PALEOBIOLOGY
A PUBLICATION OF THE

Paleontological
S O C I E T Y

Article

Assessing the impact of incomplete species sampling on estimates of speciation and extinction rates

Rachel C. M. Warnock ⓘ *, Tracy A. Heath, and Tanja Stadler*

Abstract.—Estimating speciation and extinction rates is essential for understanding past and present biodiversity, but is challenging given the incompleteness of the rock and fossil records. Interest in this topic has led to a divergent suite of independent methods—paleontological estimates based on sampled stratigraphic ranges and phylogenetic estimates based on the observed branching times in a given phylogeny of living species. The fossilized birth–death (FBD) process is a model that explicitly recognizes that the branching events in a phylogenetic tree and sampled fossils were generated by the same underlying diversification process. A crucial advantage of this model is that it incorporates the possibility that some species may never be sampled. Here, we present an FBD model that estimates tree-wide diversification rates from stratigraphic range data when the underlying phylogeny of the fossil taxa may be unknown. The model can be applied when only occurrence data for taxonomically identified fossils are available, but still accounts for the incomplete phylogenetic structure of the data. We tested this new model using simulations and focused on how inferences are impacted by incomplete fossil recovery. We compared our approach with a phylogenetic model that does not incorporate incomplete species sampling and to three fossil-based alternatives for estimating diversification rates, including the widely implemented boundary-crosser and three-timer methods. The results of our simulations demonstrate that estimates under the FBD model are robust and more accurate than the alternative methods, particularly when fossil data are sparse, as the FBD model incorporates incomplete species sampling explicitly.

Rachel C. M. Warnock. Department of Biosystems Science & Engineering, Eidgenössische Technische Hochschule Zürich, 4058 Basel, Switzerland; Swiss Institute of Bioinformatics (SIB), 1015 Lausanne, Switzerland; and Department of Paleobiology, National Museum of Natural History, Smithsonian Institution, Washington, D.C. 20560, U.S.A. E-mail: rachel.warnock@bsse.ethz.ch

Tracy A. Heath. Department of Ecology, Evolution, and Organismal Biology, Iowa State University, Ames, Iowa, 50011, U.S.A. E-mail: phylo@iastate.edu

Tanja Stadler. Department of Biosystems Science & Engineering, Eidgenössische Technische Hochschule Zürich, 4058 Basel, Switzerland; and Swiss Institute of Bioinformatics (SIB), 1015 Lausanne, Switzerland. E-mail: tanja.stadler@bsse.ethz.ch

Introduction

Speciation and extinction rates are fundamental parameters in macroevolutionary models and integral to hypothesis testing in evolutionary biology. Traditionally, the fossil record was considered the only source of evidence for species origination and extinction in deep time. Paleontological approaches to estimating speciation and extinction (together, diversification) rates use stratigraphic range data, the interval between the first and last appearance of a taxon in the fossil record (Raup 1975; Foote 2000; Alroy 2008, 2014). More recently, methods have been introduced that enable using phylogenies of extant taxa calibrated to time to estimate diversification rates (Nee et al. 1994; Gernhard 2008; Stadler 2009; Morlon et al. 2011), although this approach has been widely criticized for omitting and contradicting evidence provided by the fossil record (Rabosky 2010; Quental and Marshall 2010; Marshall 2017). The incompleteness of the rock and fossil records presents a major challenge for all existing approaches. Only a small proportion of past diversity is preserved in the rock record and recovered by paleontologists, and the fossil-sampling process is heterogenous over time, across space, and among lineages (Raup 1972; Foote and Sepkoski 1999; Smith and McGowan 2007; Holland 2016). Paleontologists have given

considerable thought to the impact of gaps in the fossil record on estimates of diversification rates, as well as ways to deal with incomplete observations (Connolly and Miller 2001; Foote 2001, 2003, 2005; Liow and Nichols 2010). Strategies for mitigating the impact of incomplete sampling often involve filtering the data before estimating parameters of interest but have the obvious drawback of reducing the amount of information actually used to calculate parameter values. For instance, taxa that are only sampled during a single geologic interval (singletons) are typically eliminated before estimating diversification rates because their inclusion has been shown to bias parameter estimates (Alroy 1996; Foote and Raup 1996; Harper 1996).

Phylogenetic approaches, on the other hand, make minimal use of available paleontological evidence, as rates are typically inferred using the phylogeny of extant taxa only (Harvey et al. 1994; Nee et al. 1994; Stadler 2009). Thus, fossil evidence only plays a role in calibrating the phylogeny to time. These methods have also been widely criticized for producing unreasonably low estimates of extinction (Quental and Marshall 2010; Marshall 2017). Although there remains some controversy about the discrepancies observed between diversification rates inferred using phylogenetic versus fossil data, it is broadly accepted that phylogenetic birth–death branching processes give rise to extinct and extant diversity (Marshall 2017). Indeed, birth–death models are the basis of paleontological approaches used to estimate diversification rates from the fossil record (Raup 1985; Foote 2000), as well as important models in analyses of diversification using phylogenies of extant species (Nee et al. 1994).

Promising developments have sought to combine phylogenetic and paleontological approaches, taking advantage of more information in a mechanistic model-based framework. Silvestro et al. (2014b) introduced a birth–death modeling framework that can be applied to estimate diversification rates from stratigraphic range data. This approach has several important advantages over previous paleontological approaches. First, it inherently considers the underlying phylogenetic tree in the estimation of diversification rates. Second, it paves the way for incomplete sampling to be incorporated explicitly. Here, we emphasize the distinction between incomplete or discontinuous *fossil* sampling versus incomplete *species* sampling. Incomplete fossil sampling allows for the possibility that some taxa may only be sampled once. Incomplete species sampling, on the other hand, allows for the possibility that some species may never be sampled. A Bayesian method for inferring diversification rates from paleontological data under a birth–death process has been implemented in the program PyRate (Silvestro et al. 2014a) and has been shown to produce reliable diversification estimates under a range of scenarios (Silvestro et al. 2014b, 2019). One potential drawback of this implementation is that although it incorporates incomplete fossil sampling, the underlying birth–death process model assumes complete species sampling, that is, that each species has been sampled at least once. This is problematic for empirical datasets, many of which are likely to be extremely incomplete and therefore exclude a large proportion of the true species diversity (Foote and Sepkoski 1999; Wagner and Marcot 2013; Soul and Friedman 2017).

Another model for combining phylogenetic and paleontological approaches is the fossilized birth–death (FBD) process (Stadler 2010; Heath et al. 2014), which explicitly accounts for speciation, extinction, and fossil recovery, thus accounting for incomplete fossil and species sampling of the fossil record. Moreover, the FBD model can incorporate extant species sampling as a distinct process, enabling us to take advantage of the information provided by extant diversity. Here, we implement a new FBD model, which we call the *fossilized birth–death range process* (Stadler et al. 2018), for estimating diversification rates from stratigraphic range data in the absence of phylogenetic information. We focus on the scenario in which we have stratigraphic ranges for a given set of species, but no molecular or morphological character data and, therefore, no information about the phylogenetic relationships among species. Applying the FBD model to stratigraphic range data allows us to estimate speciation, extinction, and fossil

recovery rates, while explicitly accounting for incomplete species sampling and uncertainty in the underlying phylogeny. We test the ability of this model to recover estimates of these parameters for the entire tree. We assess the performance of this approach using simulations and compare our method to tree-wide estimates obtained using three fossil-based alternatives, including the boundary-crosser (Foote 2000) and three-timer (Alroy 2008) approaches, and the phylogenetic birth–death model implemented in PyRate (Silvestro et al. 2014b), which does not account for incomplete species sampling (Keiding 1975). Our results show that the FBD model produces reliable estimates of key parameters (including speciation, extinction, and fossil recovery) under a range of conditions. We demonstrate the importance of explicitly accounting for incomplete species sampling and show the benefits of incorporating singletons—including extant samples—which is possible under the mechanistic FBD phylogenetic framework. The results also highlight scenarios in which violations of the assumptions of the sampling model will lead to unreliable parameter estimates. Thus we indicate key areas for further model development and research.

Methods

Approaches to Estimating Tree-Wide Speciation and Extinction Rates from Fossil Occurrences

The Fossilized Birth–Death Range Process.—The FBD range process (illustrated in Fig. 1) was introduced in Stadler et al. (2018). It begins with a single lineage at the origin time x_0 and each lineage has an instantaneous speciation (or branching) rate λ and extinction rate μ, assuming a budding mode of species origination. Fossils are sampled along each lineage with rate ψ and extant species are sampled with probability ρ. We sample n species, of which m are extinct and l are extant. The total number of sampled fossil occurrences is denoted k. A given species i branches at time b_i in the phylogeny of sampled species (i.e., the tree in which all lineages leading to nonsampled species are removed) and goes extinct at time d_i. If species i survives to the
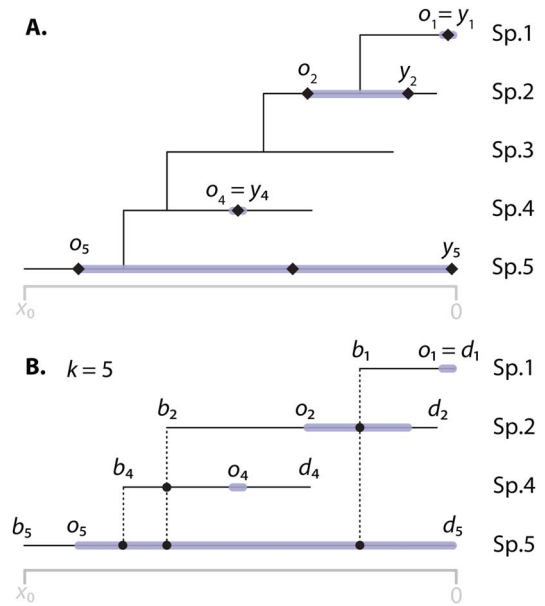

FIGURE 1. The fossilized birth–death (FBD) range process. A, The complete FBD tree, including the observed data. Diamonds represent observed samples of fossils or extant species. Points labeled o_i and y_i are the first and last appearances of range i, respectively. B, The extended sampled stratigraphic ranges, illustrating the information used to calculate the likelihood. b_i and d_i are the species attachment and extinction times, respectively. The attachment time b_i is the time at which species i attaches to other lineages in an incompletely sampled tree, which is not necessarily equivalent to speciation time of species i. For example, the attachment time b_2 for species 2 is not equivalent to the true species origin time shown in A, instead it is the origin time of its nonsampled ancestor, species 3. $b_5 = x_0$ is the origin time. In total we sample $k = 5$ occurrences and $n = 4$ ranges, with $l = 2$ extant and $m = 2$ extinct ranges. The dashed line highlights the number of possible attachment points (black circles) used to calculate γ_i, which accounts for each possible topology, given the value of b_i. γ_i is equivalent to the number of lineages that coexist with species i at time b_i. The exception is the oldest species, where b_i represents the origin and γ_i is always one. The number of possible attachment points for each species is $\gamma_1 = 2$, $\gamma_2 = 2$, $\gamma_4 = 1$, and $\gamma_5 = 1$.

present, we set $d_i = 0$. The first (or oldest) appearance of species i is sampled at time o_i and the last (or youngest) appearance at time y_i. The interval between o_i and y_i represents the stratigraphic range of species i. If a species is sampled only once (i.e., the species is a singleton), then $o_i = y_i$, and if species i is sampled only once at the present, then $o_i = y_i = d_i$. Figure 1A shows an example set of sampled species and how they relate to the true underlying tree. This is referred to as the "complete tree."

