



O'Reilly, J., & Donoghue, P. (Accepted/In press). The effect of fossil sampling on the estimation of divergence times with the fossilised birth death process. *Systematic Biology*, [syz037]. <https://doi.org/10.1093/sysbio/syz037>

Publisher's PDF, also known as Version of record

License (if available):  
CC BY

Link to published version (if available):  
[10.1093/sysbio/syz037](https://doi.org/10.1093/sysbio/syz037)

[Link to publication record in Explore Bristol Research](#)  
PDF-document

This is the final published version of the article (version of record). It first appeared online via Oxford University Press at <https://academic.oup.com/sysbio/advance-article/doi/10.1093/sysbio/syz037/5498713>. Please refer to any applicable terms of use of the publisher.

## **University of Bristol - Explore Bristol Research**

### **General rights**

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:  
<http://www.bristol.ac.uk/pure/about/ebr-terms>

# The Effect of Fossil Sampling on the Estimation of Divergence Times with the Fossilized Birth–Death Process

JOSEPH E. O'REILLY\* AND PHILIP C. J. DONOGHUE

<sup>1</sup>*School of Earth Sciences, University of Bristol, Life Sciences Building, Tyndall Avenue, Bristol BS8 1TQ, UK*

\*Correspondence to be sent to: *School of Earth Sciences, University of Bristol, Life Sciences Building, Tyndall Avenue, Bristol BS8 1TQ, UK;*  
*E-mail: joe.oreilly@bristol.ac.uk.*

*Received 22 August 2018; reviews returned 7 May 2019; accepted 17 May 2019*

*Associate Editor: Simon Ho*

**Abstract.**—Timescales are of fundamental importance to evolutionary biology as they facilitate hypothesis tests of historical evolutionary processes. Through the incorporation of fossil occurrence data, the fossilized birth–death (FBD) process provides a framework for estimating divergence times using more paleontological data than traditional node calibration approaches have allowed. The inclusion of more data can refine evolutionary timescale estimates, but for many taxonomic groups it is computationally infeasible to include all available fossil occurrence data. Here, we utilize both empirical data and a simulation framework to identify approaches to subsampling fossil occurrence data that result in the most accurate estimates of divergence times. To achieve this we assess the performance of the FBD–Skyline model when implementing multiple approaches to incorporating subsampled fossil occurrence data. Our results demonstrate that it is necessary to account for all available fossil occurrence data to achieve the most accurate estimates of clade age. We show that this can be achieved if an empirical Bayes approach, accounting for fossil sampling through time, is applied to the FBD process. Random subsampling of occurrence data can lead to estimates of clade age that are incompatible with fossil evidence if no control over the affinities of fossil occurrences is enforced. Our results call into question the accuracy of previous divergence time studies incorporating the FBD process that have used only a subsample of all available fossil occurrence data. [Divergence time estimation; fossilized birth–death process; fossil sampling; molecular clock.]

Evolutionary timescales are crucial to inference of evolutionary rate and tests of extrinsic causes and consequences of intrinsic biological evolution. Traditionally, evolutionary trees were calibrated simply to the oldest fossil representative of living lineages, but these methods have been supplanted by molecular clock techniques, which are abstracted from direct inference of the geologic record (dos Reis et al. 2016). The fossilized birth–death (FBD) process (Stadler 2010; Heath et al. 2014) promises to refine the accuracy and precision of evolutionary timescales by departing from traditional node-calibrated molecular clock methodology, instead incorporating the distribution of fossil taxa through time as a means of calibrating the rate of the molecular clock. By incorporating fossil data in this way, the FBD process avoids many of the perceived shortcomings of node-calibrated molecular clock methodology, namely, the difficulty in properly constructing the time prior, summarizing the distribution of fossil occurrences as a parametric distribution, and the reduction of available paleontological data to the ages of a handful of fossils in calibrated clades (Ho and Phillips 2009; Pyron 2011; Heled and Drummond 2012; Parham et al. 2012; Ronquist et al. 2012; Warnock et al. 2012; O'Reilly et al. 2015; Warnock et al. 2015; O'Reilly and Donoghue 2016).

