Syst. Biol. 69(2):325–344, 2020

DOI:10.1093/sysbio/syz038
Advance Access publication May 27, 2019

A Simulation-Based Evaluation of Tip-Dating Under the Fossilized Birth–Death Process

ARONG LUO[1,2,*], DAVID A. DUCHÊNE[2], CHI ZHANG[3,4], CHAO-DONG ZHU[1,5,6], AND SIMON Y.W. HO[2]

[1]*Key Laboratory of Zoological Systematics and Evolution, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China;* [2]*School of Life and Environmental Sciences, University of Sydney, Sydney, New South Wales 2006, Australia;* [3]*Key Laboratory of Vertebrate Evolution and Human Origins, Institute of Vertebrate Paleontology and Paleoanthropology, Chinese Academy of Sciences, Beijing 100044, China;* [4]*Center for Excellence in Life and Paleoenvironment, Chinese Academy of Sciences, Beijing 100044, China;* [5]*State Key Laboratory of Integrated Pest Management, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China; and* [6]*College of Life Sciences, University of Chinese Academy of Sciences, Beijing, 100049, China*
**Correspondence to be sent to: Key Laboratory of Zoological Systematics and Evolution, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China;*
E-mail: luoar@ioz.ac.cn

Received 03 October 2018; reviews returned 13 May 2019; accepted 17 May 2019
Associate Editor: Matt Friedman

Abstract.—Bayesian molecular dating is widely used to study evolutionary timescales. This procedure usually involves phylogenetic analysis of nucleotide sequence data, with fossil-based calibrations applied as age constraints on internal nodes of the tree. An alternative approach is tip-dating, which explicitly includes fossil data in the analysis. This can be done, for example, through the joint analysis of molecular data from present-day taxa and morphological data from both extant and fossil taxa. In the context of tip-dating, an important development has been the fossilized birth–death process, which allows non-contemporaneous tips and sampled ancestors while providing a model of lineage diversification for the prior on the tree topology and internal node times. However, tip-dating with fossils faces a number of considerable challenges, especially, those associated with fossil sampling and evolutionary models for morphological characters. We conducted a simulation study to evaluate the performance of tip-dating using the fossilized birth–death model. We simulated fossil occurrences and the evolution of nucleotide sequences and morphological characters under a wide range of conditions. Our analyses of these data show that the number and the maximum age of fossil occurrences have a greater influence than the degree of among-lineage rate variation or the number of morphological characters on estimates of node times and the tree topology. Tip-dating with the fossilized birth–death model generally performs well in recovering the relationships among extant taxa but has difficulties in correctly placing fossil taxa in the tree and identifying the number of sampled ancestors. The method yields accurate estimates of the ages of the root and crown group, although the precision of these estimates varies with the probability of fossil occurrence. The exclusion of morphological characters results in a slight overestimation of node times, whereas the exclusion of nucleotide sequences has a negative impact on inference of the tree topology. Our results provide an overview of the performance of tip-dating using the fossilized birth–death model, which will inform further development of the method and its application to key questions in evolutionary biology. [Bayesian phylogenetics; evolutionary simulation; fossilized birth–death process; molecular clock; tip-dating; total-evidence dating.]

Resolving the evolutionary timescale of the Tree of Life has been one of the long-standing goals of biological research. There has been remarkable progress in this area over the past few decades, driven largely by analyses of genetic sequences using the molecular clock (Zuckerkandl and Pauling 1965; Ho 2014; Donoghue and Yang 2016). Bayesian phylogenetic approaches hold particular appeal because they provide a unified framework for implementing models of nucleotide substitution, evolutionary rates, and lineage diversification (dos Reis et al. 2016; Bromham et al. 2018). At the same time, Bayesian molecular dating can incorporate calibrating information into the priors on node times. These calibration priors are usually applied to internal nodes of the tree, based on interpretations of relevant fossil evidence (Ho and Phillips 2009; Donoghue and Yang 2016) or biogeographic events (Ho et al. 2015b; De Baets et al. 2016). However, the recent introduction of fossil tip-dating enables fossils to be included as sampled taxa in the analysis, with the ages of the fossils supplying the calibrating information (Pyron 2011; Ronquist et al. 2012).

Tip-dating can make use of total-evidence data matrices that comprise both molecular sequences and morphological characters, with fossil taxa being analyzed together with their living relatives (e.g., Pyron 2011; Ronquist et al. 2012). In this approach, phylogenetic positions of the fossil taxa are informed by the morphological characters and age information is provided directly by the fossils, without the need to specify calibration priors for internal nodes in the tree. Therefore, the tip-dating framework makes it possible to include all fossils for which morphological characters are available for the group being studied (Ronquist et al. 2012). In this regard, tip-dating offers a key advantage over previous "node-dating" approaches, in which the minimum bound of each calibration is typically based on the age of the oldest known fossil that has been assigned to the clade descending from that node. Another key advantage of tip-dating is that it removes the need to specify any maximum age constraints on internal nodes; in node-dating analyses, these constraints are often chosen without strong justification but have potentially large impacts on the resulting date estimates (Hug and Roger 2007). A further benefit is that tip-dating can

eliminate the potential problem of marginal calibration priors differing from the user-specified priors, which arises when the latter are combined multiplicatively with each other and with the tree prior (Heled and Drummond 2012). Tip-dating has been used to infer the evolutionary timescales of various groups of taxa, including birds (Gavryushkina et al. 2017), fishes (e.g., Near et al. 2014; Arcila et al. 2015; Arcila and Tyler 2017), mammals (e.g., Slater 2013; Herrera and Dávalos 2016; Kealy and Beck 2017), and plants (e.g., Larson-Johnson 2016).

An important step in the evolution of tip-dating was the development of the fossilized birth–death (FBD) process (Stadler 2010). This model is designed to generate the probability density of a tree with individuals sampled through time in an epidemiological or phylogenetic context. In phylogenetic analyses that involve fossil occurrences, the FBD process provides a model of lineage diversification that accounts for speciation, extinction, fossilization, and taxon sampling (Heath et al. 2014; Gavryushkina et al. 2014). The FBD model can allow sampled ancestors, whereby sampled fossils are direct ancestors of other taxa in the data set. Extensions of the FBD model include treating the process of fossilization and sampling as a piecewise function (Gavryushkina et al. 2014) and accommodating different taxon-sampling strategies (Zhang et al. 2016), while more recent developments integrate multispecies coalescent models (Ogilvie et al. 2018) and speciation modes (Stadler et al. 2018). The FBD model can also be used without morphological characters, but under these circumstances the performance of the approach is improved when topological constraints are placed on the fossil taxa (Heath et al. 2014). Alternatively, the FBD model can be applied to data sets that exclusively comprise morphological characters from fossil taxa and their extant relatives (Bapst et al. 2016; Matzke and Wright 2016; King et al. 2017; Matzke and Irmis 2018).

The FBD model has parameters that represent the speciation rate (λ), extinction rate (μ), and fossil recovery rate (ψ), along with the start time of the process (origin time t_{or}, root age t_{mrca}, or crown age t_c) and sampling fraction of extant taxa (ρ). For mathematical convenience, the model is reparameterized using the net diversification rate ($d = \lambda - \mu$), turnover rate ($r = \mu/\lambda$), and fossil sampling proportion ($s = \psi/(\mu + \psi)$) (Heath et al. 2014). Simulation-based studies have shown that the FBD model is generally able to recover the parameters used for simulation, though with some exceptions (Gavryushkina et al. 2014; Zhang et al. 2016); for example, the uncertainty in turnover rates has been found to decrease with the size of the tree and to vary with the sampling strategy for extant taxa. Drummond and Stadler (2016) found that the FBD model was able to infer the ages of fossil samples with a good degree of accuracy, confirming its internal consistency with other models within the Bayesian framework. However, analyses of empirical data have yielded date estimates that are often considerably younger when using the FBD model than when using other tree priors (e.g., the Yule process; Pyron 2011) for tip-dating (Herrera and Dávalos 2016; Zhang et al. 2016; Gavryushkina et al. 2017). There have also been some discrepancies between the results obtained from tip-dating under the FBD model and from node-dating (Vea and Grimaldi 2016; Arcila and Tyler 2017; Gustafson and Miller 2017; Kealy and Beck 2017).

In principle, tip-dating using the FBD model provides a satisfying approach because it combines the available data from both fossil and extant taxa. However, the inclusion of morphological data and fossil taxa presents a number of complex challenges. First, fossil specimens are often incomplete or fragmentary, leading to potential difficulties in resolving their phylogenetic placements (Sansom et al. 2010; Sansom and Wills 2013). Second, morphological data are typically analyzed using the Mk model (Lewis 2001), a k-states generalization of the Jukes–Cantor model of nucleotide substitution (Jukes and Cantor 1969), but this model makes several simplifying assumptions that are likely to be violated by real data (Wright et al. 2016). Third, morphological characters are likely to have evolved in a far less clocklike manner than nucleotide sequences (dos Reis et al. 2016; Donoghue and Yang 2016; Drummond and Stadler 2016). Fourth, although the ages of fossils are usually treated as being known without error, this assumption is potentially problematic (O'Reilly et al. 2015; Barido-Sottani et al. 2019). Finally, rates of fossilization and fossil sampling might vary across clades, whereas the FBD process typically assumes homogeneity of these rates throughout the tree (Matschiner et al. 2017).

In this study, we evaluate the performance of tip-dating under a range of conditions. Our analyses are based on synthetic data generated by simulating fossil occurrences, evolution of nucleotide sequences, and evolution of morphological characters on trees generated under a birth–death process. Using the FBD model for the tree prior, we examine how Bayesian estimates of node times and tree topologies are affected by fossil occurrences, number of morphological characters, and degree of among-lineage rate heterogeneity. In addition, we examine the performance of tip-dating when using only nucleotide sequences or morphological characters, and consider the influence of the model of morphological evolution and uncertainty in fossil ages. The results of our analyses allow us to present some practical guidelines for tip-dating using the FBD model.

Materials and Methods

Species Trees and Fossil Occurrences

Using TreeSim (Stadler 2011) in R (R Core Team 2017), we simulated speciation according to the birth–death process to produce 1000 trees (Stadler 2009), each with 50 extant taxa and between 8 and 83 extinct taxa (median = 31). These simulations were performed using a constant speciation rate $\lambda = 0.05$ per myr, constant extinction rate $\mu = 0.02$ per myr, and sampling fraction

$\rho = 1$ (i.e., complete sampling of present-day taxa). The diversification process was conditioned on the number of extant species, with speciation and extinction rates chosen to generate appropriate root ages and numbers of extinct tips. From the 1000 complete trees that were produced, we selected 20 trees that had crown ages of about 100 Ma (±1 Ma). These 20 trees varied in the total number of tips (74–105), origin times t_{or} (101–196 Ma), and tree shapes as measured by the Colless index (3.7–8.5, corrected by the number of tips; Colless 1982). Ten of the 20 trees had root ages t_{mrca} equal to their crown ages t_c, whereas the others had root ages that were greater than their crown ages (Table 1; Fig. 1a and b).

We simulated fossil sampling on each of the 20 complete birth–death trees (Fig. 1c) using a single parameter, P, to represent the probability of fossil occurrence (Warnock et al. 2017). Preservation potential and sampling intensity were not specified separately (Heath et al. 2014). Fossil occurrences were modeled as a Bernoulli process in time slices of 2 myr throughout the duration of each species tree, except along the branch between t_{mrca} and t_{or} for the sake of technical convenience. We employed three uniform models of fossil occurrence (P = 0.01, 0.02, and 0.05) across the 20 trees, then considered a simple nonuniform model in which P decreases linearly with t from 0.05 ($t = 2$) to 0.005 ($t = 100$) and ultimately to zero (Table 1). Each model of P was deployed once on each complete tree. After pruning lineages without a fossil occurrence or an extant descendant, we obtained a total of 80 (20 × 4) FBD trees. We focused on these trees for our tip-dating analyses.

Because the use of P for simulating fossil sampling is not a strict match to the fossil recovery rate ψ within the FBD model, which models the occurrence of fossils as a continuous process, we investigated the relationship between the values of P and ψ. First, the above three uniform and one nonuniform models of P were again used to simulate fossil occurrences, each with 50 replicates for each of the 20 complete trees. Second, with ψ = 0.005, 0.01, and 0.025, we simulated fossil occurrences on the 20 complete trees (except between t_{mrca} and t_{or}) following a Poisson process (Heath et al. 2014), with 50 replicates for each value for each complete tree. The ψ values were chosen to be half of those of P in the uniform models (because the time increments of ψ and P here are 1 myr and 2 myr, respectively), so that our two approaches to simulating fossil sampling were expected to generate similar numbers of fossil occurrences (Supplementary Appendix 1 available on Dryad at http://dx.doi.org/10.5061/dryad.q2527ts).