To calculate the probability of observing a set of stratigraphic ranges, given the FBD parameters λ, μ, and ψ, we need to know the interval between the branching time b_i and the first appearance o_i, as well as the interval between o_i and extinction time d_i. Note that the interval between b_i and o_i is equivalent to a ghost lineage (Norell et al. 1992). The interval between o_i and d_i is referred to as the "extended sampled stratigraphic range." An example of the extended sampled stratigraphic ranges is shown in Figure 1B. Note that in the extended sampled stratigraphic ranges, the time at which species i attaches to the tree at time b_i may not correspond to the true speciation (origination) time of i in the complete tree, if species sampling is incomplete. Following Stadler et al. (2018), we refer to the branching time b_i as the "attachment time."

In reality, we typically only have information about o_i and y_i and do not know the underlying topology or the true branching times b_i and extinction times d_i. To account for uncertainty in the underlying topology, we need to define γ_i as the total number of coexisting lineages at time b_i, which represents all possible points at which species i can attach to the tree (see Fig. 1B). This value allows us to analytically integrate over all possible tree topologies (see Stadler et al. 2018). To account for uncertainty in the times, we can augment our data such that $b_i > o_i$ and $d_i < y_i$ and marginalize over all possible speciation and extinction times (b_i, d_i) using Markov chain Monte Carlo (MCMC). Summarizing $\mathcal{D} = (k, l, \{b_i, d_i, o_i\}, i = 1, \ldots, n)$, the probability of the augmented extended sampled stratigraphic ranges, as derived in Stadler et al. (2018), is

$$f[\mathcal{D}|\lambda, \mu, \psi, \rho, x_0]$$
$$= \frac{\psi^k \mu^m \rho^l (1-\rho)^{n-m-l}}{\lambda(1-p(x_0))} \prod_{i=1}^n \lambda \gamma_i \frac{\tilde{q}_{asym}(o_i) q(b_i)}{\tilde{q}_{asym}(d_i) q(o_i)}, \quad (1)$$

with

$$p(t) = 1$$
$$+ \frac{-(\lambda - \mu - \psi) + c_1 \frac{e^{-c_1 t}(1-c_2)-(1+c_2)}{e^{-c_1 t}(1-c_2)+(1+c_2)}}{2\lambda}, \quad (2)$$

$$\tilde{q}_{asym}(t) := \sqrt{e^{-t(\lambda+\mu+\psi)} q(t)}, \quad (3)$$

$$q(t) = \frac{4e^{-c_1 t}}{[e^{-c_1 t}(1-c_2)+(1+c_2)]^2}, \quad (4)$$

$$c_1 = \left| \sqrt{(\lambda - \mu - \psi)^2 + 4\lambda\psi} \right|,$$
$$c_2 = -\frac{\lambda - \mu - 2\lambda\rho - \psi}{c_1}. \quad (5)$$

Because the FBD model incorporates the fossil- and extant species-sampling processes explicitly, all observed occurrences, including all extant samples, contribute to the likelihood calculation and may be included in the analysis.

The Birth–Death Sampling Process Assuming Complete Species Sampling.—To examine the impact of incomplete species sampling we compared our approach with a birth–death model based on Keiding (1975),

$$f[\mathcal{D}|\lambda, \mu, \psi] = \lambda^B \mu^D \psi^k e^{-(\lambda+\mu+\psi)S}, \quad (6)$$

where B is the total number of speciation (birth) events, D is the total number of extinction (death) events, and S is the sum of species durations $S = \sum_{i=1}^n b_i - d_i$ (Foote and Miller 2007). We refer to this as the BD model. This model allows for the inclusion of all observed fossil samples but assumes that each species was sampled at least once. However, the inclusion of extant singletons will lead to biased estimates, as modeling incomplete species sampling through time is required to account for the increase in diversity at the present if all extant samples are included. Thus, extant singletons should be excluded from the analysis. Like the FBD range model defined earlier, the BD model also assumes a constant rate of fossil sampling. The BD model described here is similar to the birth–death model implemented in the program PyRate (Silvestro et al. 2014a), with the only difference being the PyRate BD model allows for rate variation across lineages.

Per-Taxon Rates and the Boundary-Crosser Method.—Traditional, summary statistics-based approaches used in paleontology to

estimate speciation and extinction rates are broadly based on quantifying the proportion of first and last appearances sampled during discrete geologic intervals. Following (Foote 2000), fossils sampled during a given interval t_i may be categorized as:

1. **singletons:** taxa confined to the interval, that is, taxa whose first and last appearance are both within the interval t_i (FL) (regardless of the number of occurrences found within that interval);
2. **bottom boundary crossers:** taxa that cross the bottom boundary, that is, the boundary older than interval t_i, and make their last appearance during the interval (bL);
3. **top boundary crossers:** taxa that make their first appearance during the interval and cross the top boundary, that is, the boundary more recent than interval t_i (Ft); and
4. **range-through taxa:** taxa that range through the entire interval, crossing both the top and bottom boundaries (bt).

Examples of each category are shown in Figure 2. The first three categories represent species that must be sampled within interval t_i, while the fourth category (bt) can include taxa that are sampled before and after the interval, but not necessarily within the interval.

The most straightforward approach to estimating speciation and extinction rates for a given interval using this information is to use the proportion of first and last appearances observed during the that interval,

$$\lambda = \left(\frac{N_{FL} + N_{Ft}}{N_{tot}}\right) \times \frac{1}{\Delta t_i},$$
$$\mu = \left(\frac{N_{FL} + N_{bL}}{N_{tot}}\right) \times \frac{1}{\Delta t_i}, \tag{7}$$

where N denotes the number of representatives for a given category and $N_{tot} = N_{FL} + N_{bL} + N_{Ft} + N_{bt}$ and Δt_i is interval length. This equation relies on Δt_i being a very short time interval (i.e., the probability of a single species speciating twice within this interval is negligible). Then, for the speciation process, the probability that a species does not speciate within time interval Δt_i is $\lambda \Delta t_i$, which can be estimated by

$\frac{N_{FL} + N_{Ft}}{N_{tot}}$. This approach is taken from Foote (2000: eqs. 12 and 13) and is an extension of theory presented by Raup (1985). Estimates are defined as "per-taxon origination or extinction rates"; we refer to this as the "per-taxon rates method." This approach makes no attempts to mitigate the effects of incomplete sampling and allows for the inclusion of all fossil occurrences, including singletons. However, because rates are calculated for intervals before the present, extant singletons do not contribute to the estimation of rates (this also applies to the boundary-crosser and three-timer methods described below).

More commonly applied methods in paleontology use strategies to eliminate or minimize biases created by incomplete sampling. The boundary-crosser method (Foote 2000: eqs. 22 and 23) uses the above categories but does not consider singletons (category 1) in the estimation of speciation and extinction rates,

$$\lambda = -ln\left(\frac{N_{bt}}{N_{Ft} + N_{bt}}\right) \times \frac{1}{\Delta t_i},$$
$$\mu = -ln\left(\frac{N_{bt}}{N_{bL} + N_{bt}}\right) \times \frac{1}{\Delta t_i}. \tag{8}$$

The derivation of these equations is straightforward. For example, in the case of the speciation process, the probability that a species does not speciate within time interval Δt_i is $e^{-\lambda \Delta t_i}$, which can be estimated by $\frac{N_{bt}}{N_{Ft} + N_{bt}}$. This approach can only include taxa that have been sampled during more than one interval and thus incorporates fewer data overall.

The Three-Timer Method.—The three-timer method goes further in an effort to account for incomplete sampling by defining an alternative set of categories and applying a sampling correction (Alroy 2008; Alroy et al. 2008). The three-timer categories are defined as:

1. **two-timers:** taxa sampled immediately before and within a bin ($2t_i$) *or* taxa sampled within and immediately after ($2t_{i+1}$);
2. **three-timers:** taxa sampled immediately before, within, and after a bin ($3t_i$); and
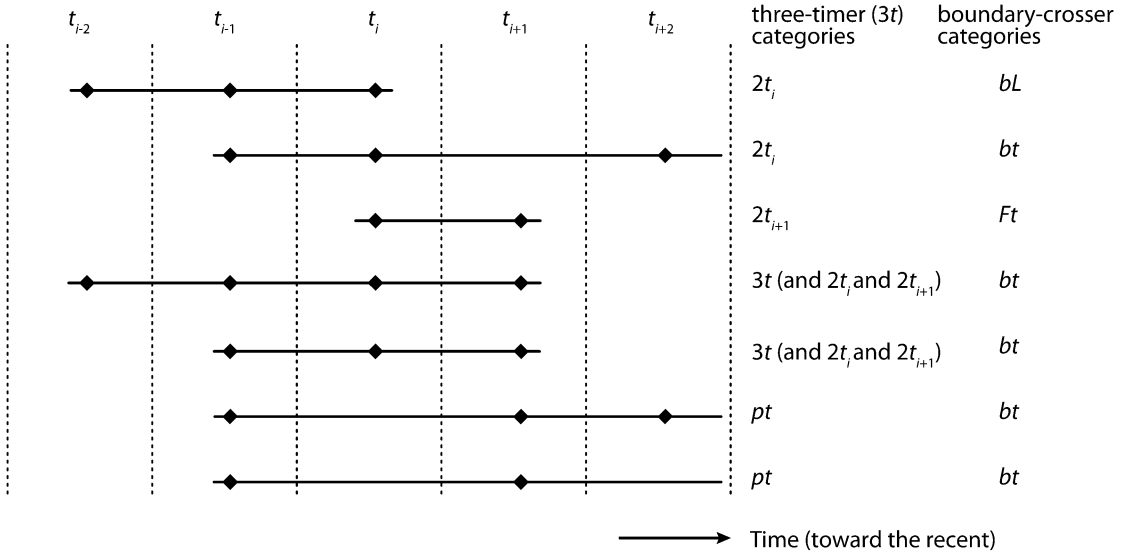3. **part-timers:** taxa sampled immediately before and after but not within the bin

FIGURE 2. Boundary-crosser versus three-timer taxon categories illustrated for a given interval t_i. Boundary-crosser categories: Ft taxa first appear and bL taxa last appear within the interval going forward in time, while the first and last appearance of FL taxa is confined to the same interval and bt taxa appear before and after the interval. The boundary-crosser method uses the bL, Ft, and bt taxa only in the estimation of λ and μ. Three-timer categories: two-timers are sampled immediately before and within bin t_i ($2t_i$) or within and immediately after t_i ($2t_{i+1}$). Three-timers are sampled immediately before, within, and after t_i ($3t$), and part-timers are sampled immediately before and after but not within t_i (pt) (note all three-timers are also two-timers). Singletons (i.e., taxa sampled from a single interval or FL taxa) are excluded before analysis. Image adapted from Alroy (2010, 2014).