The FBD process achieves its advance by integrating the process of sampling fossil occurrences through time, into the standard birth–death process that is often used to obtain the probability of a given tree (Kendall 1948; Yang and Rannala 1997; Gernhard 2008). By allowing for the inclusion of fossil occurrences, the FBD process facilitates the inclusion of all pertinent paleontological data. But, for many clades, the number of available fossil occurrences will number in the thousands to the

millions; obviously such data will require subsampling if analyses are to be computationally feasible. Given the well-characterized effects of subsampling extant taxa for phylogenetic analysis using birth–death models (Yang and Rannala 1997; Hohna et al. 2011), it is likely that different approaches to subsampling fossil occurrence data in the FBD framework will affect both the accuracy and precision of estimated divergence times. However, the impact of fossil sampling strategies on the FBD process has not been explored. Despite this uncertainty, the FBD process has been widely adopted and applied to estimate divergence times across a range of taxonomic groups (Heath et al. 2014; Grimm et al. 2015; Bapst et al. 2016; Eguchi and Tamura 2016; Matzke and Wright 2016; Zhang et al. 2016; Gavryushkina et al. 2017; Pyron 2017; Saladin et al. 2017; Slater et al. 2017; Wright 2017).

The FBD process can be applied to data sets containing both the ages of fossil occurrences and their respective morphological character data (Gavryushkina et al. 2014), or alternatively, to data sets in which fossil occurrences are represented exclusively by their age (Heath et al. 2014). In the former case, the placement of fossil taxa within the topology is estimated given their morphological data and a model of morphological character change (Lewis 2001). In the latter case, the placement of fossil taxa is integrated over by numerical methods and the times of the divergences between extant taxa are the only ones estimated. These two methods can be considered as the resolved or the unresolved cases of the FBD process, respectively. The unresolved FBD is attractive for addressing clades with a rich fossil record but few morphological characteristics (e.g., many biostratigraphically important clades). Similarly, if the chronological distribution of fossil taxa with

morphological data is markedly different from that of all known fossil occurrences, then it seems that those handful of occurrences will provide a poor approximation of the true rate of fossil sampling through time. As the rate of sampling is a fundamental parameter of the FBD process, in such situations it may be beneficial to use the unresolved FBD process. For the unresolved case, it is also possible to constrain the phylogenetic placement of fossil occurrences when estimating divergence times; this allows integration over fossil placement to be realistically restricted. If some prior information about the phylogenetic affinity of a fossil occurrence is available then this can be used to construct a topological constraint by assigning the occurrence to the node that defines the clade that the fossil is a member of. It is also possible to treat the occurrence as a crown group member only, or to allow it to be resolved in the crown or stem groups (i.e., as a total group member). It is also possible to make no assumption about the affinity of an occurrence and allow it to be resolved anywhere that is descendent of the root. An example of an implementation of these constraints is provided in Figure 1a.

The methodological developments manifest in the FBD process allow for the incorporation of far more data in divergence time estimation analyses than was previously possible. In the case of the resolved FBD process, the primary limiting factor on the number of fossil occurrences that can be included is the availability of morphological character data for those occurrences. For the unresolved case, we simply need to know the ages of the fossil occurrences to include them, which is a far simpler task than assembling a morphological character data set and is less likely to be impacted by missing character data (Sansom and Wills 2013). For clades containing only a few fossil occurrences it should be possible to include most, if not all, available fossil occurrence data. If the FBD process is applied to deeper timescales, or clades with a rich fossil record, then the number of fossil occurrences that can potentially be included grows rapidly, to the point that any analysis attempting to use all available data will become impractical (Matschiner et al. 2017). In such cases, there is no recourse except to subsample the fossil occurrence data and include only a fraction of the available occurrences, but it is not obvious how this subsample should be constructed. A similar issue is encountered when considering which extant taxa to include in an analysis (Hohna et al. 2011). As we go deeper into the tree of life the number of taxa that can be included grows rapidly and a choice must be made regarding which taxa to include. Birth–death process corrections have been developed to account for random (Yang and Rannala 1997) and nonrandom subsamples of extant taxa (Hohna et al. 2011; Zhang et al. 2016), but there are no analogous corrections for nonrandom subsamples of fossil occurrence data. Therefore, since the FBD process relies on the waiting times between fossil occurrences to be informative (Stadler 2010; Gavryushkina et al.

2014; Heath et al. 2014), any approach to subsampling must maintain the structure of the distribution of fossil occurrences through time.