Simulations of Character Evolution

For each FBD tree, we simulated the evolution of nucleotide sequences along the reconstructed history of the extant species only. Two models of among-lineage rate variation were used to transform the chronograms into phylograms in NELSI v0.2 (Ho et al. 2015a). First, we assumed a strict molecular clock with a rate of 10^{-3} subs/site/myr. Second, we used a relaxed molecular clock, the white-noise model, to allow rate variation across branches (Lepage et al. 2007; Ho et al. 2015a), with mean 10^{-3} subs/site/myr and standard deviation 2×10^{-4} subs/site/myr. Sequence evolution was simulated using Seq-Gen v1.3.4 (Rambaut and Grassly 1997) to produce five 1000 bp sequence alignments (equivalent to five "loci") for each phylogram, with relative evolutionary rates randomly sampled from a symmetric Dirichlet distribution with $\alpha = 3$. Simulations were performed using the HKY+G substitution model with base frequencies {A:0.35, C:0.15, G:0.25, T:0.25}, transition/transversion ratio $\kappa = 4.0$, and a gamma shape parameter of 0.5.

We simulated the evolution of morphological characters for both extant species and sampled fossils along each of the 80 FBD trees. To account for rates of morphological evolution being more likely to vary among lineages (dos Reis et al. 2016), we used the white-noise model of branch rates with a mean of 10^{-3} changes/character/myr and three different standard deviations: 0 (i.e., a strict clock), 2×10^{-4} changes/character/myr, and 5×10^{-4} changes/character/myr. To emulate the Mk model, we used Seq-Gen first to simulate nucleotide sequence evolution with base frequencies {A:0.0, C:0.0, G:0.5, T:0.5}, then converted the resulting nucleotides into binary characters by recoding G to 0 and T to 1 (Puttick et al. 2017). Our simulations produced full data sets of three sizes: l = 100, 200, and 1000 characters. After pruning the nonvariable characters, these data sets had l = 56–92, 122–178, and 630–891 characters, respectively.

For our core analyses, which are described in detail below, nucleotide sequences from the extant species were combined with morphological characters from both the extant species and the sampled fossils. These produced three different scenarios of rate variation among lineages: nucleotide sequences under a strict clock with morphological characters under a strict clock ("SS"); nucleotide sequences under a strict clock with morphological characters under moderate rate variation ("SM"); and nucleotide sequences under moderate rate variation with morphological characters under high rate variation ("MH"). Our simulations produced 240 data sets under each combination of clock models, differing with respect to their underlying FBD trees and/or the numbers of morphological characters. All of our data sets and input files are available in Supplementary Appendix 2 available on Dryad.

Fossil Tip-Dating

Evaluation of the FBD process.—We first evaluated the outcomes of the FBD process with fossil occurrences sampled by P and ψ on the 20 birth–death trees, paving the way for the subsequent tip-dating analyses. For each of the 7000 FBD trees (Fig. 1c), we fixed the

TABLE 1. Details of the 20 birth–death species trees and fossil occurrences sampled according to a parameter representing the probability of fossil occurrence P

Index	Tree 1	Tree 2	Tree 3	Tree 4	Tree 5	Tree 6	Tree 7	Tree 8	Tree 9	Tree 10	Tree 11	Tree 12	Tree 13	Tree 14	Tree 15	Tree 16	Tree 17	Tree 18	Tree 19	Tree 20
t_{or} (Ma)	101	102	103	105	105	105	106	107	110	113	122	132	141	142	148	159	161	166	175	196
t_{mrca} (Ma)	99	99	100	100	101	104	100	101	100	100	100	131	121	140	119	100	154	161	173	185
t_c (Ma)	99	99	100	100	99	101	100	101	100	100	100	99	100	100	101	100	100	100	100	99
No. of tips (all)	86	74	78	93	92	96	78	76	75	85	105	90	98	74	93	93	77	74	75	78
Colless index (all)	3.9	4.7	3.7	4.6	6.9	8.5	4.0	3.8	5.2	5.2	4.8	7.2	7.7	5.9	7.9	4.6	4.1	7.8	6.9	7.2
Colless index (extant)	2.1	2.0	2.7	3.2	2.3	3.9	2.9	2.4	3.9	4.9	3.6	2.8	2.5	2.9	3.0	2.6	2.5	2.8	2.8	2.8
No. of fossils, P= 0.01	5	8	4	9	13	7	6	12	5	5	14	8	11	5	8	7	8	7	5	8
No. of fossils, P= 0.02	13	12	16	18	24	17	21	16	17	17	29	14	24	12	10	19	21	15	20	13
No. of fossils, P= 0.05	37	37	35	42	40	44	35	44	42	35	54	48	64	35	36	39	43	38	34	55
No. of fossils, nonuniform P	30	18	28	41	37	20	33	29	31	20	50	30	37	24	31	38	26	24	25	30
No. of SAs, P = 0.01	4	6	3	6	11	4	5	11	5	5	12	6	10	4	5	5	7	6	5	6
No. of SAs, P = 0.02	9	10	15	14	18	12	18	14	14	12	22	8	18	10	7	16	17	11	19	11
No. of SAs, P = 0.05	25	27	28	34	33	32	32	35	39	26	41	35	50	31	33	23	34	33	27	47
No. of SAs, nonuniform P	25	13	23	37	28	14	29	24	28	14	43	26	24	18	26	29	23	22	22	23
Max. age (Ma), P= 0.01	38	92	80	40	64	94	66	38	96	52	86	78	74	52	84	78	78	66	52	166
Max. age (Ma), P= 0.02	60	66	94	72	96	102	80	42	66	78	96	88	118	136	114	68	142	102	144	178
Max. age (Ma), P= 0.05	74	82	96	100	74	100	82	84	92	98	94	124	118	132	96	86	142	144	170	176
Max. age (Ma), nonuniform P	92	64	78	64	62	84	60	74	58	74	78	62	72	78	82	66	78	56	96	76

Notes: t_{or} = the origin time of a species tree; t_{mrca} = the root age of a species tree; t_c = the age of the crown group in a species tree; No. of tips (all) = the total number of tips in a species tree; Colless index (all) = the corrected Colless index by the number of all tips in a species tree; Colless index (extant) = the corrected Colless index by the number of extant tips in a species tree; No. of fossils = the total number of fossils sampled in a species tree by the probability of fossil occurrence P = 0.01, 0.02, 0.05, or nonuniform; No. of SAs = the total number of sampled ancestors among fossils sampled in a species tree, based on the probability of fossil occurrence P = 0.01, 0.02, 0.05, or nonuniform; Max. age = the maximum age of sampled fossils in a species tree by the probability of fossil occurrence P = 0.01, 0.02, 0.05, or nonuniform.

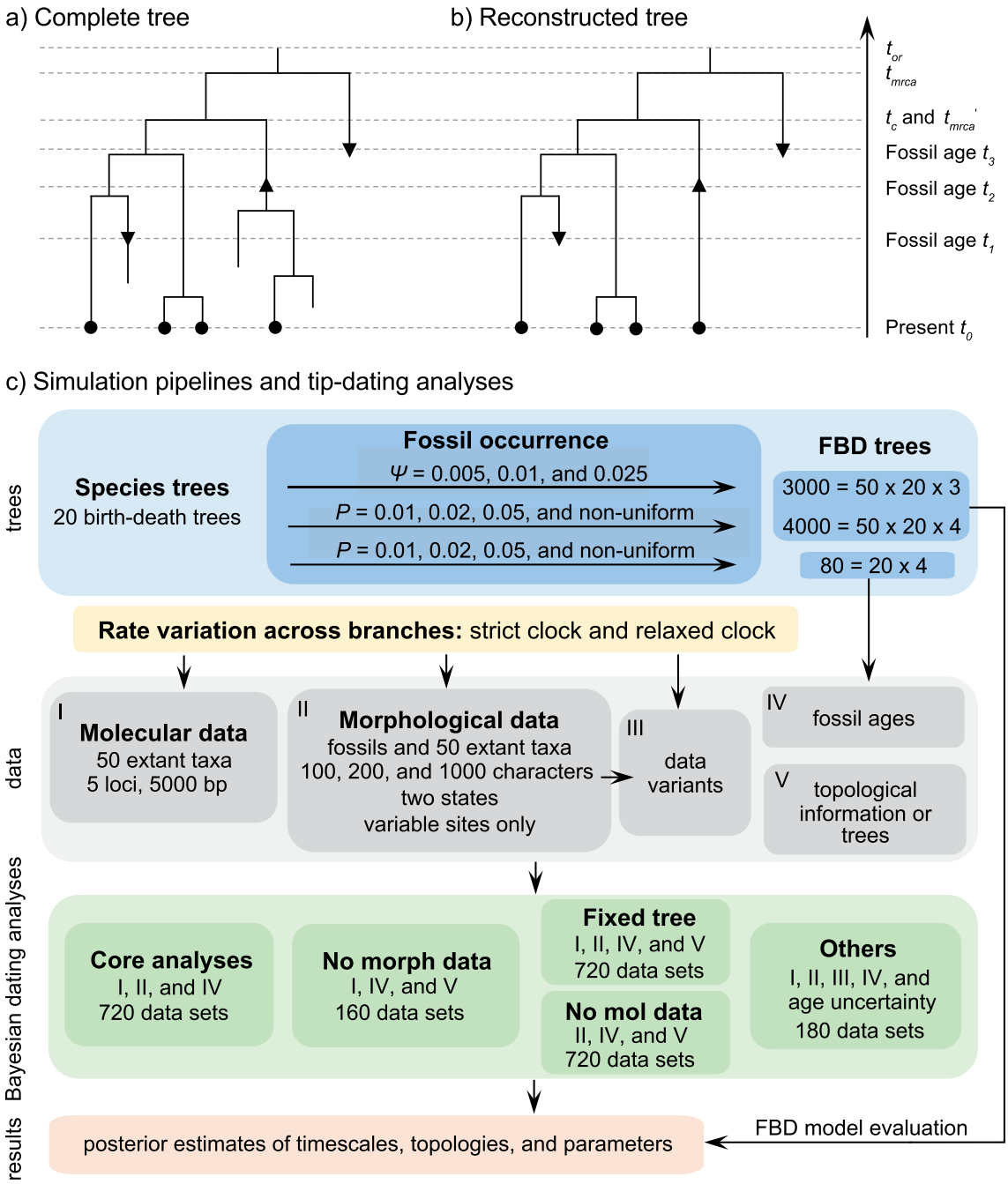

FIGURE 1. a) Illustration of a complete tree generated under the birth–death process. Lineage diversification is controlled by birth rate λ, death rate μ, and sampling fraction ρ. From the origin time (t_{or}) to the present-day (t_0), fossils have been sampled at t_1, t_2, and t_3, with one (denoted by the upward triangle) leaving an extant descendant (denoted by solid circle) and the other two (denoted by downward triangles) leaving no extant descendants. With complete sampling of present-day taxa (ρ = 1), the age of the crown group t_c remains the same, whereas the age of the root t_{mrca} depends on whether fossils are sampled between t_{mrca} and t_{mrca}'. b) The FBD tree depicting the reconstructed history of present-day taxa and sampled fossil taxa based on the complete species tree in (a). c) Flowchart showing the simulation pipelines and analyses conducted in this study. A detailed explanation of each step is provided in Materials and Methods section. Briefly, we obtained the FBD trees by simulating speciation using the birth–death process, with the probability of fossil occurrences based on either P and ψ. Among these FBD trees, the 80 trees with fossil occurrences sampled by P were the main basis of this study. These trees provided the fossil ages and topologies. We simulated the evolution of nucleotide sequences and morphological characters on these trees, under various models of rate variation among lineages. We carried out series of Bayesian dating analyses under a range of settings and using various subsets of the data. These analyses yielded estimates of the posterior distribution of tree topologies, node times, and model parameters.

tree topology, branch lengths (in units of myr), t_{or}, and ρ to their true values, and excluded the sequence data and morphological characters in turn. We adopted diffuse priors, including beta distributions $B(1,1)$ for the turnover r and fossil sampling proportion s, and an exponential distribution with mean 0.1 for the net diversification d. We estimated the posterior distribution using Markov chain Monte Carlo (MCMC) sampling with the SA (sampled ancestor) package in BEAST v2.4.8 (Bouckaert et al. 2014; Gavryushkina et al. 2014). MCMC samples were drawn every 5000 steps over a total of 20 million steps and a burn-in fraction of 0.25, which were effectively drawn from the prior distributions of d, r, and s, and the fossil occurrence ages. We used a modified version of BEAST v2.4.8 that allowed us to retain the sampled ancestors; the standard release of the software automatically adjusts zero-length branches to take small (nonzero) lengths, which would affect the inference of sampled ancestors. We ran analyses in duplicate to check for convergence, which was done by inspecting the effective sample sizes of parameters for the combined samples in Tracer v1.7 (Rambaut et al. 2018). Effective sample sizes exceeded 200 for all parameters.