(pt_i), implying at least one unsampled lineage.

Notation and definitions follow Alroy (2008, 2010), and examples are shown in Figure 2. The part-timer sampling probability is defined as,

$$P_S = \frac{N_{3t}}{N_{3t} + N_{pt}}, \tag{9}$$

where N_{3t} and N_{pt} may be summed across the entire dataset. The three-timer speciation and extinction rates are defined as

$$\lambda = \left(ln\left(\frac{N_{2t_{i+1}}}{N_{3t_i}}\right) + \ln(P_S) \right) \times \frac{1}{\Delta t_i},$$
$$\mu = \left(ln\left(\frac{N_{2t_i}}{N_{3t_i}}\right) + \ln(P_S) \right) \times \frac{1}{\Delta t_i}. \tag{10}$$

To obtain tree-wide estimates of λ and μ using all three paleontological methods, we summed the number of samples that belong to each taxon category across all intervals, rather than averaging rates calculated separately for each interval. This relies on the assumption that all time intervals (i.e., Δt) are of equal length, which we control for in our simulation design. This choice works in favor of these methods, as it increases the potential information used to calculate rates: some intervals may contain sampled species but do not have sufficient information to recover rate estimates (e.g., an interval may contain boundary crossers but no range-through taxa, or two-timers but no three-timers). We cannot estimate rates for such intervals, but we can use the sampled species in the summation used to calculate tree-wide rates. Note that the sampling parameter P_S used by the three-timer method accounts for incomplete sampling in the preceding (i.e., older) interval in the case of speciation and in the following interval (i.e., younger) in the case of extinction. Because P_S does not apply to extant species sampling, the interval before the present is excluded from tree-wide estimates of extinction rates using this method.

Note that in contrast to the phylogenetic model–based approaches, these three

paleontological methods do not require information about the total number of sampled fossil occurrences (i.e., parameter k). Instead, these methods only require information about whether a given taxon was present or absent during each interval. We used simulated data to assess and compare the performance of alternative approaches to estimating tree-wide speciation and extinction rates.

Simulations

Trees.—Trees were simulated under a constant-rate birth–death process conditioned on the number of extant species using the R package TreeSim (Stadler 2011). We simulated three sets of trees (each 100 replicates) conditioning on extant tips $n = 100$ with three different levels of turnover ($r = \mu/\lambda$)—low ($r = 0.1$), medium ($r = 0.5$), and high ($r = 0.9$)—and $\lambda = 1$ for all three r. Thus, the sets of simulations were parameterized such that $\mu = 0.1$, 0.5, and 0.9, respectively. The simulated origin time of trees ranged from 3.6 to 9.1, 5.2 to 13.7, and 3.1 to 56.2, for low, medium, and high turnover, respectively. We assumed that all speciation events occurred via budding (Wagner and Erwin 1995), such that the speciation and extinction rates obtained from stratigraphic ranges will correspond to λ and μ (Silvestro et al. 2018).

Extant Species Sampling.—The FBD model incorporates extant species sampling as a distinct process, enabling us to include all species sampled at the present, including extant singletons. Because we aim to infer diversification rates and not topology, we can include all extant samples irrespective of available phylogenetic data and fix the probability of extant species sampling $\rho = 1$. To explore the impact of including extant singletons on FBD parameter estimates, we generated two datasets for each replicate, one including all extant samples ($\rho = 1$) and one excluding extant singletons ($\rho = 0$). In the latter case, extant samples are still included for species that have a sampled fossil record. Datasets with $\rho = 0$ only were used for analysis under the BD model, because this model does not allow for the inclusion of extant singletons. For paleontological methods, extant singletons do not contribute any

information, but extant species with a fossil record can be used to categorize taxa that range through to the present. We assumed complete knowledge of extant diversity was available to maximize the amount of information used to calculate rates using these methods.

Fossil Sampling.—Fossils were simulated under a uniform sampling process using the R package FossilSim (Barido-Sottani et al. 2019). For each tree, a single fossil record was generated, and the time period between the origin x_0 and the present ($t = 0$) was divided into 20 equal stratigraphic intervals. We specified a per-interval sampling probability P_S, which was then transformed into a Poisson sampling rate ψ used for the simulation. Specifying P_S was done to keep the number of intervals constant across simulated replicates and to facilitate comparison with approaches that use sampled-in-bin data. The probability of collecting a lineage during each interval was set equal to the uniform sampling probability $P_S = 0.1$, 0.5, and 0.99. In order to simulate fossils, we transform P_S to a sampling rate ψ using $\psi = \frac{-\ln(1 - P_S)}{\Delta t}$, where Δt is interval length. This transformation was obtained because $e^{-\psi \Delta t}$ and $1 - P_S$ both equal the probability of no sample being observed within an interval. Across all tree replicates this approach generated fossils with an average value of $\psi = 0.2$, 1.6, and 10.6. The simulated fossil data were used to generate stratigraphic ranges for analysis using different methods. The total number of fossils recovered for the entire tree was used to specify the BD and FBD model parameter k.

In some empirical cases we may not know the total number of times a species is represented for a particular interval, site, or collection. Instead, we may only know that a given species was sampled during that interval but not the total number of times it was collected. This is the information used by the per-taxon rates and boundary-crosser and three-timer methods. We thus generated fossil datasets to reflect this scenario for analysis under the FBD and BD models. In these datasets, it is only known whether a species was sampled during a given interval but not how many times. We refer to this as sampled-in-bin (1/0) data sampling. The total number of

presences recorded for the entire tree was used to specify the FBD model parameter k. We note that as sampling increases, the specified value of k for a given sampling probability will tend to be lower than the actual number of fossils sampled for a given sampling probability.

The FBD model presented in this study assumes constant fossil recovery over time. To investigate the impact of violating this assumption, we generated simulated fossil records under two simple models of nonuniform preservation: one in which per-interval probability increases linearly toward the present from 0.01 to 0.99, and one in which per-interval probability decreases linearly toward the present from 0.99 to 0.01. We did this for trees with intermediate turnover and simulated a single record under each nonuniform model for each tree replicate. We analyzed these datasets using all five methods for estimating speciation and extinction rates.

Fossil Ages.—Under the Poisson sampling scenario described earlier, in all cases we assumed that the age of a fossil was known precisely. Under the sampled-in-bin strategy, fossil ages were assigned using the midpoint of the stratigraphic interval in which they were sampled. We note this approach can lead to fossil-sampling times that conflict with the true species durations (e.g., $o_i > b_i$ or $y_i < d_i$).

Finally, we simulated datasets for a subset of trees to assess the impact of underestimating the total number of occurrences (i.e., parameter k) and not the combined impact of misspecifying occurrence number and having fossil ages that conflict with species durations. If a given fossil sample represented a species that was only extant for a proportion of a given interval, the fossil age was assigned using the midpoint between when the species first and last appeared within the interval. If speciation occurs within the interval, the beginning of this age range will be younger than the start of the interval. Similarly, if extinction occurs within the interval, the end of this age range will be older than the end of the interval. This ensured that we obtained datasets with $b_i > o_i$ and $y_i > d_i$. We did this for trees with intermediate turnover and simulated a single record for each tree replicate.

Analysis of Simulated Data

Method Implementation.—Given the stratigraphic range data, meaning o_i and y_i for each of the n sampled species, together with the total number of samples k, we estimated b_i and d_i times, the FBD model parameters $\theta = (\lambda, \mu, \psi, x_0)$ and the BD model parameters $\theta = (\lambda, \mu, \psi)$ in a Bayesian framework using MCMC sampling to infer the joint posterior distribution

$$P[\theta|\mathcal{D}] = \frac{P[\mathcal{D}|\theta]P[\theta]}{P[\mathcal{D}]}.$$

Both the FBD and BD models were implemented in the phylogenetic software program DPPDiv (https://github.com/trayc7/DPPDiv; Heath 2012; Heath et al. 2012, 2014).

We specified exponential priors of mean 1.0 on the model parameters λ, μ, and ψ. The probability of extant species sampling ρ was fixed to the its true value, either 1 or 0 depending on the analysis. For analysis under the FBD model a uniform(0, 1000) prior was applied to the origin time. The MCMC sampler was run for 100,000 generations, sampling every 10 steps and discarding 10% as burn-in. Convergence was assessed initially for a small number of replicates using the program Tracer, examining trace plots and ensuring effective sample size (ESS) values >>200 for all parameters for all parameter combinations. For subsequent runs, convergence was monitored for parameters of interest (speciation, extinction and sampling rates) across all replicates and analyses using custom scripts (availability is given below). The per-taxon rates and boundary-crosser and three-timer methods were implemented in the R package fbdR (https://github.com/rachelwarnock/fbdR). For a given replicate, taxon categories were calculated for each interval using the boundary-crosser or three-timer categories, excluding the first interval for which we cannot assign categories. Tree-wide speciation and extinction rates were estimated using the total number of each category across all intervals, meaning we assume the same rates across all intervals. A summary of the analyses applied to each simulated replicate is presented in Table 1.

Performance Measures.—We focus on the ability of alternative methods to estimate speciation, extinction, and sampling rates from

TABLE 1. Overview of data used by different methods for each simulation condition. When the total number occurrences or fossil samples is not used, this refers to scenarios in which only sampled-in-bin data are available. BD, birth–death; FBD, fossilized birth–death.

