SE. 2013. Cretaceous tetrapod fossil record sampling and faunal turnover: implications for biogeography and the rise of modern clades. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 372:88–107

Benton MJ, Storrs GW. 1994. Testing the quality of the fossil record: Paleontological knowledge is improving. *Geology* 22(2):111–14

Bodenbender BE, Fisher DC. 2001. Stratocladistic analysis of blastoid phylogeny. *J. Paleontol.* 75(2):351–69

Bouckaert RR. 2010. DensiTree: making sense of sets of phylogenetic trees. *Bioinformatics* 26(10):1372–73

Brusatte SL, Benton MJ, Ruta M, Lloyd GT. 2008. Superiority, competition, and opportunism in the evolutionary radiation of dinosaurs. *Science* 321(5895):1485–88

Cau A. 2017. Specimen-level phylogenetics in paleontology using the fossilized birth-death model with sampled ancestors. *PeerJ* 5:e3055

Černý D, Madzia D, Slater GJ. 2021. Empirical and methodological challenges to the model-based inference of diversification rates in extinct clades. *Syst. Biol.* 71(1):153–71

Chamorro ML, de Medeiros BAS, Farrell BD. 2021. First phylogenetic analysis of dryophthorinae (coleoptera, curculionidae) based on structural alignment of ribosomal DNA reveals cenozoic diversification. *Ecol. Evol.* 11(5):1984–98

Clarke JA, Middleton KM. 2008. Mosaicism, modules, and the evolution of birds: results from a Bayesian approach to the study of morphological evolution using discrete character data. *Syst. Biol.* 57(2):185–201

Close RA, Evers SW, Alroy J, Butler RJ. 2018. How should we estimate diversity in the fossil record? Testing richness estimators using sampling-standardised discovery curves. *Methods Ecol. Evol.* 9(6):1386–400

Darwin C. 1859. *On the Origin of Species by Means of Natural Selection, or, the Preservation of Favoured Races in the Struggle for Life*. London: J. Murray

Drummond AJ, Stadler T. 2016. Bayesian phylogenetic estimation of fossil ages. *Philos. Trans. R. Soc. B* 371(1699):20150129

Fisher D. 1980. The role of stratigraphic data in phylogenetic inference. *Geol. Soc. Am. Abstr. Programs* 12(7):426

Fisher DC. 2008. Stratocladistics: integrating temporal data and character data in phylogenetic inference. *Annu. Rev. Ecol. Evol. Syst.* 39:365–85

Foote M. 1996. On the probability of ancestors in the fossil record. *Paleobiology* 22(2):141–51

Foote M. 2000. Origination and extinction components of taxonomic diversity: general problems. *Paleobiology* 26(S4):74–102

Foote M, Hunter JP, Janis CM, Sepkoski JJ. 1999. Evolutionary and preservational constraints on origins of biologic groups: divergence times of eutherian mammals. *Science* 283(5406):1310–14

Gavryushkina A, Heath TA, Ksepka DT, Stadler T, Welch D, Drummond AJ. 2016. Bayesian total-evidence dating reveals the recent crown radiation of penguins. *Syst. Biol.* 66(1):syw060

Gavryushkina A, Welch D, Stadler T, Drummond AJ. 2014. Bayesian inference of sampled ancestor trees for epidemiology and fossil calibration. *PLOS Comput. Biol.* 10(12):e1003919

Gould SJ. 1980. G.G. Simpson, paleontology, and the modern synthesisIn *The Evolutionary Synthesis*, ed. E Mayr, WB Provine, pp. 153–72. Cambridge, MA: Harvard Univ. Press

Green PJ. 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* 82(4):711–32

Gupta A, Manceau M, Vaughan T, Khammash M, Stadler T. 2020. The probability distribution of the reconstructed phylogenetic tree with occurrence data. *J. Theor. Biol.* 488:110115

Hasegawa M, Kishino H, Yano T. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* 22(2):160–74

Heath TA, Huelsenbeck JP, Stadler T. 2014. The fossilized birth–death process for coherent calibration of divergence-time estimates. *PNAS* 111(29):E2957–66

Höhna S, Coghill LM, Mount GG, Thomson RC, Brown JM. 2018. P3: phylogenetic posterior prediction in RevBayes. *Mol. Biol. Evol.* 35(4):1028–34

Höhna S, Landis MJ, Heath TA, Boussau B, Lartillot N, et al. 2016. RevBayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification language. *Syst. Biol.* 65(4):726–36