Core analyses.—We used the FBD model for the tree prior in our analyses of the data sets produced by our simulations, with the ages of the sampled fossils treated as point values (Fig. 1c). The HKY+G model with four rate categories was used for the nucleotide sequences (Yang 1994), whereas the Mkv model was used for the variable characters in the morphological data (Lewis 2001). We applied separate uncorrelated lognormal relaxed-clock models to the molecular and morphological data (Drummond et al. 2006), with a uniform prior $U(10^{-6},1)$ for the mean rate in all analyses. For the parameters of the FBD model, we used relatively diffuse priors: beta distributions $B(1,1)$ for r and s; an exponential distribution with mean 0.1 for d; and a uniform distribution for t_{or}, with a maximum bound of 300 Ma and a minimum bound matching the age of the oldest sampled fossil. The sampling proportion of extant species was fixed to 1 to match the settings used in our simulations.

The 720 data sets produced by our simulations were analyzed using BEAST with the SA package. For each data set, we carried out two independent MCMC analyses in order to check for convergence. Each MCMC analysis consisted of 100 million steps, with samples drawn every 5000 steps and with a discarded burn-in fraction of 0.25. We checked for sufficient sampling by ensuring that all parameters had effective sample sizes of at least 100. Maximum-clade-credibility trees were identified from the combined samples using TreeAnnotator v1.8.4. To investigate the maximum-clade-credibility tree topology, we pruned the fossils to produce annotated trees containing only extant taxa. An additional MCMC analysis was performed without data, to allow us to evaluate the combined signal from the sequence data and morphological characters.

No morphological characters.— The FBD model can be used in the absence of morphological characters, such that the diversification process is marginalized over all of the possible placements of the fossil occurrences (Heath et al. 2014). We performed analyses with the morphological characters excluded, so that the data comprised only the nucleotide sequences (generated using either a strict clock or with moderate among-lineage rate variation) and fossil occurrence times (i.e., 160 data sets in total). We used two different strategies for specifying the placements of the fossils in the tree during MCMC sampling. First, based on its parent node in the FBD trees, we imposed a monophyletic constraint for each fossil and placed it into its correct group with its extant relative(s) (or extant and extinct relatives), so that the fossil could potentially be a crown or stem fossil for its extant relative(s). Second, we did not specify any constraints on the tree, so that we sampled full trees for the fossil positions conditioned on fixed $\rho = 1$ (Gavryushkina et al. 2014). Other settings for the MCMC analyses were the same as for the core analyses (Fig. 1c).

No molecular data.—To examine the performance of Bayesian dating with the FBD model applied to data sets comprising only morphological characters, we excluded nucleotide sequences from the 720 data sets of the core analyses. Analyses were then performed using the morphological characters of the extant and fossil taxa, with the ages of the fossils being taken into account. We used the same settings as for the core analyses, but with samples drawn every 1000 steps from a total of 20 million MCMC steps.

Fixed tree topology.—We carried out analyses using the sequence data and morphological characters, with the FBD tree topologies fixed to those used for simulation (i.e., true topologies) rather than jointly estimating the topologies as in our core analyses above. Clock models, choices of priors, and the MCMC settings were the same as those used for the core analyses. The sole exception was that posterior distributions of parameters were estimated from samples drawn every 2000 steps from a total of 40 million MCMC steps.

Alternative conditions.—We performed further analyses to investigate the influence of several factors that could be influential on tip-dating in practice. These analyses were all based on the nucleotide sequences and morphological characters produced by simulation with the SS pattern of among-lineage rate variation, fossil occurrences obtained using $P = 0.05$, and other appropriate settings for each set of analyses (as described below). Given the three sizes of morphological character sets (i.e., l = 100, 200, and 1000) and 20 FBD trees, there were 60 data sets for each set of analyses (Fig. 1c).

First, to test the impact of different numbers of morphological character states, we replaced the binary morphological characters with four-state morphological characters. To generate the four-state data, we simulated the evolution of nucleotide sequences using Seq-Gen with the Jukes–Cantor model, converted the nucleotides

in the resulting sequences to numerical multistate coding (A to 0, C to 1, G to 2, and T to 3), and then pruned the nonvariable characters. Second, to examine the impacts of removing ascertainment bias (Lewis 2001), we performed analyses using the Mk model with the full data sets of binary morphological characters, rather than using the Mkv model with only the variable characters. In our third set of analyses here, we aimed to test whether accounting for uncertainty in fossil ages would have any impact on date estimates (O'Reilly et al. 2015; Barido-Sottani et al. 2019). We used uniform priors rather than point values for the fossil sampling times. The bounds of these uniform priors were chosen to match the boundaries of the stratigraphic stage from which each fossil was sampled, as defined by the International Commission on Stratigraphy (February 2017). For example, if a fossil had been sampled at 76 Ma, we instead used a uniform prior $U(71.9,83.8)$ to reflect the age boundaries of the Campanian stage of the Upper Cretaceous. Other settings for the MCMC analyses were the same as those used for our core analyses.

Evaluation of Fossil Tip-Dating

Our main objectives were to examine the estimates of node times and/or tree topologies inferred from data generated under different simulation conditions. We thus treated estimates for the 20 birth–death species trees as independent replicates under each set of conditions (e.g., 20 repeats under the SS pattern of among-lineage rate variation, P = 0.05, and l = 100). Unless noted otherwise, we did not consider in any detail the differences across replicates.

To evaluate the performance of tip-dating with the FBD model, we examined the posterior medians of model parameters and node times. To allow further comparisons of estimated node times, we used three metrics that involved standardizing the absolute estimates. First, as a measure of accuracy we computed relative bias, which is the distance between the posterior median and the true value, divided by the true value. Second, as a measure of precision we computed the relative 95% credibility interval (CI) width, which is the 95% CI width divided by the true (point) value. Third, we computed the coverage probability, which is the proportion of 95% CIs that contain the true values. Additionally, after pruning fossil taxa, we used the gamma statistic (Pybus and Harvey 2000) and stemminess rank (Fiala and Sokal 1985) to summarize relative node depths in the maximum-clade-credibility trees based on posterior medians of node times. We focused on three key time points including the origin time of the FBD process (t_{or}), the root age (or time to the most recent common ancestor of the sampled taxa, t_{mrca}), and the crown age (or time to the most recent common ancestor of the extant taxa, t_c). We also examined date estimates for the youngest and median nodes in each tree, which we chose from the trees used for simulation and were conditioned on the presence of these nodes in the resultant maximum-clade-credibility trees.

To evaluate the differences between each maximum-clade-credibility tree and the true topology, we used two measures of topological distance. First, we computed the absolute Robinson–Foulds topology distance, which is defined as twice the number of internal branches defining different bipartitions of the tips (Robinson and Foulds 1981; Penny and Hendy 1985). Second, we corrected this distance by the total number of tips in the tree. Distance calculations were performed using the R package ape (Paradis et al. 2004; Popescu et al. 2012). To measure the performance of tip-dating in placing fossils into their correct phylogenetic positions, we split the sampled fossils into two categories, based on whether they had extant descendants or not (Fig. 1a and b). For fossils that left extant descendants, we measured whether their positions were correctly inferred or not by using two criteria: monophyletic grouping, which depends on whether the fossil is grouped with its extant and/or extinct relative(s); and being a sampled ancestor, which depends on whether its terminal branch length is zero or not. For fossils that did not leave extant descendants, we recorded whether they were correctly identified as sampled ancestors or not.

Results

Recovery of the FBD Parameters

The parameters of the FBD model were generally well recovered when the tree topology and branch lengths were fixed to those of the trees used for simulation (Fig. 2). As expected, our models with uniform P(P = 0.01, 0.02, and 0.05) led to fossil sampling proportions s similar to those from the three ψ values (ψ = 0.005, 0.01, and 0.025), supporting the mathematical approximation $P \approx 2 \times \psi$ that we have used in our study. The median estimates of s across each set of 1000 (50 × 20) FBD trees approached the true values (0.2, 0.33, and 0.56, respectively). The model with nonuniform P yielded estimates of s most similar to those produced with P = 0.05, being generally consistent with the numbers of sampled fossils. The rates of net diversification (d = 0.03) and turnover (r = 0.4) were generally estimated accurately. However, we identified some biases at finer scales, such as a tendency to underestimate d and to overestimate r when P = 0.05. We also observed an increase in the estimated turnover rate with increasing ψ and P.

Impacts of Rate Variation, Fossil Occurrences, and Number of Morphological Characters

Based on our core analyses, we examined the impacts of the three main factors that we varied across our simulations: degree of rate variation across branches, probability of fossil occurrence P, and number of binary morphological characters l (Table 2). To evaluate their

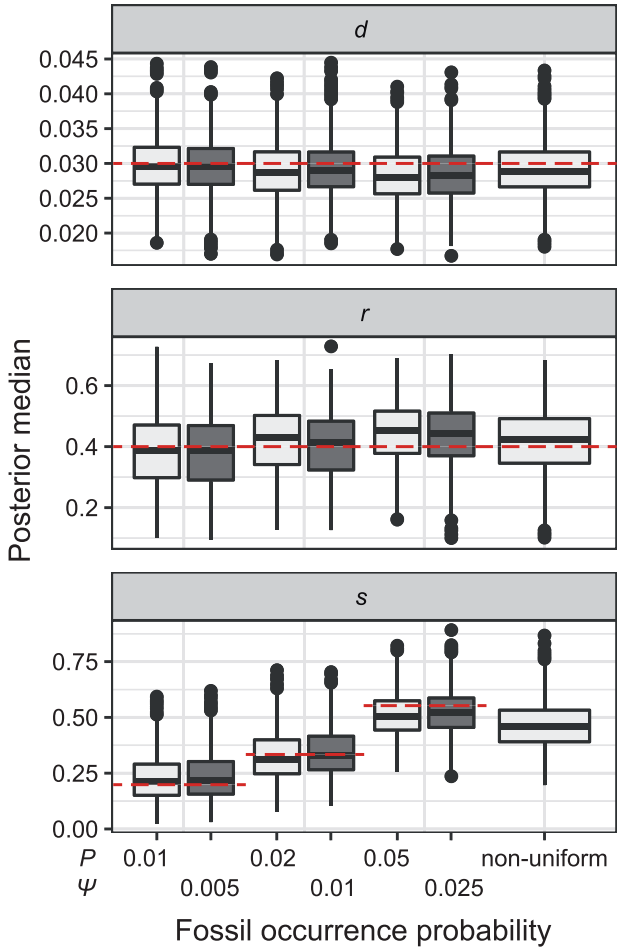

FIGURE 2. Posterior medians of the FBD model parameters from our evaluation of the FBD process, while conditioning on fixed tree topologies and branch lengths. The three panels show boxplot summaries of posterior estimates of net diversification rate ($d = \lambda - \mu$), turnover rate ($r = \mu/\lambda$), and fossil sampling proportion ($s = \psi/(\mu + \psi)$). Each summary is based on a set of 1000 FBD trees, which were derived from fossil occurrences sampled by P (light grey shading) or ψ (dark grey shading) on our 20 simulated species trees. The dashed horizontal lines indicate the true values of d, r, and s that were used for simulation.

impacts via standardized metrics, we focused on the topological distance between the inferred topology and the true topology, as well as the date estimates for the three key time points (Fig. 1a and b): the origin time (t_{or}), the root age (t_{mrca}), and the crown age (t_c).

The performance of topological inference depended on whether or not the sampled fossils were taken into account. When fossil taxa were pruned, the maximum-clade-credibility trees were very similar to those used for simulation (median of corrected Robinson–Foulds distances = 0.04; Fig. 3a). When fossil taxa were retained, however, the differences between the maximum-clade-credibility trees and true trees were much larger (median of corrected Robinson–Foulds distances >0.1 if P = 0.01); corrected Robinson–Foulds distances increased with P in the models with uniform probabilities of fossil occurrence, while those for the analyses of data generated with the nonuniform P model fell between the values found in analyses with P = 0.02 and P = 0.05 (Fig. 3b). Topological distances showed weaker trends with the degree of rate variation and l, especially when the fossil taxa were pruned (Fig. 3a and b).

Coverage probabilities (percentage of cases in which the 95% CIs contained the true values) were 86.9%, 85.7%, and 82.1% for t_{or}, t_{mrca}, and t_c, respectively. They did not show clear associations with the three main factors across the results of our core analyses, except for a lower coverage probability for t_{or} when the probability of fossil occurrence P was nonuniform (Supplementary Appendix 3 available on Dryad). We thus focus on relative bias and relative 95% CI width as respective measures of the accuracy and precision of our time estimates. We found that low rate variation across branches (SS or SM patterns of rate variation) led to estimates that had slightly better accuracy and precision than those from scenarios with higher rate variation across branches (MH pattern of rate variation; Fig. 4). The impacts of the fossil occurrence probability P and the number of morphological characters l varied among simulation treatments.