	Method	Data		
		Fossil singletons	Extant singletons	Total occurrence number
1	FBD model ($\rho = 0$)	Yes	No	Yes
2	FBD model ($\rho = 1$)	Yes	Yes	Yes
3	FBD model ($\rho = 0$)	Yes	No	No
4	FBD model ($\rho = 1$)	Yes	Yes	No
5	BD model	Yes	No	Yes
6	BD model	Yes	No	No
7	Per-taxon rates method	Yes	No	No
8	Boundary-crosser method	No	No	No
9	Three-timer method	No	No	No

stratigraphic ranges. Bayesian approaches to estimating parameters were assessed using percentage error of the median, average width of the 95% credible intervals, and coverage (the proportion of replicates for which the true parameter values falls within the 95% credible intervals). For non-Bayesian approaches, we used the percentage error of point estimates and constructed 95% confidence intervals to approximate the variance using the point estimates obtained across all simulated replicates. Precision was measured using the width of the 95% confidence interval. From these intervals we cannot estimate coverage, but we can compute consistency for each set of simulation conditions. The method is consistent ($= 1$) if the true parameter value falls within the confidence interval or otherwise not ($= 0$). Because different approaches utilize variable amounts of available data (e.g., some methods exclude singletons), we also quantified the average proportion of species, relative to the total sampled species number per replicate that directly contribute to the estimation of speciation and extinction rates under different simulation schemes.

Data Availability.—All code necessary to reproduce the results and figures, along with all input files for analysis using DPPDiv, is available on Dryad (https://doi.org/10.5061/dryad.34tmpg4g9).

Results

The FBD Model Performs Well under a Wide Range of Conditions.—Figure 3 shows the coverage obtained for speciation (λ), extinction (μ), and sampling (ψ) for different turnover and sampling scenarios using the FBD model, assuming the total number of fossil samples is known. The results obtained excluding and including extant singletons are also presented in Supplementary Tables S1 and S2, respectively. Across all scenarios coverage was ≥ 0.91 for λ and μ. Increased fossil sampling led to higher precision and decreased percentage error for both λ and μ. Including extant singletons had the largest impact when turnover was high ($r = 0.9$); in particular, the inclusion of extant singletons improved estimates of P_S across all metrics (e.g., percentage error $= 21\%$ vs. 55%).

Assuming Complete Species Sampling Leads to Biased Parameter Estimates.—Figure 4 shows the coverage obtained using the BD model, which assumes all species are sampled at least once, under the same simulation conditions applied in the previous section (i.e., the total number of samples is known). Further details are also presented in Supplementary Table S3. Overall, coverage was lower and percentage error was higher than the values obtained using the FBD model for the equivalent turnover and sampling scenarios. The worst results were obtained at both lower levels of sampling ($P_S = 0.1$ or $P_S = 0.5$), in particular when coupled with high turnover ($r = 0.9$)—higher turnover results in species with shorter durations and increases the probability that a given species will not be sampled. Thus, as expected, the BD model performs worst when its assumption of complete species sampling is strongly violated. In particular, the BD model leads to erroneous estimates of the
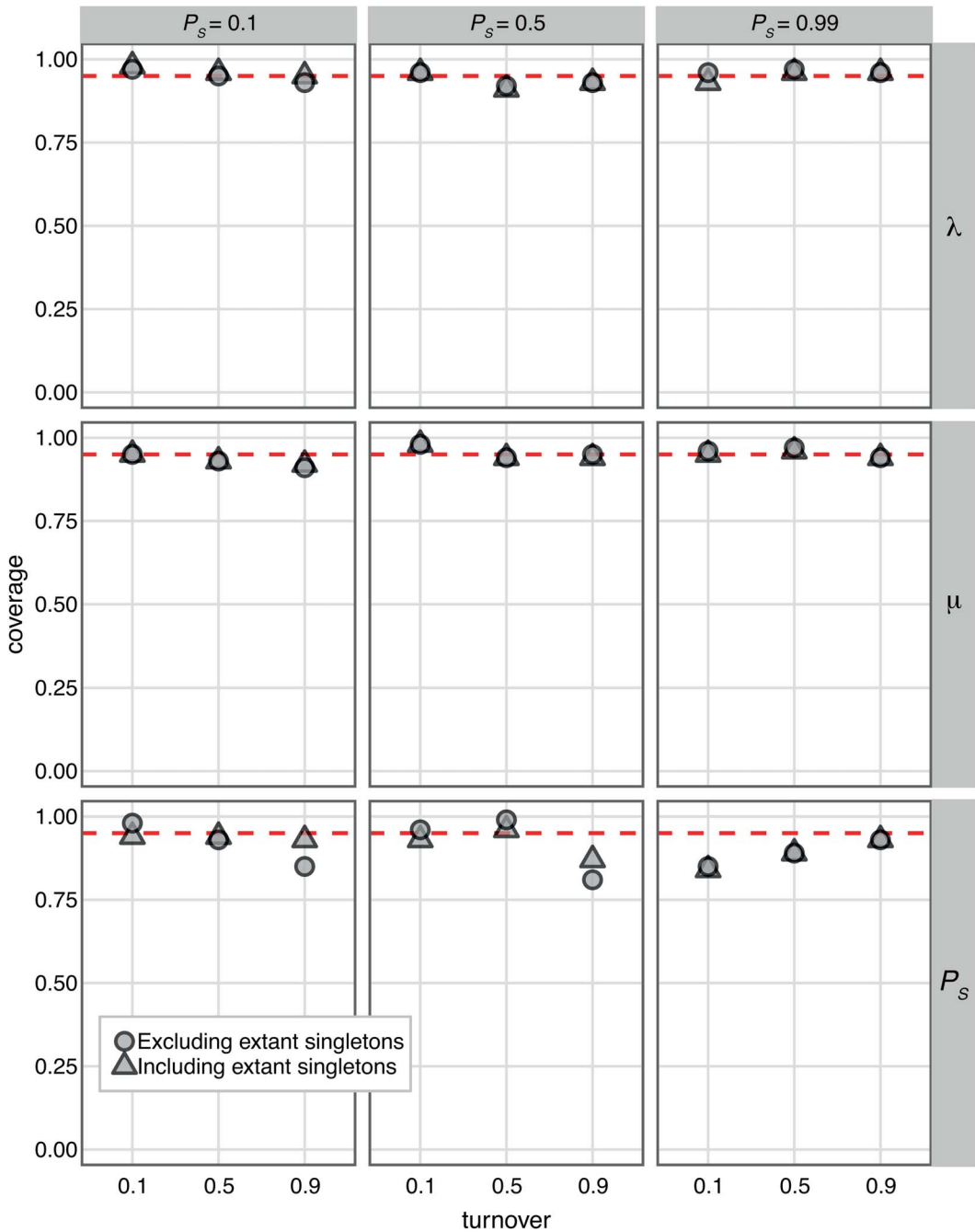
FIGURE 3. Performance of the fossilized birth–death (FBD) model obtained for different turnover and sampling scenarios assuming the number of fossil samples is known. In this set of experiments, the model used for simulation and inference is the same. Results are shown for analyses excluding ($\rho = 0$) and including ($\rho = 1$) extant singletons. Rows show results obtained for speciation (λ), extinction (μ), and sampling (P_S). Columns show results obtained at different sampling levels (per-interval sampling probability $P_S = 0.1$, 0.5, or 0.99). In each box, the x-axis represents turnover and the y-axis represents coverage (the proportion simulation replicates out of 100 that contain the true value). The dashed line highlights the 0.95 coverage level.
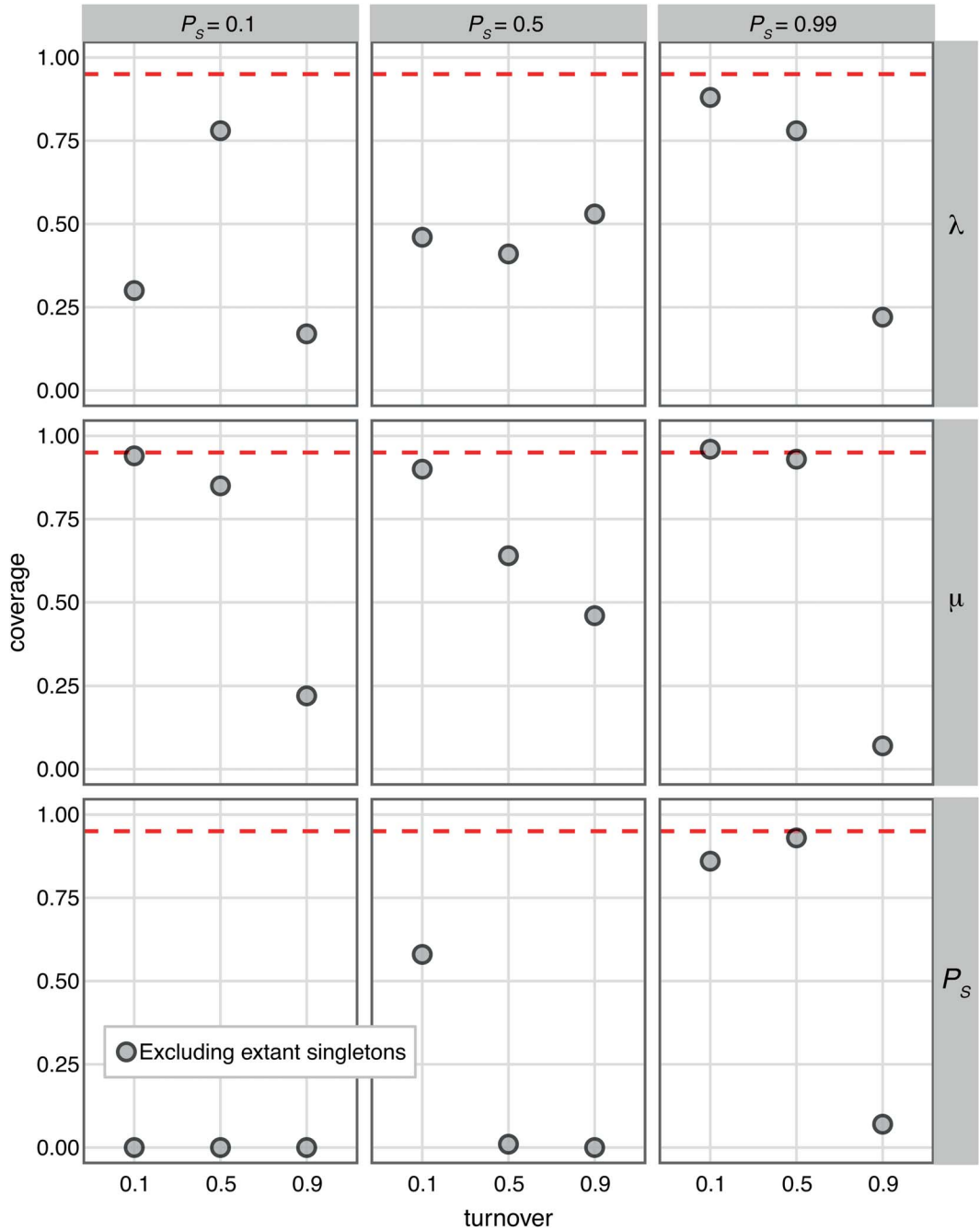
FIGURE 4. Performance of the birth–death (BD) model obtained for different turnover and sampling scenarios assuming the number of fossil samples is known. This model assumes that each species has been sampled at least once. Analysis using this model always excludes extant singletons ($\varrho = 0$). Rows show results obtained for speciation (λ), extinction (μ), and sampling (P_S). Columns show results obtained at different sampling levels (per-interval sampling probability $P_S =$ 0.1, 0.5, or 0.99). In each box, the x-axis represents turnover and the y-axis represents coverage (the proportion simulation replicates out of 100 that contain the true value). The dashed line highlights the 0.95 coverage level.

sampling parameter when sampling is low (P_S = 0.1, percentage error ≫100). As fossil sampling increased, estimates of λ, μ, and P_S overall improved, in terms of percentage error, for all turnover scenarios (Supplementary Table S3).