Hopkins MJ, Bapst DW, Simpson C, Warnock RC. 2018. The inseparability of sampling and time and its influence on attempts to unify the molecular and fossil records. *Paleobiology* 44(4):561–74

Huelsenbeck JP. 1994. Comparing the stratigraphic record to estimates of phylogeny. *Paleobiology* 20(4):470–83

Hunt G, Carrano MT. 2010. Models and methods for analyzing phenotypic evolution in lineages and clades. *Paleontol. Soc. Papers* 16:245–69

Jukes TH, Cantor CR, et al. 1969. Evolution of protein molecules. *Mamm. Protein Metab.* 3:21–132

Klopfstein S, Ryer R, Coiro M, Spasojevic T. 2019. Mismatch of the morphology model is mostly unproblematic in total-evidence dating: insights from an extensive simulation study. bioRxiv 679084. **https://doi.org/10.1101/679084**

Kocsis AT, Reddin CJ, Alroy J, Kiessling W. 2019. The R package divDyn for quantifying diversity dynamics using fossil sampling data. *Methods Ecol. Evol.* 10(5):735–43

Ksepka DT, Parham JF, Allman JF, Benton MJ, Carrano MT, et al. 2015. The fossil calibration database—a new resource for divergence dating. *Syst. Biol.* 64(5):853–59

Kühnert D, Stadler T, Vaughan TG, Drummond AJ. 2016. Phylodynamics with migration: a computational framework to quantify population structure from genomic data. *Mol. Biol. Evol.* 33(8):2102–16

Laurin M. 2004. The evolution of body size, Cope's rule and the origin of amniotes. *Syst. Biol.* 53(4):594–622

Legried B, Terhorst J. 2021. A class of identifiable phylogenetic birth-death models. bioRxiv 2021.10.04.463015. **https://doi.org/10.1101/2021.10.04.463015**

Lepage T, Bryant D, Philippe H, Lartillot N. 2007. A general comparison of relaxed molecular clock models. *Mol. Biol. Evol.* 24(12):2669–80

Lepage T, Lawi S, Tupper P, Bryant D. 2006. Continuous and tractable models for the variation of evolutionary rates. *Math. Biosci.* 199(2):216–33

Lewis PO. 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. *Syst. Biol.* 50(6):913–25

Liow LH, Nichols JD. 2010. Estimating rates and probabilities of origination and extinction using taxonomic occurrence data: capture-mark-recapture (CMR) approaches. *Paleontol. Soc. Papers* 16:81–94

Lloyd GT, Slater GJ. 2021. A total-group phylogenetic metatree for Cetacea and the importance of fossil data in diversification analyses. *Syst. Biol.* 70(5):922–39

Lloyd GT, Wang SC, Brusatte SL. 2012. Identifying heterogeneity in rates of morphological evolution: discrete character change in the evolution of lungfish (Sarcopterygii;Dipnoi). *Evol.: Int. J. Org. Evol.* 66(2):330–48

Louca S, McLaughlin A, MacPherson A, Joy JB, Pennell MW. 2021. Fundamental identifiability limits in molecular epidemiology. *Mol. Biol. Evol.* 38(9):4010–24

Louca S, Pennell MW. 2020. Extant timetrees are consistent with a myriad of diversification histories. *Nature* 580:502–5

Luo A, Duchêne DA, Zhang C, Zhu CD, Ho SYW. 2020. A simulation-based evaluation of tip-dating under the fossilized birth–death process. *Syst. Biol.* 69(2):325–44

Luo ZX, Yuan CX, Meng QJ, Ji Q. 2011. A Jurassic eutherian mammal and divergence of marsupials and placentals. *Nature* 476(7361):442–45

MacPherson A, Louca S, McLaughlin A, Joy JB, Pennell MW. 2021. A general birth-death-sampling model for epidemiology and macroevolution. bioRxiv 2020.10.10.334383. **https://doi.org/10.1101/2020.10.10.334383**

Manceau M, Gupta A, Vaughan T, Stadler T. 2021. The probability distribution of the ancestral population size conditioned on the reconstructed phylogenetic tree with occurrence data. *J. Theor. Biol.* 509:110400

Marcot JD, Fox DL. 2008. StrataPhy: a new computer program for stratocladistic analysis. *Palaeontol. Electron.* 11(1):11.1.5A