The accuracy of date estimates was not substantially affected by P or l (Fig. 4a), with the relative biases of t_{or}, t_{mrca}, and t_c being close to 0. However, the accuracy of estimates of t_{or} slightly increased with P in the uniform models. For the nonuniform P model, the spread of date estimates across the simulation replicates was similar to that when P = 0.01. There were some large overestimates for the three time points, although the proportions of estimates that were greater than the true dates were 58.6%, 58.8%, and 57.2% for t_{or}, t_{mrca}, and t_c, respectively. Combinations of P and l had clear impacts on the precision of time estimates (Fig. 4b). For t_{or}, t_{mrca}, and t_c, the relative 95% CI widths decreased with increasing P in the uniform models, while those with the nonuniform P model were smaller than those when P = 0.02 but larger than those when P = 0.05. As expected, the precision of time estimates generally increased with l. The relative 95% CI widths of the estimates of t_{or} were generally greater and more variable across replicates than those of t_{mrca} and t_c. For example, given P = 0.01, the means of the relative 95% CI widths were 0.78, 0.35, and 0.34 for t_{or}, t_{mrca}, and t_c, respectively.

Relative Node Times and Placements of Fossil Taxa

We used the gamma statistic and stemminess rank to summarize the relative node times in the maximum-clade-credibility trees without fossils. When these were plotted against the corresponding metrics for the trees used for simulation, the lines of best fit had slopes close to 1.00 for all scenarios of rate variation (Supplementary Appendix 4 available on Dryad). However, some estimation errors were apparent with higher levels of among-lineage rate heterogeneity, as seen in the MH pattern of rate variation (Pearson's correlation coefficient R = 0.93 for the gamma statistics). This result was

TABLE 2. Summaries of the effects of the main factors that we considered on the tree topologies and three key time points estimated by the tip-dating analyses in this study

Index			Core analyses						Alternative conditions[a]		
			Fossil occurrence possibility P	Number of morphological characters l	Scenario of among-lineage rate variation	No morph data[a]	No mol data[a]	Fixed tree[a]	Four-state morph data	Mk model	Fossil age uncertainty
Corrected R-F distance		Excluding fossils	No effect	No effect	No effect	Similar	Larger values, and decreasing with l	—	Similar	Similar	Similar
		Including fossils	Increasing with P = 0.01, 0.02, nonuniform, and 0.05	Weak effect	Weak effect	Smaller values	Similar, but larger values and decreasing with l	—	Similar	Similar	Similar
t_{or}		Accuracy	Increasing with P = 0.01, nonuniform, 0.02, and 0.05	No effect	Weak effect	Poorer	Similar	Similar, but with no extreme relative biases	Similar	Similar	Similar
		Precision	Increasing with P = 0.01, 0.02, nonuniform, and 0.05	Increasing with l	Weak effect	Poorer	Similar	Similar	Similar	Similar	Similar
t_{mrca}		Accuracy	No effect	No effect	Weak effect	Poorer	Similar	Similar, but with no extreme relative biases	Similar	Similar	Similar
		Precision	Increasing with P = 0.01, 0.02, nonuniform, and 0.05	Increasing with l	Weak effect	Poorer	Similar	Similar, and smaller relative 95% CI widths	Similar	Similar	Poorer
t_c		Accuracy	No effect	No effect	Weak effect	Poorer	Similar	Similar, but with no extreme relative biases	Similar	Similar	Similar
		Precision	Increasing with P = 0.01, 0.02, nonuniform, and 0.05	Increasing with l	Weak effect	Poorer	Similar	Similar, and smaller relative 95% CI widths	Similar	Similar	Similar

Notes: Corrected R-F distance = the Robinson–Foulds topology distance, corrected by the total number of tips; excluding fossils = on the condition of excluding fossil taxa; including fossils = on the condition of including fossil taxa; accuracy = measured by relative bias (distance between posterior median and true value, divided by the true value); precision = measured by relative 95% credibility interval (CI) width (95% CI width divided by the true value); Four-state morph model = the analyses of four-state morphological characters; Mk model = the analyses using the Mk model for the full morphological data sets; Fossil age uncertainty = the analyses in which the uncertainty in fossil ages was taken into account.

[a]Interpretations of the results of these analyses are based on comparisons with the factor effects in the core analyses.

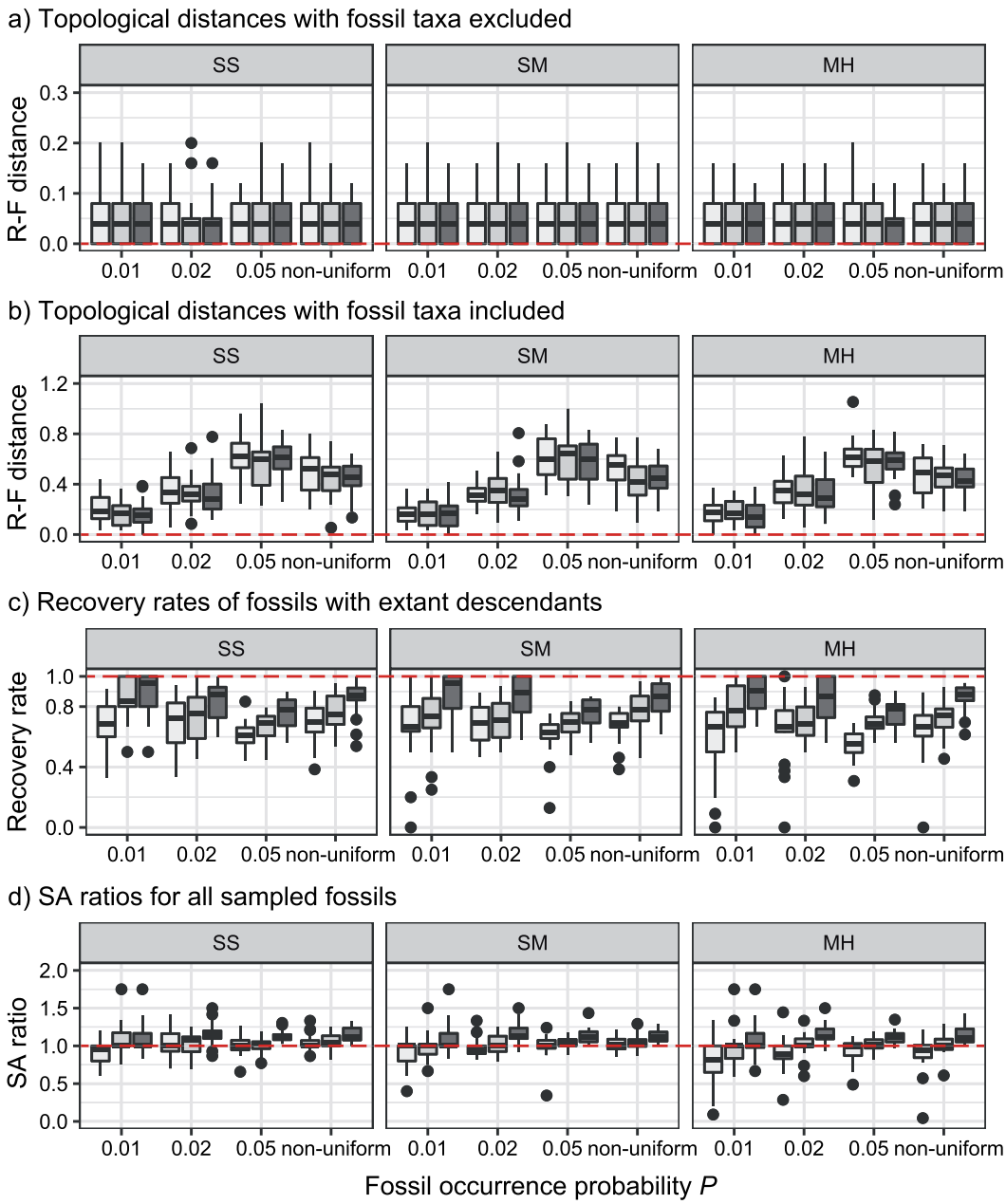

FIGURE 3. Performance of tip-dating in our core analyses in topological inference. Dashed horizontal lines indicate the target values for the metrics in each of the plots. Plots show corrected R-F distances between maximum-clade-credibility trees and true trees, while (a) excluding fossil taxa or (b) including fossil taxa. c) Recovery rates of correct phylogenetic positions for fossil taxa that have left extant descendants. d) Ratios of placing all sampled fossils as SA in maximum-clade-credibility trees to the true numbers of sampled ancestors. Each panel shows the results from a different model of among-lineage rate variation for the molecular and morphological data: strict clock and strict clock (SS); strict clock and moderate rate variation (SM); and moderate rate variation and high rate variation (MH). Within each panel, boxplot summaries are shown for the 20 FBD trees under each model of fossil occurrence probability ($P = 0.01$, 0.02, 0.05, and nonuniform). For each fossil occurrence probability, results are shown for three different sizes of morphological characters ($l = 100$, 200, 1000 from left to right, in increasingly dark shades of grey).

consistent with the outcomes of the date estimation described above. To examine the date estimates in further detail, we inspected the estimates for the youngest and median nodes. Posterior medians of the ages of the two nodes were close to the true values whether the fossils were pruned or not. However, the date estimates for the youngest node in the tree had smaller biases than those for the nodes with median ages (Supplementary Appendix 4 available on Dryad).

The large topological distances when the fossil taxa were retained in the maximum-clade-credibility trees revealed the difficulty in placing fossils correctly. Fossils with extant descendants were usually placed in the expected phylogenetic positions. Their recovery

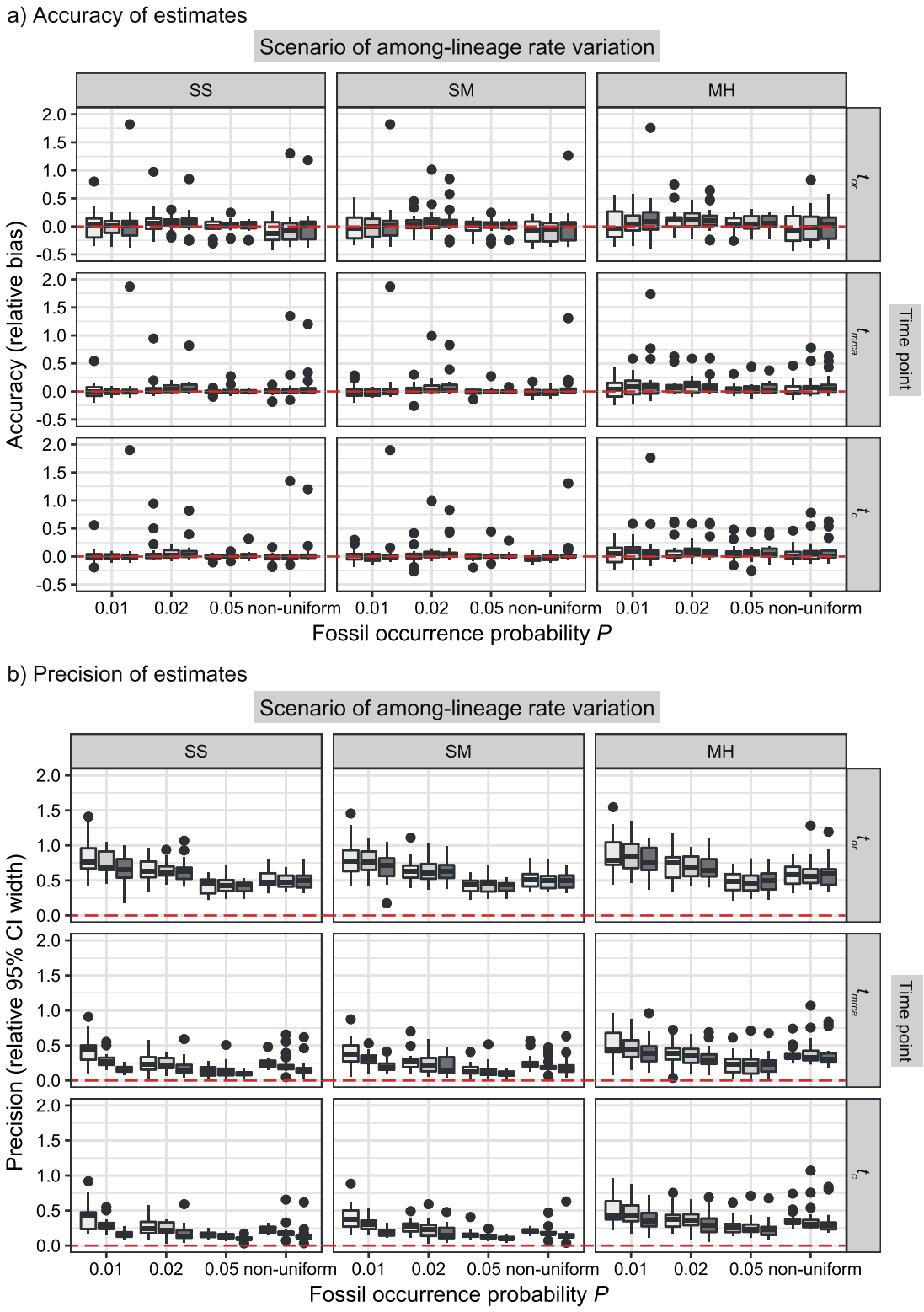

FIGURE 4. Performance of tip-dating in our core analyses in estimating origin time (t_{or}), root age (t_{mrca}), and crown age (t_c) Dashed horizontal lines indicate the target values in each of the plots. a) Accuracy of estimates, as measured by relative bias (distance between posterior median and true value, divided by the true value). b) Precision in estimates, as measured by relative 95% CI width (posterior 95% CI width divided by the true value). Each column of panels shows the results from a different model of among-lineage rate variation for the molecular and morphological data: strict clock and strict clock (SS); strict clock and moderate rate variation (SM); and moderate rate variation and high rate variation (MH). Within each panel, boxplot summaries are shown for the 20 FBD trees under each model of fossil occurrence probability ($P = 0.01$, 0.02, 0.05, and nonuniform). For each fossil occurrence probability, results are shown for three different sizes of morphological characters ($l = 100$, 200, 1000 from left to right, in increasingly dark shades of grey)