Performance of the FBD Model Decreases when the Fossil-Sampling Process Assumptions Are Violated.—Figure 5 shows the coverage obtained for λ, μ, and P_S, using the FBD model, assuming we have sampled-in-bin data but the total number of fossil samples in not known (i.e., we only know whether a given species was sampled during each interval but not how many times), which violates the model assumptions. The results obtained excluding and including extant singletons are also presented in Supplementary Tables S4 and S5, respectively. These scenarios led to poorer performance of the FBD model in terms of coverage and percentage error, as expected. In particular, P_S was underestimated and became worse with increased fossil sampling. This is because the number of observations used to inform the model parameter k increasingly deviates from the true number of samples collected. For intermediate and high sampling (P_S = 0.5 or 0.99) coverage was zero for this parameter. This led λ and μ to be underestimated, and the worst results were obtained when turnover was also high (r = 0.9, percentage error of around 20% for λ and μ; Supplementary Table S4). Similar results were obtained using simulated datasets in which we ensured there was no conflict between the fossil ages and true species durations, demonstrating that these patterns are largely driven by misspecification of occurrence number (Supplementary Table S6). Similar effects are also observed for the BD model (i.e., performance decreased with increased fossil sampling given sampled-in-bin data; Supplementary Table S7, Supplementary Fig. S1).

Figure 6 shows the coverage obtained for λ and μ under medium turnover and two nonuniform sampling scenarios, in which sampling either increased or decreased toward the present. The results obtained excluding and including extant singletons are also presented in Supplementary Tables S8 and S9, respectively. Violating the assumption of constant fossil recovery led to overall poorer performance of the FBD model, as expected, although

including extant singletons had a positive impact on the results under these simulation conditions. Nonuniform fossil recovery also led to poorer performance of the BD model (Supplementary Table S10, Supplementary Fig. S2).

The FBD Model Performs Well in Comparison to Alternative Approaches, Even When the Model Is Violated.—Figure 7 shows the percentage error obtained for λ and μ using five alternative methods across different turnover and sampling scenarios, assuming the total number of fossil and extant samples is known. Results obtained using the per-taxon rates and boundary-crosser and three-timer methods are also presented in Supplementary Tables S11–S13. Overall, with regard to percentage error, the FBD model is the best model for medium and high turnover, while the BD model (with the FBD model producing similar results) is the best model under low turnover. The per-taxon rates method, which takes no steps to account for incomplete sampling, is the worst method when sampling proportion is high. This is in contrast to the boundary-crosser and three-timer methods, which produced the highest errors at the lowest levels of sampling, meaning the performance of these methods improves with the addition of more data. The difference in performance observed for high turnover between these methods and the FBD model is linked to the proportion of singleton taxa associated with these datasets. At high turnover, taxa tend to have shorter durations, and an effect of our simulation design (keeping interval number constant across trees) is that these datasets also have longer on-average time bins, resulting in datasets with a larger proportion of singletons. Because singletons are excluded by the boundary-crosser and three-timer methods, this reduces the information available to estimate rates using these methods.

At the lowest sampling level (P_S = 0.1), the results of the per-taxon rates method were similar to those obtained using the FBD model in terms of percentage error and precision (Supplementary Tables S1, S11). The boundary-crosser method produced similar results to the FBD model at the highest levels of sampling in terms of percentage error and precision, but
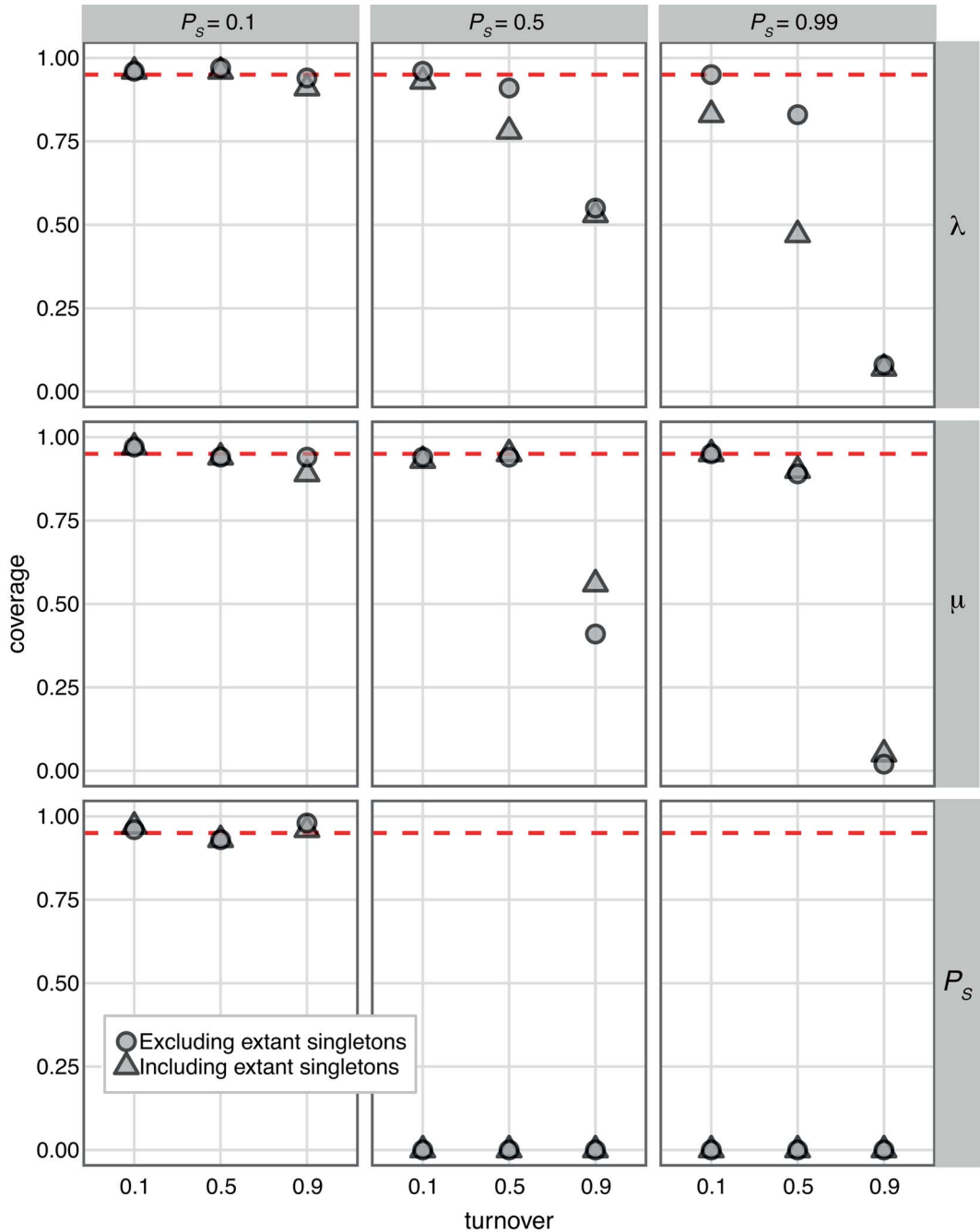
FIGURE 5. Performance of the fossilized birth–death (FBD) model obtained for different turnover and sampling scenarios assuming only sampled-in-bin data are available. In this set of experiments, we use sampled-in-bin data, which is a violation of the model used during inference. Results are shown for analyses excluding ($\rho = 0$) and including ($\rho = 1$) extant singletons. Rows show results obtained for speciation (λ), extinction (μ), and sampling (P_S). Columns show results obtained at different sampling levels (per-interval sampling probability $P_S = 0.1$, 0.5, or 0.99). In each box, the x-axis represents turnover and the y-axis represents coverage (the proportion simulation replicates out of 100 that contain the true value). The dashed line highlights the 0.95 coverage level.
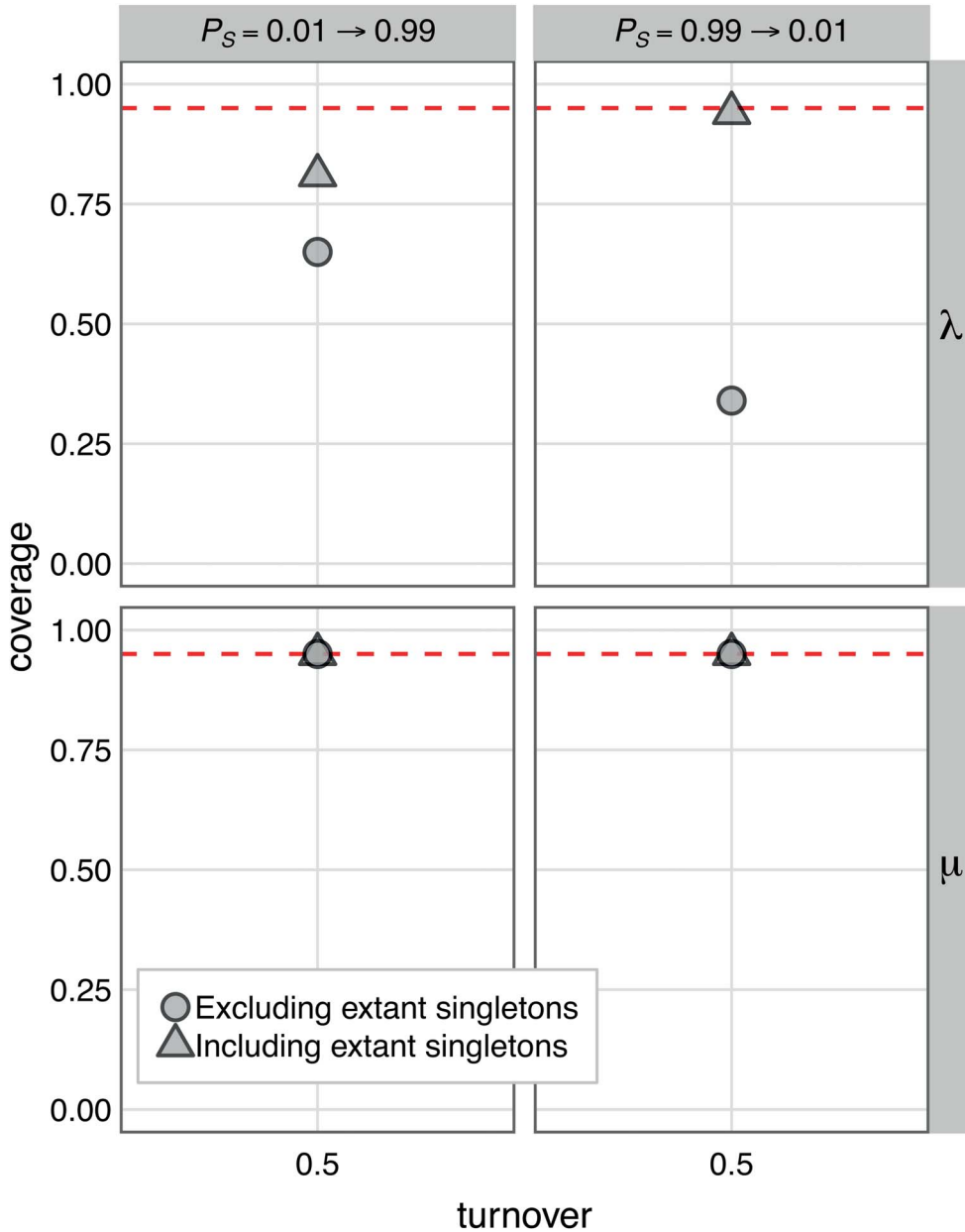
FIGURE 6.   Performance of the fossilized birth–death (FBD) model obtained under nonuniform fossil recovery assuming the number of fossil samples is known. In this set of experiments, the model used for simulation violates the model used for inference. Results are shown for analyses excluding ($\rho = 0$) and including ($\rho = 1$) extant singletons. Rows show results obtained for speciation (λ) and extinction (μ). Columns show results obtained under different nonuniform sampling scenarios, where sampling increases ($0.01 \rightarrow 0.99$) or decreases ($0.99 \rightarrow 0.01$) linearly toward the present. In each box, the x-axis represents turnover and the y-axis represents coverage (the proportion simulation replicates out of 100 that contain the true value). The dashed line highlights the 0.95 coverage level.