Given the decidedly nonuniform rate of fossil sampling (Holland 2016), it is not intuitively obvious how a subsample should be constructed to result in the most accurate age estimates. Subsampling fossil occurrences with uniform probability seems likely to preserve the structure of fossil sampling while proportionally altering the rate, but whether or not the oldest or youngest members of clades also need to be included in such a subsample is unclear. If the placement of fossil occurrences is constrained correctly then the age of the oldest member of a clade will effectively provide a minimum bound on the age of the clade. If a uniform subsample of occurrences is taken then there is no guarantee that the oldest fossil member of each clade will be included. In such cases the minimum possible age of the clade will be moved to a younger fossil or, if no fossils from the clade are sampled, removed entirely. If this occurs, it is possible for age estimates to extend to times that are younger than the oldest fossil occurrence belonging to the clade, even though evidence is available to better constrain clade age. This phenomenon is demonstrated in Figure 1b. It is therefore possible that including the oldest fossils assigned to each node in the tree in a subsample will enforce realistic age estimates.

When dealing with subsamples from the fossil record, the FBD process must estimate fossil sampling rate with only a fraction of the available fossil occurrence information. In such cases, it may be best to employ an empirical Bayes approach, in which a maximum-likelihood estimate (MLE) of at least one parameter of the model is obtained *a priori* using the available data (Casella 1985; Kolaczkowski and Thornton 2007). The parameter of interest can then be fixed to this MLE or an informative prior can be placed on it. For the case of estimating divergence times with the FBD process, this can be achieved by obtaining MLEs of fossil sampling rate using all available occurrence data. Similarly, when there is obvious heterogeneity in the fossil sampling rate, an empirical Bayes approach with MLEs of fossil sampling rate within time slices could be employed. This approach deviates slightly from the traditional approach to empirical Bayesian analysis as the fossil sampling rate is estimated from a set of data that is subsequently subsampled before the estimation of the other model parameters.

Which application of the FBD process results in the most accurate divergence time estimates when subsampling fossil occurrence data is currently unknown. Through the analysis of both empirical and simulated data, we establish the relative performance of different applications of the unresolved FBD process when fossil occurrence data are subsampled.

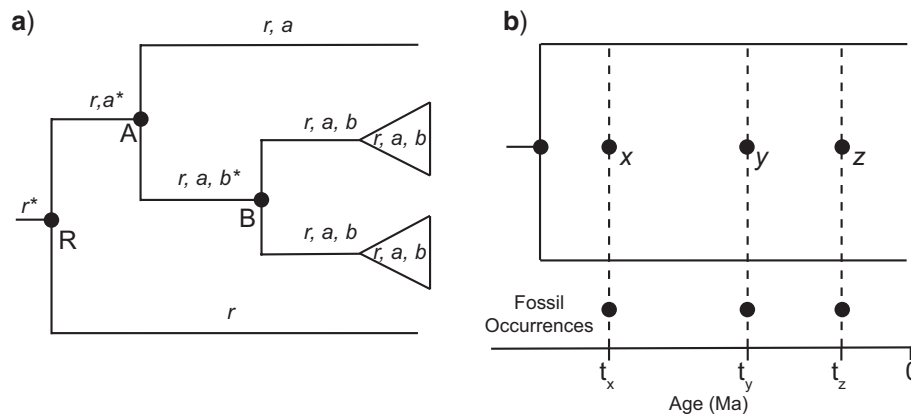


FIGURE 1. An example of topological constraints on fossil occurrence placement (a), and their effect on the minimum bound on the age of clades (b). Fossil occurrences can be assigned to a specific node in the tree (A, B) or to the root (R), depending on this assignment the fossil can then be placed on any descendent lineage of this node. For example, for any node in the tree of (a) represented by an uppercase letter, the possible lineages in which the fossils assigned to that node can be placed are presented with the corresponding lower case letter (e.g., fossils assigned to node “R” can be placed on any lineage possessing “r”). It is also possible to construct a constraint on fossil occurrence placement such that an occurrence can also be resolved in a total-group position (i.e., directly ancestral to the node it has been assigned to). The additional positions that occurrences assigned to A, B, or R can be resolved if a total-group constraint is employed are represented with (\*). The inclusion of different fossil occurrences in a subsample can alter the hard minimum bound on the age of a clade (b). If topological constraints are applied to fossil occurrences then the oldest occurrence in each clade will act as a minimum bound on its age. This is because the clade can be no younger than the age of its oldest descendent fossil ( $t_x$  in this case). If the subsample does not include the oldest occurrence (fossil  $x$ ), then the hard bound on the minimum age of the clade will be pushed forward in time to  $t_y$  or  $t_z$  depending on which fossil is included in the subsample. If no fossil is sampled in this clade then there will be no minimum bound on its age, despite the fact that any age younger than fossil  $x$  is incompatible with the fossil record.