rates increased with l, but decreased with increasing probability of fossil occurrence in the models with uniform P. For the nonuniform P model, the recovery rates were similar to those when P = 0.02 (Fig. 3c). Among all sampled fossils, the ratios of the number of fossils placed as sampled ancestors to the true numbers of sampled ancestors were greater than 1.0 for 368 out of 720 cases, with this ratio increasing with l (Fig. 3d). Thus, for the fossils without extant descendants, the numbers of sampled ancestors were generally overestimated, while absolute numbers of sampled ancestors being placed incorrectly tended to increase with l (Supplementary Appendix 5 available on Dryad). These numbers of sampled ancestors were obtained from the maximum-clade-credibility trees, but the numbers increased considerably when based on the posterior medians from MCMC samples (the ratio of being sampled ancestors >1.0 for 63.5% of cases; Supplementary Appendix 5 available on Dryad).

We further evaluated potential differences across the species/FBD trees, with reference to the age estimates for key nodes (i.e., t_{or}, t_{mrca}, and t_c; Supplementary Appendix 6 available on Dryad). With respect to l (100, 200, and 1000) and different scenarios of among-lineage rate variation (SS, SM, and MH), we treated estimates under these conditions as independent replicates. Thus, we had nine repeats (3 × 3) for each of the 80 FBD trees. We found that the large overestimates of dates tended to occur for three particular species trees (Trees 6, 10, and 11; Table 1), which were among the four trees that had the most imbalanced topologies for extant taxa (corrected Colless index >3.5). However, this pattern was not always clear, especially when $P = 0.05$, and was only moderate for the relative 95% CI widths. The accuracy and precision of the estimates of t_{mrca} and t_c were broadly similar across the other species trees, for each value of P. For t_{or}, however, relative biases and relative 95% CI widths were found to decrease with true t_{or}, with the magnitude of these changes varying across the P models (Supplementary Appendix 6 available on Dryad).

We also evaluated potential differences across the species trees with respect to the performance of topological inference. When fossils were pruned, distributions of topological distances were uneven across the 20 species trees. The absolute Robinson–Foulds distances ranged from 0 (e.g., Trees 8 and 9) to 10 (Trees 3 and 14).

Effects of Excluding Morphological Characters

We performed two sets of analyses without morphological characters, either with or without constraints on the placements of the fossil taxa. When we used monophyly constraints to restrict the placements of the fossil taxa, inference of the tree topology showed similar performance to the core analyses (Supplementary Appendix 7 available on Dryad). However, there were slight improvements in the placement of fossil taxa, as reflected by smaller topological distances between inferred and true topologies (e.g., when P = 0.05, the median of the corrected Robinson–Foulds distances was <0.3 compared with >0.6 for the core analyses). The overestimation of the number of sampled ancestors was somewhat mitigated, with 96 out of 160 cases yielding sampled ancestor ratios not exceeding 1.0. Most of the fossils that left extant descendants were correctly placed (mean recovery rate 79.7% overall).

The accuracy of the posterior medians for t_{or}, t_{mrca}, and t_c was poorer than when morphological data were included, although with fewer instances of extreme overestimations (when relative bias > 1.0). There was a greater tendency for the posterior medians to exceed the true values, which occurred in 65.0%, 75.0%, and 83.8% of analyses for t_{or}, t_{mrca}, and t_c, respectively (Fig. 5a). Accuracy was slightly poorer when the sequence data had evolved with a moderate degree of rate variation among lineages. Coverage probabilities remained relatively high overall, however, with the 95% CIs containing the true values between 78% and 91% of instances for the three time points. Compared with the analyses that included morphological data, there was also a reduction in the precision of the date estimates (Fig. 5b). Using the gamma statistic and stemminess rank to summarize the overall estimates of relative node times, the inferred values generally matched those for the trees used for simulation. Posterior medians for the ages of the youngest and median nodes in the maximum-clade-credibility trees were close to the true values, but the lines of best fit had slightly greater slopes than in the core analyses, consistent with the overestimations of the three key time points t_{or}, t_{mrca}, and t_c described above (Supplementary Appendix 7 available on Dryad).

For the analyses in which morphological data were excluded and in which we did not specify any constraints on the tree topology, we experienced substantial problems with MCMC mixing and found that independent MCMC replicates usually failed to converge. These problems were most pronounced when the data included large numbers of fossil taxa. Consequently, we do not report the detailed results of these analyses here.

Effects of Excluding Molecular Data

We carried out analyses of the morphological data, with the nucleotide sequence data excluded. Topological inferences were generally similar to those of the core analyses. When fossils were pruned from the maximum-clade-credibility trees, corrected Robinson–Foulds distances from true topologies did not vary clearly across the P models (Fig. 6a). When fossils were retained, topological distances were larger and increased with P (Fig. 6b). In contrast with the results of the core analyses, however, larger l greatly reduced the topological distances whether the fossils were retained or not. Nevertheless, the distance estimates were greater overall than in the core analyses. For example, the overall

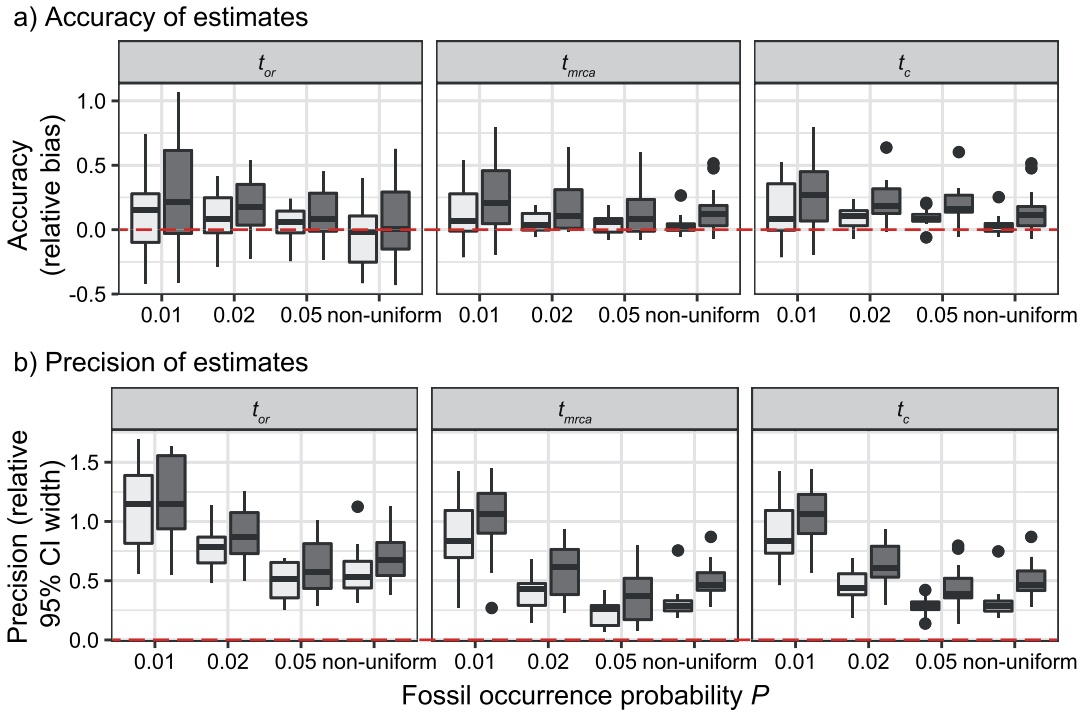

FIGURE 5. Posterior estimates for origin time (t_{or}), root age (t_{mrca}), and crown age (t_c) in the analyses when morphological data were excluded. For each fossil occurrence probability ($P = 0.01$, 0.02, 0.05, and nonuniform), the left boxplot (light grey shading) shows estimates for molecular data that have evolved under a strict clock, whereas the right boxplot (dark grey shading) shows estimates that have evolved under moderate rate variation across branches. Dashed horizontal lines indicate the target values in each of the plots. a) Accuracy of estimates, as measured by relative bias. b) Precision in estimates, as measured by relative 95% credibility interval width.

median of the corrected Robinson–Foulds distances was 0.40 when fossils were pruned, 10 times larger than that for the core analyses. Summaries of fossil positions in the maximum-clade-credibility trees did not differ much from those of the core analyses (Supplementary Appendix 8 available on Dryad).

The accuracy and precision of the posterior estimates for t_{or}, t_{mrca}, and t_c were generally consistent with those when nucleotide sequences were included, with coverage probabilities of 90.6%, 89.6%, and 83.8% for t_{or}, t_{mrca}, and t_c, respectively (Supplementary Appendix 8 available on Dryad). However, the summarized relative node depths via the gamma statistic and stemminess rank were less accurate under all degrees of rate variation, though the lines of best fit still had slopes close to 1.00. Whether the fossils were retained or not, posterior medians of the age estimates for median nodes were close to the true values. In contrast, there was a tendency to overestimate the ages of the youngest nodes in the trees (lines of best fit with slopes around 2.0; Supplementary Appendix 8 available on Dryad).

Effect of Fixing Tree Topology

We carried out analyses with both the morphological and molecular data, while fixing tree topologies during MCMC sampling. The overall ratio of sampled ancestors in maximum-clade-credibility trees to the true numbers of sampled ancestors was close to 1.00. In light of the fact that not all fossils that left extant descendants were recovered as sampled ancestors, overestimation of the number of sampled ancestors was still problematic sometimes for the fossils that did not have extant descendants (Supplementary Appendix 9 available on Dryad). The accuracy and precision of the estimates of t_{or}, t_{mrca}, and t_c were similar to those from the core analyses, except in two respects: there were almost no estimates with extreme relative biases; and the relative 95% CI widths were slightly smaller for t_{mrca} and t_c (Supplementary Appendix 9 available on Dryad). Estimates for the summarized relative node depths, youngest nodes, and median nodes were generally close to the true values.