differences were observed at the lowest sampling levels. The three-timer method also produced similar results to the FBD and boundary-crosser methods at the highest sampling level ($P_S = 0.99$; Supplementary Table S13). When turnover was high ($r = 0.9$) and sampling was low ($P_S = 0.1$), the BD model produced the worst estimates (Fig. 7).
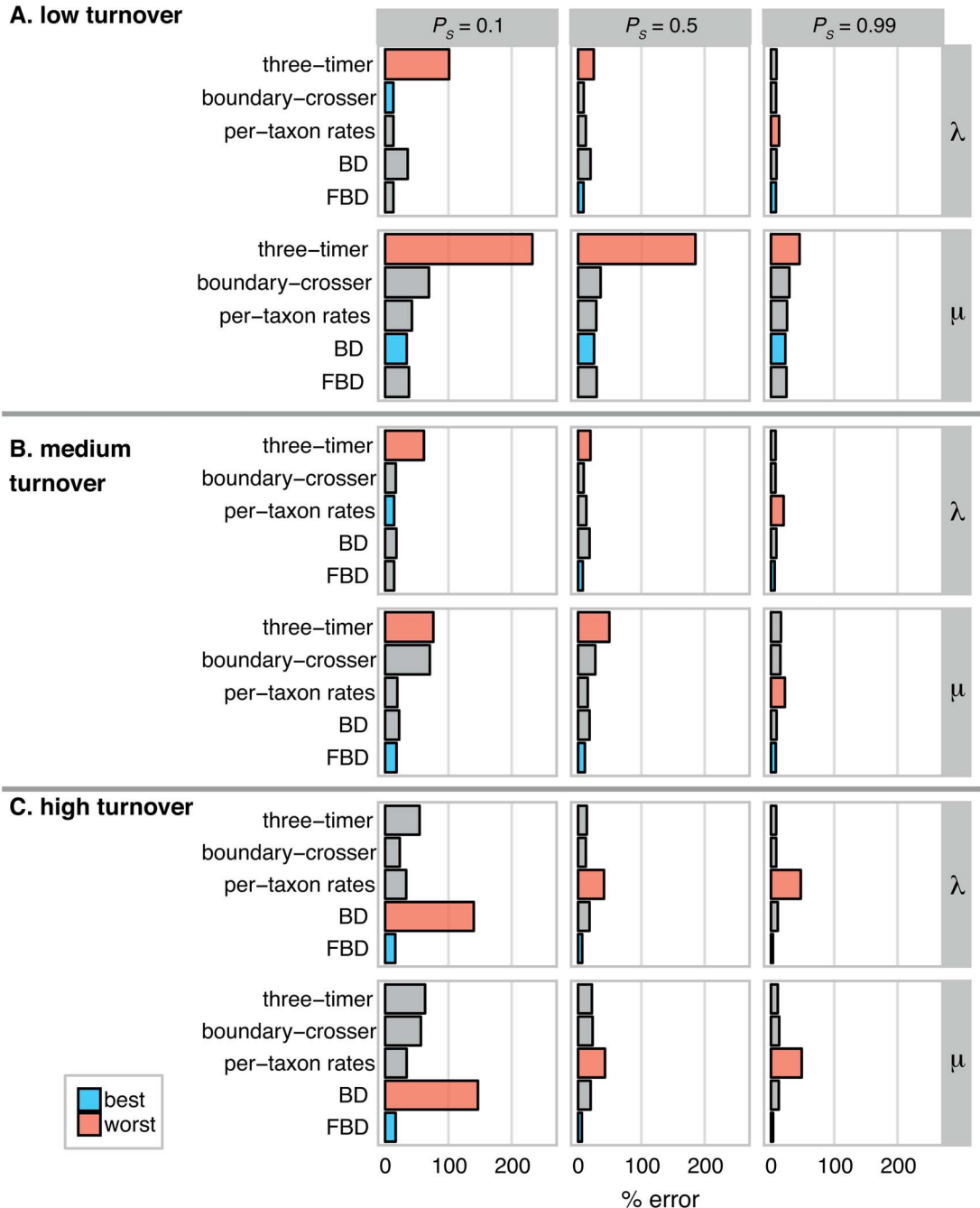
FIGURE 7. Comparison between different approaches to estimating speciation (λ) and extinction (μ) rates assuming the number of fossil samples is known. Results are shown for low (A, $r = 0.1$), medium (B, $r = 0.5$), and high (C, $r = 0.9$) turnover. Each column shows results obtained at different sampling probabilities (per-interval sampling probability $P_S = 0.1$, 0.5, or 0.99). For each box, the y-axis represents the percentage error of the median estimate (birth–death [BD] and fossilized birth–death [FBD] approaches) or the point estimate (all other approaches) averaged across 100 simulation replicates. The best (lowest) and worst (highest) percentage error are highlighted for each turnover and sampling scenario. Analysis using the FBD model includes extant singletons ($\rho = 1$). Results obtained excluding extant singletons are shown in Supplementary Fig. S3.

In contrast to the phylogenetic models, the paleontological methods considered here (the per-taxon rates and boundary-crosser and three-timer methods) use only sampled-in-bin data. Figure 8 shows the percentage error obtained for λ and μ assuming sampled-in-bin data only also for the FBD and BD methods (i.e., we violate the assumptions made about the fossil-sampling process by these models). These results show that even when the FBD model is violated, it still recovered parameter estimates that are closer to their true values than alternative methods when turnover and/or sampling is low. As sampling and turnover increases, the boundary-crosser method overall recovers the best estimates given sampled-in-bin data. Similar results are obtained when knowledge of extant diversity is not known (see Supplementary Figs. S3, S4) or when the assumption of uniform fossil recovery is violated (see Supplementary Fig. S5, Supplementary Tables S14–S16).

The FBD Model Uses a Higher Proportion of Sampled Species in the Estimation of Rates Than Alternative Methods.—Figure 9 and Supplementary Table S17 present the proportion of the total number of sampled species that are directly included in the estimation of speciation and extinction rates using different methods. As sampling increases, the proportion of sampled species included in the estimation of rates increases for all methods, but some methods always excluded more sampled species. Overall, the FBD model including extant singletons always used the largest proportion of sampled species, followed by the FBD and BD models excluding extant singletons and the per-taxon rates method. Fewer sampled species were used by the boundary-crosser method, and the least were used by the three-timer method. These two latter methods also showed high percentage errors under low-medium turnover (see Figs. 6, 7). However, more data do not necessarily mean higher accuracy, for example, if there are underlying model violations. For example, the per-taxon rates method includes singletons in the estimation of rates at the expense of accuracy, leading to poor performance at higher sampling probabilities compared with other approaches (see Figs. 6, 7).

Discussion

We scrutinized the performance of a new model—the FBD range process (Stadler et al. 2018)—for recovering tree-wide estimates of speciation and extinction rates from stratigraphic ranges using simulated data and compared it with alternative approaches. Our results demonstrate the advantages and importance of explicitly modeling incomplete fossil and species sampling. Additionally, our simulations highlight critical areas for further model development.

Recent simulation studies have demonstrated that estimates of diversification metrics become challenging at low levels of fossil recovery (Soul and Friedman 2017; Smiley 2018). The sampling levels explored in our simulations reflect a broad range of scenarios. The lowest level (P_S = 0.1, $E[\psi]$ = 0.2) may correspond to the scenario among some poorly sampled terrestrial groups (Wagner and Marcot 2013), while the highest level (P_S = 0.99, $E[\psi]$ = 10) may correspond to the scenario among some marine invertebrate groups densely sampled from a single locality (Bapst and Hopkins 2017).

The FBD model has a clear advantage when sampling is low, which will be the case for many groups of interest (Foote and Sepkoski 1999). In this case, the model's ability to take into account incomplete species sampling and use all available data, meaning all extinct and extant singletons, is particularly advantageous. While the phylogenetic BD model, which is equivalent to the birth–death process assumed by the program PyRate, can also include all extinct singletons, it assumes complete species sampling, leading to high errors at low sampling proportions and high turnover. We note that based on our paper, PyRate has been extended to include an implementation of the FBD process presented here (PyRate commit b171ee1; Silvestro et al. 2014a). Analysis under the FBD process is the only available approach that can take advantage of extant singletons and may be valuable for clades in which extant diversity is high relative to the number of sampled fossils, such as insects (Labandeira 2005).