## MATERIALS AND METHODS

### Empirical Analyses

We used the hymenopteran data set employed by Ronquist et al. (2012) and Zhang et al. (2016) to demonstrate the effects of different approaches to incorporating and constraining fossil occurrence data in FBD analyses. We removed extant nonhymenopteran taxa from the fixed-tree topology used by Ronquist et al. (2012) and fixed the relationships between extant taxa to this reduced tree for all analyses. To obtain fossil occurrence data we used the Paleobiology Database API (Peters and McClennen 2016) to collect any occurrence assigned to the order Hymenoptera. Of these occurrences, all of those older than 200 Ma were also assigned to specific families within crown Hymenoptera, suggesting that none of the collected occurrences were stem hymenopterans. Fossil occurrence data were uniformly subsampled (1%) and analyzed with the rates of fossil sampling fixed to an *a priori* estimate obtained from the full set of occurrences. The random subsample of fossil occurrence data was then supplemented with the oldest unequivocal occurrences that could be assigned to seven key hymenopteran clades. These oldest occurrences were taken to be the fossil taxa that defined the minima of the hymenopteran node calibrations in Ronquist et al. (2012).

The fixed rate of fossil sampling approach was also combined with three separate methods for constraining the position of fossil occurrences. For the first approach, there was no constraint on fossil placement, with fossils being placed anywhere in the tree without restriction.

In the second approach, each fossil occurrence was constrained to total-group membership of the most derived clade it is assigned to in the Paleobiology Database (PBDB) that was also present in the fixed tree topology. The restriction of a total-group placement for a fossil means that it can be resolved in either a stem or crown group position. In the third approach, the seven node calibration fossils were constrained to be resolved only in clades that were descendants of the direct ancestral node of the occurrence, effectively forcing them to be crown members of the clade they were assigned to and consequently providing hard minimum age constraints. All other fossil occurrences were constrained to be total-group members as in the second approach. For each of these three separate approaches, 10 replicate data sets were obtained, each consisting of a unique uniform subsample of the total available fossil occurrence data. We also performed a further set of analyses in which an arbitrary subsample of fossil occurrences are included; for these analyses, we included only those that inform the minimum age constraints of node calibrations in Ronquist et al. (2012). We also performed all of the above analyses with the rate of fossil sampling as an estimated model parameter or with the rate of fossil sampling assigned an informative prior based on the *a priori* estimate of fossil sampling, but initial tests showed that these analyses suffered from poor Markov chain Monte Carlo (MCMC) mixing for birth and death parameters of the FBD model.

To obtain a MLE of the rate of fossil sampling the method of Solow and Smith (1997) was employed, as implemented in the PaleoTree R Package (Bapst 2012).

Cutting points for the skyline FBD process were applied at 66 Ma and 252 Ma, after Zhang et al. (2016). Unique values for fossil sampling rate within each time slice were estimated by binning fossil occurrences by strata and applying the Solow and Smith (1997) method to these bins individually. Estimates of fossil sampling rate were inferred from the total set of occurrences retrieved from PBDB. The MLE of fossil sampling rate was multiplied by the proportion of the total number of fossil occurrences included in the subsample (0.01) to account for the increased waiting times induced by subsampling.

To analyze these data sets we employed the FBD skyline model in BEAST2 (Stadler et al. 2013; Bouckaert et al. 2014; Gavryushkina et al. 2014) using the Sampled Ancestors package (Gavryushkina et al. 2014) and the bdsky package (Stadler et al. 2013). Birth rate and death rate were each assigned  $U(0, \infty)$  priors. Birth and death rates were modeled as time homogeneous, but fossil sampling rate was modeled as being time heterogeneous, with time slices at 66 Ma and 252 Ma, as employed by Zhang et al. (2016). Molecular data was partitioned as in Zhang et al. (2016), with the GTR+I+G model applied to each partition and the substitution model parameters of each partition being unlinked. A single uncorrelated lognormal clock model was shared by all partitions with rate multipliers modeling the relative rate between partitions. The prior on the mean rate was an exponential distribution with a mean of 0.0008, which is equal to the mean of the prior used by Ronquist et al. (2012). The proportion of sampled extant taxa was fixed at 0.0005, as in Zhang et al. (2016). For each occurrence, age uncertainty was accounted for with a uniform prior with minimum and maximum bounds defined by the occurrence's entry in PBDB. The FBD process was conditioned on rho-sampling. For each analysis a uniform prior was placed on the origin of the FBD process with a minimum equal to the maximum age of the oldest fossil occurrence included in the analysis, and a maximum equal to the minimum age (302 Ma) of the more inclusive holometabolan clade (Ronquist et al. 2012). Each analysis consisted of 125,000,000 MCMC generations, sampling every 12,500. The resulting data sets each possessed 5096 nucleotide sites for 58 extant taxa and 51 fossil occurrences with no molecular data.