Variations on the Conditions of the Core Analyses

We examined three variations on the conditions of the core analyses, with their counterparts from the core analyses being treated as the controls (Table 2). First, we replaced the binary morphological characters with four-state morphological characters. Second, we used the Mk model to analyze the full sets of morphological characters rather than using the Mkv model to analyze only the variable morphological characters. Neither of these variations led to any appreciable impacts on the accuracy and precision of date estimates, nor on the

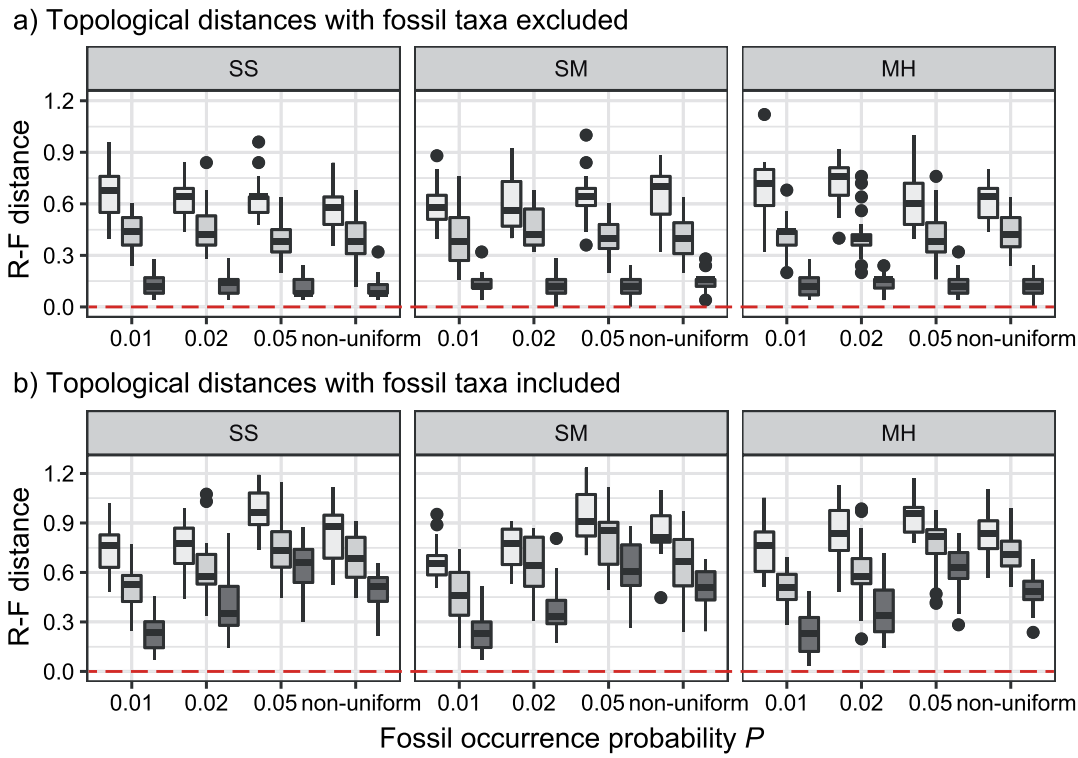

FIGURE 6. Posterior estimates in topological inference when molecular data were excluded. Dashed horizontal lines indicate the target values for the metrics in each of the plots. a) Corrected Robinson–Foulds distances between maximum-clade-credibility trees and true trees with fossil taxa excluded. b) Corrected Robinson–Foulds distances with fossil taxa included. Each panel shows the results from a different model of among-lineage rate variation for the molecular and morphological data: strict clock and strict clock (SS); strict clock and moderate rate variation (SM); and moderate rate variation and high rate variation (MH). Within each panel, boxplot summaries are shown for the 20 FBD trees for each model of fossil occurrence probability ($P = 0.01$, 0.02, 0.05, and nonuniform). For each fossil occurrence probability, results are shown for three different sizes of morphological characters ($l = 100$, 200, 1000 from left to right, in increasingly dark shades of grey).

estimates of the tree topology (Fig. 7). In contrast, our third variation on the conditions of the core analyses, which involved incorporating uncertainty in the fossil sampling times, resulted in slightly wider relative 95% CI widths in estimates for t_{mrca} (Fig. 7b). In this scenario, the inferred topologies of extant taxa did not differ substantially from the results of the core analyses. However, numerous discrepancies appeared when fossils were retained in the trees, and the absolute Robinson–Foulds distances decreased with l (Fig. 7d).

Estimates of FBD Model Parameters

Estimates of the net diversification rate (d), turnover rate (r), and fossil occurrence probability (s) varied across the different conditions for simulation and analysis explored in this study (Fig. 8). When the tree topologies were fixed, posterior medians were similar to those from the aforementioned model recovery in our evaluation of the FBD process. In other cases (e.g., our core analyses), r tended to be underestimated and s tended to be overestimated, whereas the estimates of d showed inconsistent variations from the true value.

The degree of evolutionary rate variation across branches did not produce clear impacts on the estimates of the three FBD parameters. However, the number of morphological characters (l) had a small impact in analyses conditioned on fixed tree topologies, but much larger effects in the other related analyses where d and s increased and r decreased with l. Compared with the results of our core analyses, using four-state morphological characters or using the Mk model to analyze the morphological data had no apparent impact on the estimates of the FBD parameters. When we incorporated uncertainty in the fossil ages, however, s and r were slightly overestimated and underestimated, respectively (Supplementary Appendix 10 available on Dryad).

DISCUSSION

Joint Estimation of Node Times and Tree Topology

Our simulation study provides a range of insights into the performance of Bayesian tip-dating with the FBD

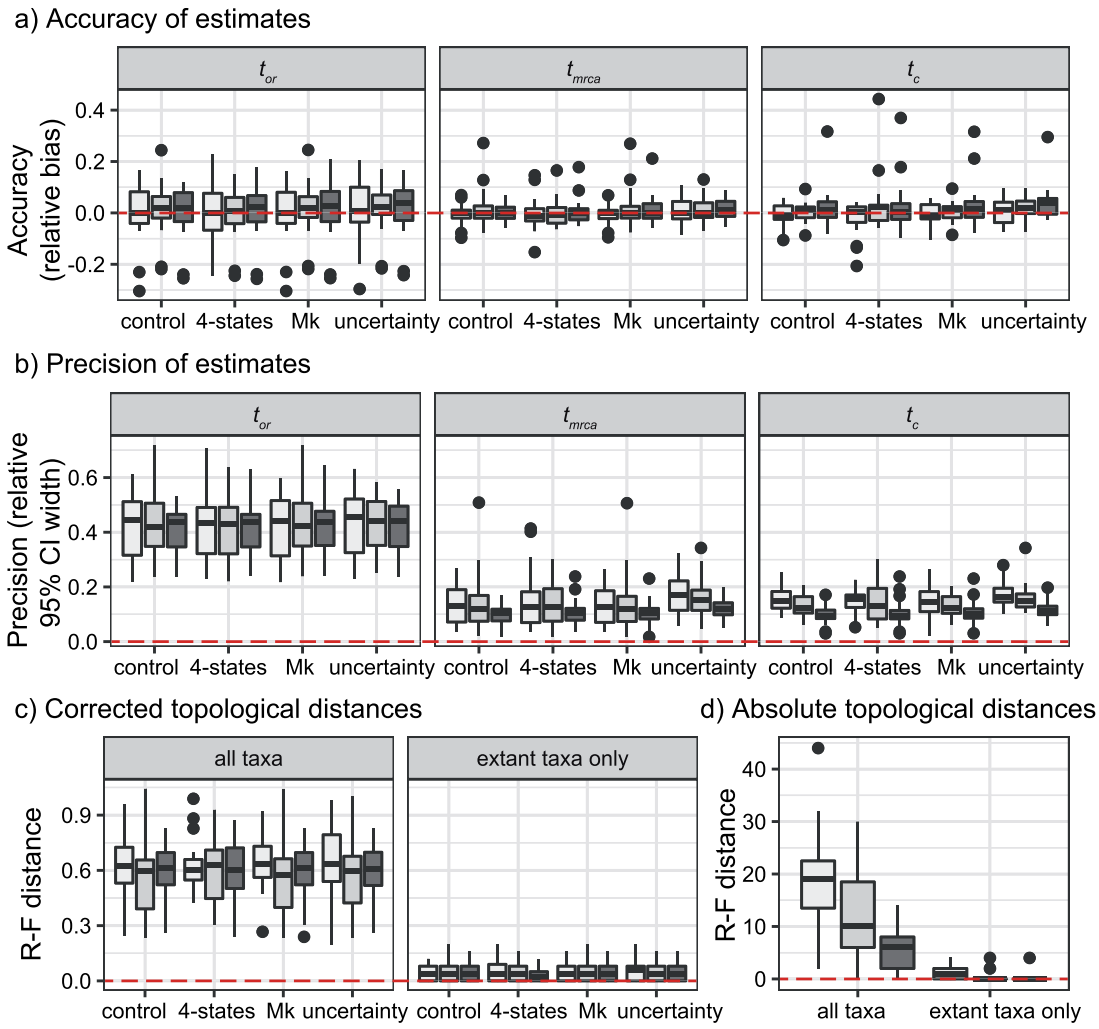

FIGURE 7. Performance of tip-dating under variations on the conditions of the core analyses. Results are shown for: counterpart analyses in the core analyses (denoted by "control"); those when binary morphological characters were replaced by four-state morphological characters (denoted by "four states"); those when the Mk model was used to analyse the full morphological data sets, rather than using the Mkv model to analyse only the variable morphological characters (denoted by "Mk"); and those when the uncertainty in fossil ages was taken into account (denoted by "uncertainty"). Dashed horizontal lines indicate the target values for the metrics in each of the plots. a) Accuracy of posterior medians, as measured by relative bias, for origin time (t_{or}), root age (t_{mrca}), and crown age (t_c). b) Precision in date estimates, as measured by relative 95% credibility interval width. c) Corrected Robinson–Foulds distances between the maximum-clade-credibility trees and the trees used for simulation. d) Absolute Robinson–Foulds distance between maximum-clade-credibility trees derived from control analyses and those derived from the analyses taking into account fossil age uncertainty, based on either all taxa or only extant taxa. For each of the four treatments within each panel in (a), (b), and (c) and for the two treatments in (d), boxplots summarize the results for three different sizes of morphological characters ($l = 100$, 200, 1000 from left to right, in increasingly dark shades of grey).

model. We found that divergence times were accurately estimated under most of the conditions investigated in our core analyses, with relative biases being close to zero and posterior medians approaching the true ages. Relative 95% CI widths were usually below 1.0, indicating a moderate degree of precision in the divergence-time estimates. Our use of these measures of performance is consistent with those in previous studies (e.g., Gavryushkina et al. 2014; Zhang et al. 2016), but different from the use of other metrics such as the coverage probability (Heath et al. 2014) and the mean of the posterior distribution (Warnock et al. 2017).

The results of our core analyses revealed that the estimates of the origin time of the FBD process (t_{or}) were less accurate and less precise than those of the ages of the root (t_{mrca}) and the crown group (t_c). This is consistent with expectations, given that taxa were not sampled during the interval between t_{mrca} and t_{or}, and that we used a diffuse prior for t_{or}. The association between estimates of t_{or} and other factors, such as the probability of fossil occurrence P and number of morphological characters l, was also different from those seen for t_{mrca} and t_c. These results highlight the difficulty in estimating t_{or}, even under the most benign conditions explored in

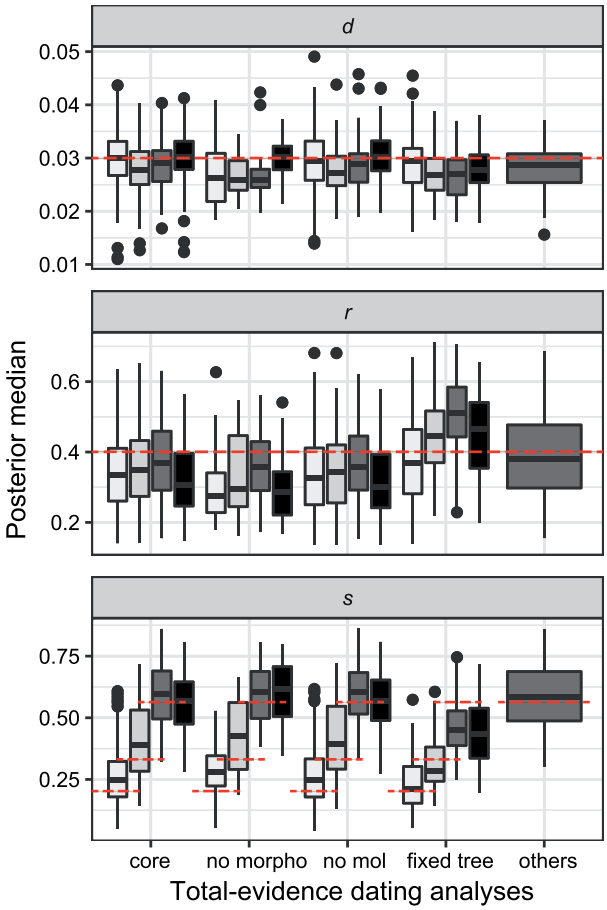

FIGURE 8. Posterior medians of the fossilized birth–death model parameters net diversification rate (d), turnover rate (r), and fossil sampling proportion (s) from all dating analyses. Results are shown for the core analyses ("core"), analyses without morphological characters ("no morpho"), analyses without nucleotide sequences ("no mol"), analyses conditioned on fixed tree topologies ("fixed tree"), and analyses under other variations on the conditions of the core analyses ("others"). Boxplots summarize the estimates from analyses grouped according to the P models used for simulation ($P = 0.01$, 0.02, 0.05, and nonuniform from left to right, in increasingly dark shades of grey). The dashed horizontal lines indicate the true values of d, r, and s that were used for simulation.

this study. However, this parameter is rarely of direct interest in tip-dating analyses, where there tends to be a much greater focus on the age of the root or the crown group.