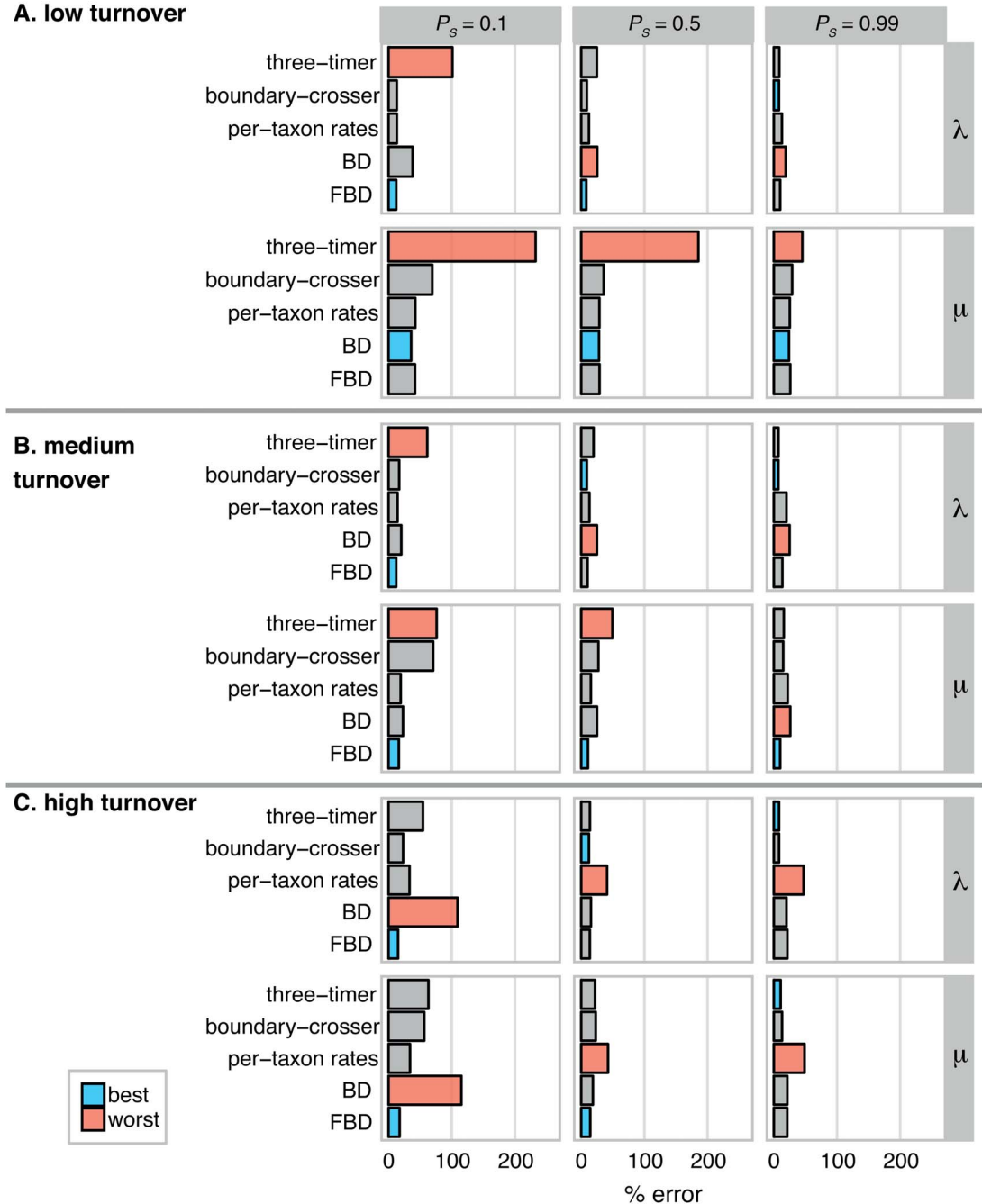Perhaps surprisingly, the per-taxon rates method—the only method we examined that

FIGURE 8. Comparison between different approaches to estimating speciation (λ) and extinction (μ) rates assuming only sampled-in-bin data are available. Results are shown for low (A, $r = 0.1$), medium (B, $r = 0.5$), and high (C, $r = 0.9$) turnover. Each column shows results obtained at different levels (per-interval sampling probability $P_S = 0.1$, 0.5, or 0.99). For each box, the y-axis represents the percentage error of the median estimate (birth–death [BD] and fossilized birth–death [FBD] approaches) or the point estimate (all other approaches) averaged across 100 simulation replicates. The best (lowest) and worst (highest) percentage error are highlighted for each turnover and sampling scenario. Analysis using the FBD model includes extant singletons ($\rho = 1$). Results obtained excluding extant singletons are shown in Supplementary Fig. S4.
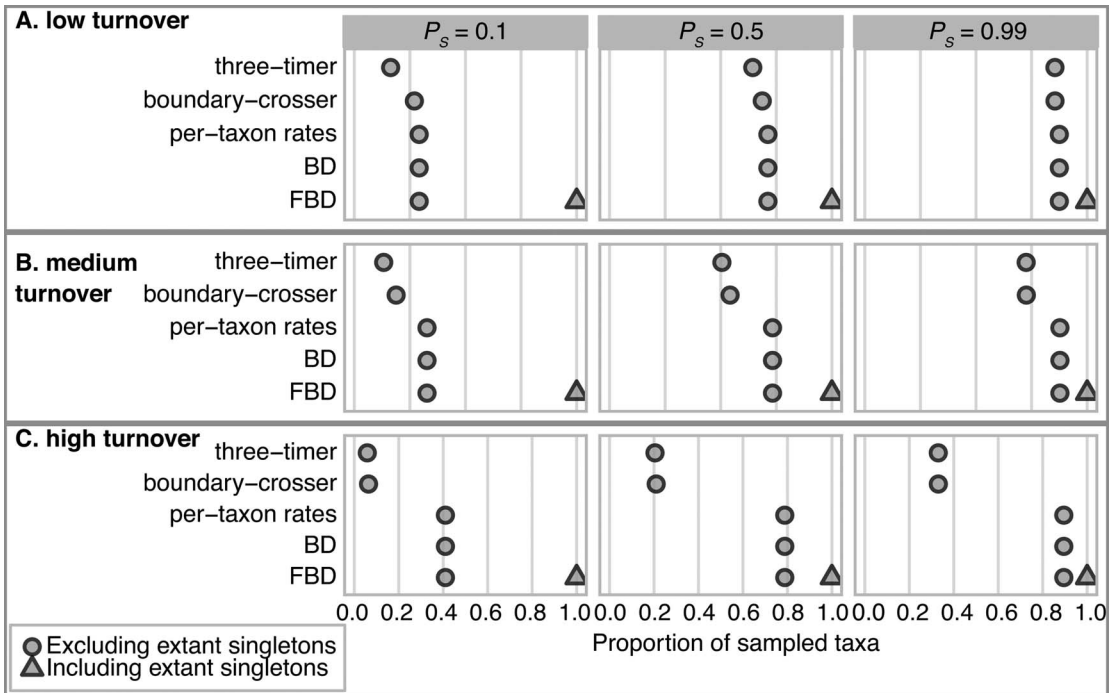
FIGURE 9. Comparison between the proportion of sampled taxa incorporated into rate estimation using different methods. BD, birth–death; FBD, fossilized birth–death. Relts are shown for low (A, $r = 0.1$), medium (B, $r = 0.5$), and high (C, $r = 0.9$) turnover. Each column shows results obtained at different sampling probabilities (per-interval sampling probability P_S = 0.1, 0.5, or 0.99). For each box, the y-axis represents the proportion of the total number of species included in the analysis averaged across 100 replicates. For the FBD analysis, results are shown for analyses excluding ($\rho = 1$) and including ($\rho = 0$) extant singletons.

does nothing to mitigate the potential impact of incomplete sampling—does well across a wide range of scenarios. However, this method is statistically inconsistent, as its performance decreases with increased sampling. Although all other methods converge on similar estimates of speciation and extinction as fossil sampling increases, the FBD model (assuming the underlying sampling process is correct) still often outperforms alternative methods in terms of percentage error and precision (Fig. 7). In addition, the BD and FBD approaches are probabilistic models that have been implemented within a Bayesian framework. This has the advantage of producing interpretable measures of uncertainty (the Bayesian credible interval) that reflect underlying statistical uncertainty in parameter estimates.

We also present scenarios in which the performance of the phylogenetic approaches is expected to decrease, namely, when the Poisson sampling process assumed by the model

is violated. In particular, we show that having sampled-in-bin data rather than information about the total number of occurrences (i.e., we know whether each taxon was sampled or not during each interval but not how many times) has a detrimental impact on the performance of the BD and FBD models, especially when sampling is high (Fig. 8). In theory it is possible to marginalize over the number of fossils within a given stratigraphic range or stratigraphic interval (Stadler et al. 2018), which would be useful in accounting for uncertainty in both per-interval specimen number and fossil ages. Further work is required to ascertain the performance of these model variations and their impact on estimates of speciation, extinction, and sampling. Nonuniform sampling and temporal variation in diversification are also important factors worth consideration. In particular, temporal variation in fossil recovery is a predominant feature of the paleontological record, and we show that performance

of the FBD model decreases when we violate this assumption. Accounting for temporal variation in sampling can be achieved using birth–death skyline models (Stadler et al. 2013), and previous applications of FBD skyline models have been shown to recover reliable parameter estimates when rates vary over time (Gavryushkina et al. 2014; Zhang et al. 2015). Smiley (2018) recently demonstrated that temporal variation in sampling can lead to spurious estimates of diversification patterns using traditional paleontological methods, including the boundary-crosser and three-timer methods. Further, Silvestro et al. (2019) showed that a phylogenetic approach that allows the timing of diversification shifts to be coestimated along with other model parameters can recover more reliable estimates applied to these same scenarios. We anticipate the application of the FBD range model accounting for temporal rate variation would lead to further benefits when sampling is both low and nonuniform and the goal is to recover per-interval rather than tree-wide estimates of speciation and extinction.

The methods we compared our approach with are typically used to recover per-interval estimates of speciation and extinction, rather than tree-wide estimates. For example, the program PyRate combines nonuniform fossil recovery with the BD model, while we implement a version of this model that assumes constant fossil recovery. Our analyses allowed us to explore the impact of incomplete species sampling without the confounding effects of temporal rate variation. However, our finding that incomplete species sampling decreases the performance of this model will remain relevant even under more complex sampling scenarios that incorporate rate heterogeneity. Similarly, our finding that phylogenetic model–based approaches can incorporate more data than alternative methods will still apply when we relax the assumption of constant fossil recovery.