Sufficient sampling of the posterior distribution was assessed with the R package coda (Plummer et al. 2006), with an effective sample size of 100 for each parameter after a 25% burn-in considered to be indicative of a sufficient sample from the posterior distribution. We combined the 10 individual replicates together and obtained the resulting aggregate distributions of the crown group age estimates. We also obtained the widths of the 95% highest posterior density (HPD) for these age estimates.

#### Simulation Analyses

We used a simulation framework to assess the relative accuracy and precision of different approaches to

incorporating subsamples of a structured record of fossil occurrences in FBD analyses. The simulation framework also facilitates tests of the effect of constraining the placement of fossil occurrences in the tree and, further, the influence of relaxing these constraints. Replicate data sets consisting of a simulated tree, fossil occurrences, and extant molecular data were generated. Fossil occurrence data were subsampled and analyzed alongside the simulated extant molecular data following the three separate approaches of accounting for fossil sampling rate that were applied to the empirical data. These approaches were also combined with the three methods for constraining the position of fossil occurrences also employed previously in our analysis of the empirical data. In the simulation framework, we identified which fossil occurrences were members of each crown group without error and this allowed us to apply topological constraints such that occurrences were sampled as crown group members of their true clade only, without allowing them to be sampled on the stem of that clade. Another case was also considered in which the constraints on the placement of fossil taxa were incorporated with decreasing precision such that the assigned nodes of all occurrences were 1, 2, or 4 nodes ancestral to their true ancestral node, as three separate cases. Figure 2 demonstrates how these constraints on fossil occurrence placement were constructed. One further case in which a uniform subsample of fossil occurrences was analyzed was also explored, in which the rate of fossil sampling in each time slice was fixed to a random value drawn from a  $Uniform(0, 10)$  distribution, allowing for the fixing of sampling rate to values that are far larger than the true generating rate. This case was used to identify the importance of accurate *a priori* estimates of fossil sampling rate in an empirical Bayes approach.

To investigate the effect of including an arbitrary subsample of fossil occurrences we explored two further subsampling regimes, one in which only the oldest fossil member of each clade was included in analyses, and one in which only the youngest fossil that could be assigned to each node in the tree was included in analyses. Both of these cases were analyzed with all three approaches to accounting for fossil sampling rate and with both no constraint on the placement of fossil occurrences and the constraint of crown group membership of fossil occurrences as two separate cases. For each of these approaches 100 unique simulated data sets were generated. An overview of all the combined approaches to constraining the placement of subsampled simulated fossil occurrences and accounting for the rate of fossil sampling applied in this study are presented in Table 1.

#### Simulating trees and fossil occurrence data

Each replicate involved the simulation of a unique tree topology, molecular sequence data for extant taxa, and fossil occurrence data. To generate trees the TreeSim package (Stadler 2011) was used to simulate complete