Compared with the true values, the ages of deep nodes (e.g., the crown age t_c) were overestimated in 50–60% of cases in our core analyses. This suggests that tip-dating with the FBD model is not particularly susceptible to "deep-root attraction," a problem associated with unreasonably ancient divergence-time estimates (O'Reilly et al. 2015; Ronquist et al. 2016). In this respect, our results are consistent with the findings of previous studies (Herrera and Dávalos 2016; Zhang et al. 2016; Gavryushkina et al. 2017). However, we did observe some cases in which there were large overestimates of the ages of deep nodes. With reference to the results from differences in date estimates across the species/FBD trees and analyses with fixed tree topologies, these cases appeared to be the combined outcome of incorrect fossil placement and tree imbalance. This result partly echoes the findings of previous investigations of the impact of tree shape (Duchêne et al. 2015). Unfortunately, tree imbalance and problematic fossil placements are difficult to avoid in practice. Using informative priors that place a penalty on unobserved ghost lineages (i.e., lineages whose existence is postulated but which are unobserved in the fossil record) might help to mitigate the impacts of deep-root attraction on tip-dating (see Ronquist et al. 2016).

In Bayesian phylogenetic dating, the inferred evolutionary relationships are often also of interest. We found that tip-dating performed well in terms of inferring the relationships among extant taxa, which underscores the role of molecular data in providing a strong phylogenetic signal for these taxa. In contrast, the most notable problem observed in our analyses is the incorrect phylogenetic placement of fossil taxa in the maximum-clade-credibility trees. This outcome illustrates the challenges of topological inference and the resolution of deep nodes based on morphological characters, as identified in previous work (Puttick et al. 2017). We found that errors were reduced by increasing the number of morphological characters, but not by switching from binary to multistate character coding or including invariable sites.

Impacts of the Probability of Fossil Occurrence

A key benefit of tip-dating is that it allows the fossil record to be used more effectively in the estimation of evolutionary rates and timescales. However, the quality and completeness of the fossil record is subject to variations in depositional environments, taphonomy, time depth, and sampling intensity (Donoghue and Benton 2007; Holland 2016). Our simulations greatly simplified the process of fossil occurrence by ignoring the distinction between preservation potential and sampling intensity. Instead, the heterogeneous and incomplete nature of the fossil record was reflected in our use of different models of the probability of fossil occurrence, based on a single parameter P. The impact of this parameter on the estimates of divergence times and the tree topology was found to outweigh the effects of the other factors explored in our simulation study, such as the degree of among-lineage rate variation.

Our different models of the probability of fossil occurrence imparted multiple aspects of fossil occurrences along lineages in a birth–death species tree, which were associated with some distinct patterns among our results. With an increasing number of sampled fossils, the estimates of divergence times improved in precision. In our models with a low or nonuniform probability of fossil occurrence, there was

a smaller chance of sampling older lineages and older fossils. This effectively led to a reduction in the range of sampled fossil ages. In turn, there was a negative impact on the accuracy of estimates of the origin time t_{or} for the FBD process, but not of the age estimates for the root and crown group. Additionally, varying P can change the distribution of fossil occurrences on the tree topology, because imbalanced topologies tend to have a greater number of long terminal branches. The limited number of replicates and the stochasticity of the fossil sampling procedures in our study preclude further exploration and interpretation of this effect.

The probability of fossil occurrences had no apparent impacts on the inference of the relationships among extant species. This outcome is presumably due to the rich information content of the nucleotide sequences in our simulations. A somewhat unexpected pattern in our analyses, however, was that phylogenetic inferences tended to become worse overall with increasing numbers of fossils, due to the difficulty in placing fossil taxa correctly in the maximum-clade-credibility trees. For fossils without extant descendants, in particular, we found a tendency to overestimate the number of sampled ancestors. This potentially reduces the number of ghost lineages across the full species tree. These fossils sometimes clustered in a separate group in the maximum-clade-credibility tree, with their phylogenetic placements partly reflecting their ages (results not shown; O'Reilly et al. 2015; Donoghue and Yang 2016).

When we accounted for uncertainty in the ages of the fossil occurrences, we found a measurable decline in the precision of the divergence-time estimates. This differs from the results of previous analyses based on tip-dating methods, which found that incorporating uncertainty in the ages of ancient DNA samples did not noticeably affect the estimates of the root age (Molak et al. 2013). Moreover, we found that uncertainty in fossil ages affected the placement of these fossils in the maximum-clade-credibility trees, which supports the notion that the placement of fossils should be informed by both their ages and their morphological characters (O'Reilly et al. 2015; Lee and Yates 2018). Our investigation of the observed impacts of fossil-age uncertainty is based on the relatively narrow age ranges of the stratigraphic stages to specify the uncertainty in fossil sampling times. In reality, the uncertainty in fossil ages can be much greater and there are likely to be benefits in taking this uncertainty into account explicitly. Such a practice should lead to better estimates than simply using an arbitrarily chosen and potentially incorrect point value (e.g., the midpoint of the stratigraphic interval) to represent the fossil age (Barido-Sottani et al. 2019).

Dating with Restricted Data Sets or Conditions

The full potential of tip-dating with the FBD prior is realized in joint analyses of morphological and molecular data sets (i.e., total-evidence dating in Ronquist et al. 2012 and Zhang et al. 2016), but the method can also be used when only one of these types of data is available. In the absence of morphological data, the fossil occurrence times can be used to inform the FBD model (Heath et al. 2014). Although we found that this approach has the potential to mitigate the problem of deep-root attraction (Ronquist et al. 2016), the inclusion of monophyly constraints on the fossil occurrences was essential to the tractability of the dating analyses. Even with optimal constraints on monophyly, however, the divergence times tended to be overestimated. This was possibly because the fossil occurrences were still able to jump between placements on the stem lineage or into the crown group defined by the monophyly constraints. When a stem fossil is incorrectly placed in the crown group, the age of the crown group will tend to be overestimated. The extent of age overestimation was exacerbated by the presence of rate variation across branches.

The FBD model is increasingly being used to analyze data sets comprising only morphological characters (e.g., Bapst et al. 2016; Matzke and Irmis 2018). We found that the exclusion of molecular data led to large reductions in the performance of phylogenetic inference, which confirms the substantial challenges facing Bayesian phylogenetic inference with morphological characters alone. Our method of tree summarization using the maximum-clade-credibility topology might have led to less accurate performance than using majority-rule-consensus topologies for morphological data (O'Reilly and Donoghue 2018), but we chose to use the former for the sake of consistency across our analyses. Our results pointed to some decoupling of the inferences of divergence times and tree topology, given that date estimates for key nodes remained accurate even while the quality of topological inference declined. However, we did observe greater inaccuracy in estimates of relative node times, along with overestimation of the age of the youngest node in the tree.

When we fixed the tree topology, we found a substantial improvement in the performance of tip-dating. There were far fewer cases of extremely inaccurate estimates of node times, while the number of sampled ancestors for all sampled fossils was correctly recovered. The greatest improvements in performance were seen for highly imbalanced trees. However, given that the true tree topology is almost never known in practice, our results do not necessarily provide support for the sequential inference of the tree topology and divergence times when compared with a joint estimation procedure.

Morphological Characters in Fossil Tip-Dating

Total-evidence dating has gained considerable traction since the earliest studies were undertaken in the early 2010s (Pyron 2011; Ronquist et al. 2012). Although the availability of morphological data sets containing characters shared between living and fossil taxa has often been a limiting factor (Guillerme and Cooper

2016), recent studies have seen the assembly of relatively large data sets (e.g., 4541 characters in O'Leary et al. 2013). Our results highlight the benefits of using large morphological data matrices, which serves to improve the precision of node-time estimates and to reduce topological distance from the true tree. However, the quality of morphological character data is an additional consideration that is potentially important but outside the scope of our simulations here (Simões et al. 2017).

Our simulations of the evolution of morphological characters involved a number of simplifications (Goloboff et al. 2018; O'Reilly et al. 2018). For example, we assumed independence among characters, a simple model of character replacement, and relatively simple patterns of among-lineage rate variation (Wright et al. 2016). In reality, however, morphological traits are subject to a range of selective pressures and potentially display differing modes and degrees of evolutionary rate heterogeneity across lineages (Lee and Palci 2015; Goloboff et al. 2019). Additionally, morphological characters are often incompletely coded for fossil taxa (Sansom et al. 2010; Sansom and Wills 2013), but we did not include missing data in our simulations. Despite these simplifications, the key assumptions involved in our simulations were matched by those in the methods used to analyze the morphological data. The design of our study allowed us to avoid potential problems arising from model misspecification and model inadequacy, which are likely to be important problems for analyses of empirical data (Ronquist et al. 2016). A final point to note is that we only simulated the evolution of discrete morphological characters, but in recent years there have been promising developments in evolutionary models for continuous morphological characters (Felsenstein 1973; Parins-Fukuchi 2018; Álvarez-Carretero et al. 2019).

Macroevolution and the FBD Process

The FBD process provides a convenient tree prior for Bayesian analyses of combined paleontological and neontological data, but can itself also provide valuable information about the diversification rates of the lineages being studied. Our results from the evaluation of the FBD process showed good accuracy in recovering the parameters of the FBD model: the net diversification rate d, turnover rate r, and fossil sampling proportion$s$. However, perhaps partly due to the use of diffuse priors for these parameters, we found some unexpected variation in estimates of d and r for different probabilities of fossil occurrence. Biases in estimation of the FBD model parameters are potentially problematic because they can mislead interpretations of the macroevolutionary process. For example, using the formula $\lambda = d/(1-r)$ and $\mu = rd/(1-r)$ (Heath et al. 2014), we found that both speciation and extinction rates were somewhat underestimated across our analyses (overall means were 0.04 and 0.01, respectively). Among all of the dating analyses that we carried out, only those with fixed tree topologies yielded estimates of the FBD model parameters that were comparable to those obtained from our initial evaluations of the model. These results draw attention to the influence of the phylogenetic placements of fossil taxa and the numbers of sampled ancestors on the estimates of the FBD model parameters.

Conclusions

Fossil tip-dating offers a powerful means of understanding macroevolutionary processes, especially when coupled with the FBD process as a tree prior in Bayesian analysis. Using a simulation-based approach, we have performed a comprehensive evaluation of the performance of Bayesian tip-dating with the FBD process. Our results have demonstrated that the evolutionary relationships of extant taxa are well estimated, while the precision of divergence-time estimates tended to increase with the number of sampled fossils. However, we encountered considerable difficulty in identifying the correct phylogenetic placements of fossil taxa, even with good sampling of morphological characters.

Our study has revealed the considerable challenges posed by the absence of morphological data when analyzing a combination of extant and fossil taxa. Even though the FBD model can be used to infer evolutionary timescales using fossil occurrence times alone (Heath et al. 2014), the date estimates were sensitive to the presence of rate variation across branches even when topological constraints were applied to all fossils.

Overall, the results of our simulation study have demonstrated the general utility of the FBD model in tip-dating. Further studies involving comprehensive analyses of empirical data sets will provide deeper insights into the performance of these methods when using morphological and molecular data that have evolved under more complex conditions. Continued development and extension of the FBD model will help to unlock the potential of using the combined information in morphological, molecular, neontological, and paleontological data for resolving evolutionary timescales across the diversity of life.

Supplementary Material

Data available from the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.q2527ts.

Funding

This work was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (XDB31000000) and the Youth Innovation Promotion Association of the Chinese Academy of Sciences

(2017118). A.L. was funded by a visiting scholarship from the Chinese Academy of Sciences to carry out research at the University of Sydney. D.A.D. was funded by the Australian Research Council (DP160104173). C.Z. was supported by the 100 Young Talents Program of the Chinese Academy of Sciences. C-D.Z. acknowledges the support of the National Science Fund for Distinguished Young Scholars (grant number 31625024). S.Y.W.H. was supported by a Future Fellowship (FT160100167) from the Australian Research Council.

Acknowledgments

We acknowledge the University of Sydney's high-performance computing cluster, Artemis, for providing computing resources that have contributed to the research reported here. We also sincerely thank Matt Friedman, Thomas Guillerme, Rachel Warnock, Joëlle Barido-Sottani, and an anonymous reviewer for their constructive comments and suggestions that helped to improve this article

References

Álvarez-Carretero S., Goswami A., Yang Z., dos Reis M. 2019. Bayesian estimation of species divergence times using correlated quantitative characters. Syst. Biol. https://doi.org/10.1093/sysbio/syz015

Arcila D., Pyron R.A., Tyler J.C., Orti G., Betancur-R R. 2015. An evaluation of fossil tip-dating versus node-age calibrations in tetraodontiform fishes (Teleostei: Percomorphaceae). Mol. Phylogenet. Evol. 82:131–145.

Arcila D., Tyler J.C. 2017. Mass extinction in tetraodontiform fishes linked to the Palaeocene-Eocene thermal maximum. Proc. R. Soc. B. 284:20171771.

Bapst D.W., Wright A.M., Matzke N.J., Lloyd G.T. 2016. Topology, divergence dates, and macroevolutionary inferences vary between different tip-dating approaches applied to fossil theropods (Dinosauria). Biol. Lett. 12:20160237.