Nonuniform fossil recovery within lineages will be less straightforward to incorporate into the constant-rate FBD range model presented here. In this respect, a birth–death model that assumes complete species sampling has some advantages over the FBD model. For example, the program PyRate incorporates a sampling model that recognizes species may be rare at the beginning and end of species ranges and peak somewhere in the middle, which has been demonstrated for some empirical datasets (Liow and Stenseth 2007). Such nonuniform sampling models require making the assumption that species sampling is complete. In particular, these models assume that the speciation times are known (or can be estimated) for all species. Because the attachment time of a given range (b_i) (i.e., the branching time at which the species attaches to other lineages in an incompletely sampled FBD tree) is not necessarily equivalent to the speciation time of species i (Fig. 1), it will be challenging to incorporate this nonuniform sampling process within an incomplete species-sampling framework. The BD model implemented in PyRate may therefore be more appropriate than the FBD model when (approximately) complete species sampling can be assumed. We note that when analyses under the BD and FBD models (assuming a uniform sampling process) produce very different rate estimates, this is a potential indicator that species sampling is incomplete.

The FBD range model presented here assumes a budding speciation process. Although there is increasing recognition that alternative speciation modes (e.g., anagenesis or bifurcating cladogenesis) play an important role in diversification (Ezard et al. 2012; Bapst 2013; Silvestro et al. 2018), different modes cannot easily be incorporated into this version of the FBD model. This is because we do not sample the tree and instead use the number of coexisting lineages to integrate out analytically all possible trees. This analytic integration assumes budding speciation (Stadler et al. 2018). Presently, if stratigraphic range datasets include taxa that evolved via anagenesis or bifurcation, then estimated rates should be closer to the rates predicted by the birth–death chronospecies model (Silvestro et al. 2018). Further extensions of the FBD range model that incorporate other speciation modes and allow us to sample the tree topology will make it possible to more directly explore the relative contributions of different speciation modes to evolution (Stadler et al. 2018).

Conclusions

In this study we focused on demonstrating the impact of incomplete fossil and species data on speciation and extinction rates estimated from stratigraphic ranges. We show that a process-based approach to accommodating incomplete sampling allows for the inclusion of more data and leads to more robust inference. Further work focusing on key aspects of paleontological databases (e.g., sampled-in-bin data, temporal variation in fossil recovery) will expand the applicability of phylogenetic models to the study of macroevolution.

Acknowledgments

Literature Cited

Alroy, J. 1996. Constant extinction, constrained diversification, and uncoordinated stasis in North American mammals. Palaeogeography, Palaeoclimatology, Palaeoecology 127:285–311.

Alroy, J. 2008. Dynamics of origination and extinction in the marine fossil record. Proceedings of the National Academy of Sciences USA 105:11536–11542.

Alroy, J. 2010. Fair sampling of taxonomic richness and unbiased estimation of origination and extinction rates. *In* John Alroy and Gene Hunt, eds. Quantitative methods in paleobiology. Paleontological Society Papers 16:55–80.

Alroy, J. 2014. Accurate and precise estimates of origination and extinction rates. Paleobiology 40:374–397.

Alroy, J., M. Aberhan, D. J. Bottjer, *et al*. (35 coauthors). 2008. Phanerozoic trends in the global diversity of marine invertebrates. Science 321:97–100.

Bapst, D. W. 2013. When can clades be potentially resolved with morphology? PLoS ONE 8:e62312.

Bapst, D. W., and M. J. Hopkins. 2017. Comparing *cal*3 and other a posteriori time-scaling approaches in a case study with the pterocephaliid trilobites. Paleobiology 43:49–67.

Barido-Sottani, J., W. Pett, J. E. O'Reilly, and R. C. M. Warnock. 2019. FossilSim: an R package for simulating fossil occurrence data under mechanistic models of preservation and recovery. Methods in Ecology and Evolution 10:835–840.

Connolly, S. R., and A. I. Miller. 2001. Joint estimation of sampling and turnover rates from fossil databases: capture-mark-recapture methods revisited. Paleobiology 27:751–767.

Ezard, T. H., P. N. Pearson, T. Aze, and A. Purvis. 2012. The meaning of birth and death (in macroevolutionary birth–death models). Biology Letters 8:139–142.

Foote, M. 2000. Origination and extinction components of taxonomic diversity: general problems. Paleobiology 26:74–102.

Foote, M. 2001. Inferring temporal patterns of preservation, origination, and extinction from taxonomic survivorship analysis. Paleobiology 27:602–630.

Foote, M. 2003. Origination and extinction through the Phanerozoic: a new approach. Journal of Geology 111:125–148.

Foote, M. 2005. Pulsed origination and extinction in the marine realm. Paleobiology 31:6–20.

Foote, M., and A. I. Miller. 2007. Principles of paleontology. Freeman, New York.

Foote, M., and D. M. Raup. 1996. Fossil preservation and the stratigraphic ranges of taxa. Paleobiology 22:121–140.

Foote, M., and J. J. Sepkoski. 1999. Absolute measures of the completeness of the fossil record. Nature 398:415.

Gavryushkina, A., D. Welch, T. Stadler, and A. J. Drummond. 2014. Bayesian inference of sampled ancestor trees for epidemiology and fossil calibration. PLoS Computational Biology 10: e1003919.

Gernhard, T. 2008. The conditioned reconstructed process. Journal of Theoretical Biology 253:769–778.

Harper, C. W. 1996. Patterns of diversity, extinction and origination in the Ordovician–Devonian stropheodontacea. Historical Biology 11:267–288.

Harvey, P. H., R. M. May, and S. Nee. 1994. Phylogenies without fossils. Evolution 48:523–529.

Heath, T. A. 2012. A hierarchical Bayesian model for calibrating estimates of species divergence times. Systematic Biology 61:793–809.

Heath, T. A., M. T. Holder, and J. P. Huelsenbeck. 2012. A Dirichlet process prior for estimating lineage-specific substitution rates. Systematic Biology 29:939–255.

Heath, T. A., J. P. Huelsenbeck, and T. Stadler. 2014. The fossilized birth–death process for coherent calibration of divergence-time estimates. Proceedings of the National Academy of Sciences USA 111:E2957–E2966.

Holland, S. M. 2016. The non-uniformity of fossil preservation. Philosophical Transactions of the Royal Society of London B 371:20150130.

Keiding, N. 1975. Maximum likelihood estimation in the birth-and-death process. Annals of Statistics 3:363–372.

Labandeira, C. C. 2005. The fossil record of insect extinction: new approaches and future directions. American Entomologist 51:14–29.

Liow, L. H., and J. D. Nichols. 2010. Estimating rates and probabilities of origination and extinction using taxonomic occurrence data: capture-mark-recapture (CMR) approaches. Paleontological Society Papers 16:81–94.

Liow, L. H., and N. C. Stenseth. 2007. The rise and fall of species: implications for macroevolutionary and macroecological studies. Proceedings of the Royal Society of London B 274:2745–2752.

Marshall, C. R. 2017. Five paleobiological laws needed to understand the evolution of the living biota. Nature Evolution and Ecology 1:165.

Morlon, H., T. L. Parsons, and J. B. Plotkin. 2011. Reconciling molecular phylogenies with the fossil record. Proceedings of the National Academy of Sciences USA 108:16327–16332.

Nee, S., R. M. May, and P. H. Harvey. 1994. The reconstructed evolutionary process. Philosophical Transactions of the Royal Society of London B 344:305–311.

Norell, M. A., M. Novacek, and Q. Wheeler. 1992. Taxic origin and temporal diversity: the effect of phylogeny. Pp. 89–118 in M. Novacek, ed. Extinction and phylogeny. Columbia University Press, New York.

Quental, T., and C. R. Marshall. 2010. Diversity dynamics: molecular phylogenies need the fossil record. Trends in Ecology and Evolution 25:434–441.

Rabosky, D. L. 2010. Extinction rates should not be estimated from molecular phylogenies. Evolution 64:1816–1824.

Raup, D. M. 1972. Taxonomic diversity during the Phanerozoic. Science 177:1065–1071.

Raup, D. M. 1975. Taxonomic survivorship curves and Van Valen's law. Paleobiology 1:82–96.

Raup, D. M. 1985. Mathematical models of cladogenesis. Paleobiology 11:42–52.

Silvestro, D., N. Salamin, and J. Schnitzler. 2014a. PyRate: a new program to estimate speciation and extinction rates from incomplete fossil data. Methods in Ecology and Evolution 5:1126–1131.

Silvestro, D., J. Schnitzler, L. H. Liow, A. Antonelli, and N. Salamin. 2014b. Bayesian estimation of speciation and extinction from incomplete fossil occurrence data. Systematic Biology 63:349–367.

Silvestro, D., R. C. Warnock, A. Gavryushkina, and T. Stadler. 2018. Closing the gap between palaeontological and neontological speciation and extinction rate estimates. Nature Communications 9:5237.

Silvestro, D., A. Antonelli, N. Salamin, and X. Meyer. 2019. Improved estimation of macroevolutionary rates from fossil data using a Bayesian framework. Paleobiology 45:546–570.

Smiley, T. M. 2018. Detecting diversification rates in relation to preservation and tectonic history from simulated fossil records. Paleobiology 44:1–24.

Smith, A. B., and A. J. McGowan. 2007. The shape of the Phanerozoic marine palaeodiversity curve: how much can be predicted from the sedimentary rock record of western Europe? Palaeontology 50:765–774.

Soul, L. C., and M. Friedman. 2017. Bias in phylogenetic measurements of extinction and a case study of end-Permian tetrapods. Palaeontology 60:169–185.

Stadler, T. 2009. On incomplete sampling under birth–death models and connections to the sampling-based coalescent. Journal of Theoretical Biology 261:58–66.

Stadler, T. 2010. Sampling-through-time in birth–death trees. Journal of Theoretical Biology 267:396–404.

Stadler, T. 2011. Simulating trees with a fixed number of extant species. Systematic Biology 60:675–684.

Stadler, T., D. Kühnert, S. Bonhoeffer, and A. J. Drummond. 2013. Birth–death skyline plot reveals temporal changes of epidemic spread in HIV and hepatitis C virus (HCV). Proceedings of the National Academy of Sciences USA 110:228–233.

Stadler, T., A. Gavryushkina, R. C. Warnock, A. J. Drummond, and T. A. Heath. 2018. The fossilized birth-death model for the analysis of stratigraphic range data under different speciation modes. Journal of Theoretical Biology 447:41–55.

Wagner, P. J., and D. H. Erwin. 1995. Phylogenetic patterns as tests of speciation models. Pp. 87–122 in R. C. Anstey, ed. New approaches to speciation in the fossil record. Columbia University Press, New York.

Wagner, P. J., and J. D. Marcot. 2013. Modelling distributions of fossil sampling rates over time, space and taxa: assessment and implications for macroevolutionary studies. Methods in Ecology and Evolution 4:703–713.

Zhang, C., T. Stadler, S. Klopfstein, T. A. Heath, and F. Ronquist. 2015. Total-evidence dating under the fossilized birth-death process. Systematic Biology 65:228–249.