Marshall C. 1995. Stratigraphy, the true order of species originations and extinctions, and testing ancestor-descendant hypotheses among Caribbean neogene bryozoans. In *New Approaches to Speciation in the Fossil Record*, ed. D Erwin, R Anstey, pp. 208–35. New York: Columbia Univ. Press

Marshall CR. 2010. Using confidence intervals to quantify the uncertainty in the end-points of stratigraphic ranges. *Paleontol. Soc. Papers* 16:291–316

Matschiner M. 2019. Selective sampling of species and fossils influences age estimates under the fossilized birth–death model. *Front. Genet.* 10:1064

Matschiner M, Musilová Z, Barth JM, Starostová Z, Salzburger W, et al. 2017. Bayesian phylogenetic estimation of clade ages supports trans-Atlantic dispersal of cichlid fishes. *Syst. Biol.* 66(1):3–22

May MR, Contreras DL, Sundue MA, Nagalingum NS, Looy CV, Rothfels CJ. 2021. Inferring the total-evidence timescale of marattialean fern evolution in the face of model sensitivity. *Syst. Biol.* 70(6):1232–55

May MR, Rothfels CJ. 2021. Mistreating birth-death models as priors in phylogenetic analysis compromises our ability to compare models. bioRxiv 2021.07.12.452074. **https://doi.org/10.1101/2021.07.12.452074**

Monarrez PM, Zimmt JB, Clement AM, Gearty W, Jacisin JJ, et al. 2021. Our past creates our present: a brief overview of racism and colonialism in Western paleontology. *Paleobiology*. **https://doi.org/10.1017/pab.2021.28**

Norell MA. 1992. Taxic origin and temporal diversity: the effect of phylogeny. In *Extinction and phylogeny*, ed. M Novacek, Q Wheeler, pp. 89–118. New York: Columbia Univ. Press

Ogilvie HA, Bouckaert RR, Drummond AJ. 2017. StarBEAST2 brings faster species tree inference and accurate estimates of substitution rates. *Mol. Biol. Evol.* 34(8):2101–14

Ogilvie HA, Mendes FK, Vaughan TG, Matzke NJ, Stadler T, et al. 2021. Novel integrative modeling of molecules and morphology across evolutionary timescales. *Syst. Biol.* 71(1):208–20

O'Reilly JE, Donoghue PCJ. 2018. The efficacy of consensus tree methods for summarizing phylogenetic relationships from a posterior sample of trees estimated from morphological data. *Syst. Biol.* 67(2):354–62

O'Reilly JE, Donoghue PCJ. 2020. The effect of fossil sampling on the estimation of divergence times with the fossilized birth–death process. *Syst. Biol.* 69(1):124–38

Parham JF, Donoghue PCJ, Bell CJ, Calway TD, Head JJ, et al. 2011. Best practices for justifying fossil calibrations. *Syst. Biol.* 61(2):346–59

Parins-Fukuchi C. 2018. Use of continuous traits can improve morphological phylogenetics. *Syst. Biol.* 67(2):328–39

Parins-Fukuchi C. 2021. Morphological and phylogeographic evidence for budding speciation: an example in hominins. *Biol. Lett.* 17(1):20200754

Parins-Fukuchi C, Brown JW. 2017. What drives results in Bayesian morphological clock analyses? bioRxiv 219048. **https://doi.org/10.1101/219048**

Parins-Fukuchi C, Greiner E, MacLatchy LM, Fisher DC. 2019. Phylogeny, ancestors, and anagenesis in the hominin fossil record. *Paleobiology* 45(2):378–93

Patzkowsky ME, Holland SM. 2012. *Stratigraphic Paleobiology*. Chicago: Univ. Chicago Press

Pearson PN. 1998. Speciation and extinction asymmetries in paleontological phylogenies: evidence for evolutionary progress. *Paleobiology* 24(3):305–35

Pennell MW, Harmon LJ, Uyeda JC. 2014. Is there room for punctuated equilibrium in macroevolution? *Trends Ecol. Evol.* 29(1):23–32

Rabosky DL, Goldberg EE. 2015. Model inadequacy and mistaken inferences of trait-dependent speciation. *Syst. Biol.* 64(2):340–55

Raja NB, Dunne EM, Matiwane A, Khan TM, Nätscher PS, et al. 2022. Colonial history and global economics distort our understanding of deep-time biodiversity. *Nat. Ecol. Evol.* 6:145–54

Raup DM. 1972. Taxonomic diversity during the Phanerozoic. *Science* 177(4054):1065–71