Barido-Sottani J., Aguirre-Fernández G., Hopkins M., 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. 286:20190685.

Bouckaert R., Heled J., Kühnert D., Vaughan T., Wu C-H., Xie D., Suchard M.A., Rambaut A., Drummond A.J. 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Comput. Biol. 10:e1003537.

Bromham L., Duchêne S., Hua X., Ritchie A.M., Duchêne D.A., Ho S.Y.W. 2018. Bayesian molecular dating: opening up the black box. Biol. Rev. 93:1165–1191.

Colless D.H. 1982. Review of phylogenetics: the theory and practice of phylogenetic systematics. Syst. Zool. 31:100–104.

De Baets K., Antonelli A., Donoghue P.C.J. 2016. Tectonic blocks and molecular clocks. Phil. Trans. R. Soc. B. 371:20160098.

Donoghue P.C.J., Benton M.J. 2007. Rocks and clocks: calibrating the Tree of Life using fossils and molecules. Trends Ecol. Evol. 22:424–431.

Donoghue P.C.J., Yang Z. 2016. The evolution of methods for establishing evolutionary timescales. Phil. Trans. R. Soc. B. 371:20160020.

dos Reis M., Donoghue P.C.J., Yang Z. 2016. Bayesian molecular clock dating of species divergences in the genomics era. Nat. Rev. Genet. 17:71–80.

Drummond A.J., Ho S.Y.W., Phillips M.J., Rambaut A. 2006. Relaxed phylogenetics and dating with confidence. PLoS Biol. 4:e88.

Drummond A.J., Stadler T. 2016. Bayesian phylogenetic estimation of fossil ages. Phil. Trans. R. Soc. B. 371:20150129.

Duchêne D., Duchêne S., Ho S.Y.W. 2015. Tree imbalance causes a bias in phylogenetic estimation of evolutionary timescales using heterochronous sequences. Mol. Ecol. Resour. 15:785–794.

Felsenstein J. 1973. Maximum-likelihood estimation of evolutionary trees from continuous characters. Am. J. Hum. Genet. 25:471–492.

Fiala K.L., Sokal R.R. 1985. Factors determining the accuracy of cladogram estimation: evaluation using computer simulation. Evolution. 39:609–622.

Gavryushkina A., Heath T.A., Ksepka D.T., Stadler T., Welch D., Drummond A.J. 2017. Bayesian total-evidence dating reveals the recent crown radiation of penguins. Syst. Biol. 66:57–73.

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

Goloboff P.A., Galvis A.T., Arias J.S. 2018. Parsimony and model-based phylogenetic methods for morphological data: comments on O'Reilly *et al*. Palaeontology 61:625–630.

Goloboff P.A., Pittman M., Pol D., Xu X. 2019. Morphological data sets fit a common mechanism much more poorly than DNA sequences and call into question the Mkv model. Syst. Biol. 68:494–504.

Guillerme T., Cooper N. 2016. Effects of missing data on topological inference using a total evidence approach. Mol. Phylogenet. Evol. 94:146–158.

Gustafson G.T., Miller K.B. 2017. Systematics and evolution of the whirligig beetle tribe Dineutini (Coleoptera: Gyrinidae: Gyrininae). Zool. J. Linn. Soc. 181:118–150.

Heath T.A., Huelsenbeck J.P., Stadler T. 2014. The fossilized birth-death process for coherent calibration of divergence-time estimates. Proc. Natl. Acad. Sci. USA. 111:E2957–E2966.

Heled J., Drummond A.J. 2012. Calibrated tree priors for relaxed phylogenetics and divergence time estimation. Syst. Biol. 61:138–149.

Herrera J.P., Dávalos L.M. 2016. Phylogeny and divergence times of lemurs inferred with recent and ancient fossils in the tree. Syst. Biol. 65:772–791.

Ho S.Y.W. 2014. The changing face of the molecular evolutionary clock. Trends Ecol. Evol. 29:496–503.

Ho S.Y.W., Duchêne S., Duchêne D. 2015a. Simulating and detecting autocorrelation of molecular evolutionary rates among lineages. Mol. Ecol. Resour. 15:688–696.

Ho S.Y.W., Phillips M.J. 2009. Accounting for calibration uncertainty in phylogenetic estimation of evolutionary divergence times. Syst. Biol. 58:367–380.

Ho S.Y.W., Tong K.J., Foster C.S.P., Ritchie A.M., Lo N., Crisp M.D. 2015b. Biogeographic calibrations for the molecular clock. Biol. Lett. 11:20150194.

Holland S.M. 2016. The non-uniformity of fossil preservation. Phil. Trans. R. Soc. B. 371:20150130.

Hug L.A., Roger A.J. 2007. The impact of fossils and taxon sampling on ancient molecular dating analyses. Mol. Biol. Evol. 24:1889–1897.

Jukes T.H., Cantor C.R. 1969. Evolution of protein molecules. In: Munro H.N., editor. Mammalian protein metabolism. New York: Academic Press. p. 21–123.

Kealy S., Beck R. 2017. Total evidence phylogeny and evolutionary timescale for Australian faunivorous marsupials (Dasyuromorphia). BMC Evol. Biol. 17:240.

King B., Qiao T., Lee M.S.Y., Zhu M., Long J.A. 2017. Bayesian morphological clock methods resurrect placoderm monophyly and reveal rapid early evolution in jawed vertebrates. Syst. Biol. 66:499–516.

Larson-Johnson K. 2016. Phylogenetic investigation of the complex evolutionary history of dispersal mode and diversification rates across living and fossil Fagales. New Phytol. 209:418–435.

Lee M.S.Y., Palci A. 2015. Morphological phylogenetics in the genomic age. Curr. Biol. 25:R922–R929.

Lee M.S.Y., Yates A.M. 2018. Tip-dating and homoplasy: reconciling the shallow molecular divergences of modern gharials with their long fossil record. Proc. R. Soc. B. 285:20181071.

Lepage T., Bryant D., Philippe H., Lartillot N. 2007. A general comparison of relaxed molecular clock models. Mol. Biol. Evol. 24:2669–2680.

Lewis P.O. 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. Syst. Biol. 50:913–925.

Matschiner M., Musilová Z., Barth J.M.I., Starostová Z., Salzburger W., Steel M., Bouckaert R. 2017. Bayesian phylogenetic estimation of clade ages supports trans-Atlantic dispersal of cichlid fishes. Syst. Biol. 66:3–22.

Matzke N.J., Irmis R.B. 2018. Including autapomorphies is important for paleontological tip-dating with clocklike data, but not with non-clock data. PeerJ. 6:e4553.

Matzke N.J., Wright A. 2016. Inferring node dates from tip dates in fossil Canidae: the importance of tree priors. Biol. Lett. 12: 20160328.

Molak M., Lorenzen E.D., Shapiro B., Ho S.Y.W. 2013. Phylogenetic estimation of timescales using ancient DNA: the effects of temporal sampling scheme and uncertainty in sample ages. Mol. Biol. Evol. 30:253–262.

Near T.J., Dornburg A., Friedman M. 2014. Phylogenetic relationships and timing of diversification in gonorynchiform fishes inferred using nuclear gene DNA sequences (Teleostei: Ostariophysi). Mol. Phylogenet. Evol. 80:297–307.

Ogilvie H.A., Vaughan T.G., Matzke N.J., Slater G.J., Stadler T., Welch D., Drummond A.J. 2018. Inferring species trees using integrative models of species evolution. bioRxiv https://doi.org/10.1101/242875.

O'Leary M.A., Bloch J.I., Flynn J.J., Gaudin T.J., Giallombardo A., Giannini N.P., Goldberg S.L., Kraatz B.P., Luo Z.X., Meng J., Ni X., Novacek M.J., Perini F.A., Randall Z.S., Rougier G.W., Sargis E.J., Silcox M.T., Simmons N.B., Spaulding M., Velazco P.M., Weksler M., Wible J.R., Cirranello A.L. 2013. The placental mammal ancestor and the post–K-Pg radiation of placentals. Science. 339:662–667.

O'Reilly J.E., Donoghue P.C.J. 2018. The efficacy of consensus tree methods for summarizing phylogenetic relationships from a posterior sample of trees estimated from morphological data. Syst. Biol. 67:354–362.

O'Reilly J.E., dos Reis M., Donoghue P.C.J. 2015. Dating tips for divergence-time estimation. Trends Genet. 31:637–650.

O'Reilly J.E., Puttick M.N., Pisani D., Donoghue P.C.J. 2018. Empirical realism of simulated data is more important than the model used to generate it: a reply to Goloboff *et al*. Palaeontology. 61:631–635.

Paradis E., Claude J., Strimmer K. 2004. APE: analyses of phylogenetics and evolution in R language. Bioinformatics. 20:289–290.

Parins-Fukuchi C. 2018. Use of continuous traits can improve morphological phylogenetics. Syst. Biol. 67:328–339.

Penny D., Hendy M.D. 1985. The use of tree comparison metrics. Syst. Zool. 34:75–82.

Popescu A., Huber, K.T., Paradis E. 2012 ape 3.0: new tools for distance-based phylogenetics and evolutionary analysis in R. Bioinformatics. 28:1536–1537.

Puttick M.N., O'Reilly J.E., Tanner A.R., Fleming J.F., Clark J., Holloway L., Lozano-Fernandez J., Parry L.A., Tarver J.E., Pisani D., Donoghue P.C.J. 2017. Uncertain-tree: discriminating among competing approaches to the phylogenetic analysis of phenotype data. Proc. R. Soc. B. 284:20162290.

Pybus O.G., Harvey P.H. 2000. Testing macro-evolutionary models using incomplete molecular phylogenies. Proc. R. Soc. B. 267:2267–2272.

Pyron R.A. 2011. Divergence time estimation using fossils as terminal taxa and the origins of Lissamphibia. Syst. Biol. 60:466–481.

R Core Team. 2017. R: a language and environment for statistical computing. v3.3.3. Vienna (Austria): R Foundation for Statistical Computing. Available from: URL https://www.R-project.org/.

Rambaut A., Drummond A.J., Xie Dong., Baele G., Suchard M.A. 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Syst. Biol. 67:901–904.

Rambaut A., Grassly N.C. 1997. Seq-Gen: an application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. Comput. Appl. Biosci. 13:235–238.

Robinson D.F., Foulds L.R. 1981. Comparison of phylogenetic trees. Math. Biosci. 53:131–147.

Ronquist F., Klopfstein S., Vilhelmsen L., Schulmeister S., Murray D.L., Pasnitsyn A.P. 2012. A total-evidence approach to dating with fossils, applied to the early radiation of the Hymenoptera. Syst. Biol. 61:973–999.

Ronquist F., Lartillot N., Phillips M.J. 2016. Closing the gap between rocks and clocks using total-evidence dating. Phil. Trans. R. Soc. B. 371:20150136.

Sansom R.S., Gabbott S.E., Purnell M.A. 2010. Non-random decay of chordate characters causes bias in fossil interpretation. Nature. 463:797–800.

Sansom R.S., Wills M.A. 2013. Fossilization causes organisms to appear erroneously primitive by distorting evolutionary trees. Sci. Rep. 3:2545.

Simões T.R., Caldwell M.W., Palci A., Nydam R.L. 2017. Giant taxon-character matrices: quality of character constructions remains critical regardless of size. Cladistics. 33:198–219.

Slater G.J. 2013. Phylogenetic evidence for a shift in the mode of mammalian body size evolution at the Cretaceous-Palaeogene boundary. Methods Ecol. Evol. 4:734–744.

Stadler T. 2009. On incomplete sampling under birth-death models and connections to the sampling-based coalescent. J. Theor. Biol. 261:58–66.

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

Stadler T. 2011. Simulating trees with a fixed number of extant species. Syst. Biol. 60:676–684.

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

Vea I.M., Grimaldi D.A. 2016. Putting scales into evolutionary time: the divergence of major scale insect lineages (Hemiptera) predates the radiation of modern angiosperm hosts. Sci. Rep. 6:23487.

Warnock R.C.M., Yang Z., Donoghue P.C.J. 2017. Testing the molecular clock using mechanistic models of fossil preservation and molecular evolution. Proc. R. Soc. B. 284:20170227.

Wright A.M., Lloyd G.T., Hillis D.M. 2016. Modelling character change heterogeneity in phylogenetic analyses of morphology through the use of priors. Syst. Biol. 65:602–611.

Yang Z. 1994. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. J. Mol. Evol. 39:306–314.

Zhang C., Stadler T., Klopfstein S., Heath T.A., Ronquist F. 2016. Total-evidence dating under the fossilized birth-death process. Syst. Biol. 65:228–249.

Zuckerkandl E., Pauling L. 1965. Evolutionary divergence and convergence in proteins. In: Bryson V., Vogel H.J., editors. Evolving genes and proteins. New York (NY): Academic Press. p. 97–166.