Raup DM. 1979. Conflicts between Darwin and paleontology. *Field Mus. Nat. Hist. Bull.* 50(1):22–29

Ronquist F, Lartillot N, Phillips MJ. 2016. Closing the gap between rocks and clocks using total-evidence dating. *Philos. Trans. R. Soc. B* 371(1699):20150136

Ruta M, Wagner PJ, Coates MI. 2006. Evolutionary patterns in early tetrapods. I. Rapid initial diversification followed by decrease in rates of character change. *Proc. R. Soc. B* 273(1598):2107–11

Siddall ME. 1998. Stratigraphic fit to phylogenies: a proposed solution. *Cladistics* 14(2):201–8

Silvestro D, Schnitzler J, Liow LH, Antonelli A, Salamin N. 2014. Bayesian estimation of speciation and extinction from incomplete fossil occurrence data. *Syst. Biol.* 63(3):349–67

Silvestro D, Warnock RC, Gavryushkina A, Stadler T. 2018. Closing the gap between palaeontological and neontological speciation and extinction rate estimates. *Nat. Commun.* 9(1):5237

Smith A. 1998. Is the fossil record adequate? *Nature*. **https://doi.org/10.1038/nature28134**

Smith AB. 1994. *Systematics and the Fossil Record: Documenting Evolutionary Patterns*. Hoboken, NJ: John Wiley & Sons

Smith AB. 2001. Large–scale heterogeneity of the fossil record: implications for phanerozoic biodiversity studies. *Philos. Trans. R. Soc. B* 356(1407):351–67

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

Stadler T, Gavryushkina A, Warnock RC, Drummond AJ, Heath TA. 2018. The fossilized birth-death model for the analysis of stratigraphic range data under different speciation modes. *J. Theor. Biol.* 447:41–55

Tavaré S. 1986. Some probabilistic and statistical problems in the analysis of DNA sequences. *Lect. Math. Life Sci.* 17:57–86

Terry RC, Novak M. 2015. Where does the time go?: Mixing and the depth-dependent distribution of fossil ages. *Geology* 43(6):487–90

Vaughan TG, Leventhal GE, Rasmussen DA, Drummond AJ, Welch D, Stadler T. 2019. Estimating epidemic incidence and prevalence from genomic data. *Mol. Biol. Evol.* 36:1804–16

Vilhena DA, Smith AB. 2013. Spatial bias in the marine fossil record. *PLOS ONE* 8(10):e74470

Wagner PJ. 1995. Stratigraphic tests of cladistic hypotheses. *Paleobiology* 21(2):153–78

Wagner PJ. 1998. A likelihood approach for evaluating estimates of phylogenetic relationships among fossil taxa. *Paleobiology* 24(4):430–49

Wagner PJ, Erwin DH. 1995. Phylogenetic patterns as tests. In *New Approaches to Speciation in the Fossil Record*, ed. D Erwin, R Anstey, pp. 87. New York: Columbia University Press

Warnock RC, Heath TA, Stadler T. 2020. Assessing the impact of incomplete species sampling on estimates of speciation and extinction rates. *Paleobiology* 46(2):137–57

Warnock RC, Wright AM. 2020. *Understanding the Tripartite Approach to Bayesian Divergence Time Estimation*. Cambridge, UK: Cambridge Univ. Press

Warnock RC, Yang Z, Donoghue PC. 2012. Exploring uncertainty in the calibration of the molecular clock. *Biol. Lett.* 8(1):156–59

Wills MA. 1999. Congruence between phylogeny and stratigraphy: randomization tests and the gap excess ratio. *Syst. Biol.* 48(3):559–80

Wright A, Wagner PJ, Wright DF. 2021. *Testing Character Evolution Models in Phylogenetic Paleobiology: A Case Study with Cambrian Echinoderms*. Cambridge, UK: Cambridge Univ. Press

Wright AM, Hillis DM. 2014. Bayesian analysis using a simple likelihood model outperforms parsimony for estimation of phylogeny from discrete morphological data. *PLOS ONE* 9(10):e109210

Zhang C, Stadler T, Klopfstein S, Heath TA, Ronquist F. 2016. Total-evidence dating under the fossilized birth–death process. *Syst. Biol.* 65(2):228–49

Zuckerkandl E, Pauling L. 1965. Evolutionary divergence and convergence in proteins. In *Evolving Genes and Proteins*, ed. V Bryson, HJ Vogel, pp. 97–166. New York: Academic