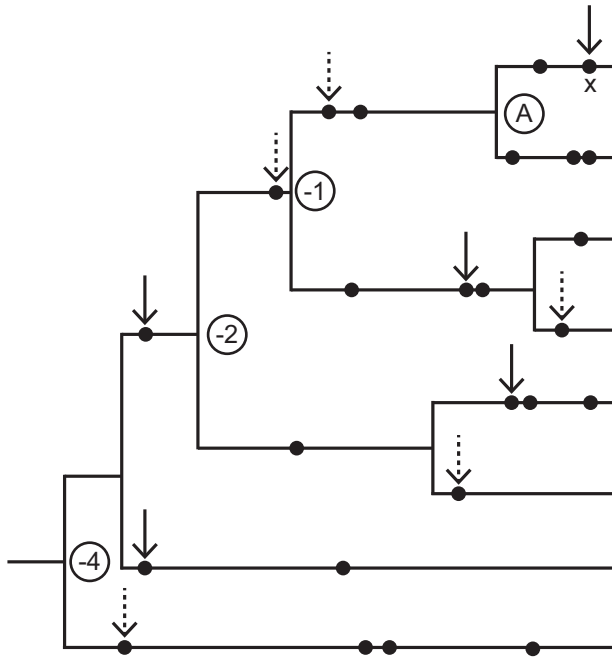


FIGURE 2. An example of the construction of the multiple approaches to subsampling fossil occurrence data, and the constraints on their topological placement, employed in the simulation and analysis framework. Fossil occurrences are represented by black circles on the lineages of the tree, those included in a 5% uniform subsample are highlighted with solid black arrows. To demonstrate the different subsampling and constraint procedures we consider the placement of the sampled fossil marked “x”. The position of fossil “x” can be entirely unconstrained by assigning it to the root of the tree, this is the first approach to constraint applied in the analysis of simulated data. Alternatively, the fossil can be assigned to its directly ancestral node “A” and only sampled on lineages that are descendants of “A”, this is the second approach applied in the analyses. In addition to the 5% subsample, the oldest members of each clade in the tree can be sampled too. These occurrences are highlighted with dashed arrows. Including these fossils in analyses in which the placement of occurrences is constrained to descendants of their directly ancestral node is the third approach applied in the analyses. Finally, subsampled occurrences (solid arrows) and oldest occurrences (dotted arrows) can be placed in the tree with reduced precision, at either 1, 2, or 4 nodes below their directly ancestral node, as three separate cases. For fossil “x” the nodes it is assigned to for these approaches are marked “-1”, “-2”, and “-4”. This approach is the fourth one applied in the analysis of simulated data.

topologies conditioned on 500 extant taxa, with a birth rate of 1.5 and a death rate of 0.5, with both the extant and extinct lineages maintained. Only tree topologies with extant descendants on both sides of the root bifurcation were retained. This procedure results in no stem lineages in the simulated trees. The model of Holland (1995) was used to simulate fossil occurrences on the tree of 500 extant tips. This method uses a number of parameters to model the probability of fossil taxa being preserved and sampled through time and then converts these probabilities to Poisson sampling rates. To achieve this, the model uses water depth as a sampling proxy, with the depth following a sine wave function through time. The model also uses three further parameters to define sampling probability: the peak abundance, the preferred depth, and the depth tolerance. A Gaussian distribution

TABLE 1. An overview of the approaches to constraining the placement of fossil occurrences and accounting for the rate of fossil sampling applied when analyzing simulated data sets in this study

Fossil subsample	Fossil placement constraints	Fossil sampling treatment	Placement imprecision
Uniform	NC, CC	F $\psi$ , E $\psi$ , I $\psi$ , R $\psi$	0, 1, 2, 4
Uniform + oldest clade members	CC	F $\psi$ , E $\psi$ , I $\psi$	0, 1, 2, 4
Uniform + oldest clade members for five clades only	CC	F $\psi$ , E $\psi$ , I $\psi$	0
Oldest clade members only	NC, CC	F $\psi$ , E $\psi$ , I $\psi$	0
Youngest clade members only	NC, CC	F $\psi$ , E $\psi$ , I $\psi$	0

Note: CC = crown constraint; E $\psi$  = estimated $\psi$ ; F $\psi$  = fixed $\psi$ ; I $\psi$  = informative prior; NC = no constraint; R $\psi$  = fixed random value. Placement imprecision denotes the number of nodes ancestral to the true topological placement that fossil occurrences were assigned to for the purposes of numerical integration over the phylogenetic relationships amongst fossil occurrences and extant taxa.

is used to model the probability of collection, with the depth tolerance defining the standard deviation of the distribution and the peak abundance defining the upper limit on the probability of sampling. Importantly, this method assumes time heterogeneous fossil sampling rates.

Fossil occurrence data sets were simulated using this model as implemented in the FossilSim R package (Barido-Sottani et al. 2018). For each simulation replicate independently drawn unique parameter values were obtained. The number of strata modeled in each replicate was randomly sampled from the range 3 to 6; the preferred depth parameter was drawn from a Uniform(2, 2.75) distribution, depth tolerance was drawn from a Uniform(2, 2.25) distribution and peak abundance was fixed to 2.25. The procedure was repeated until at least 1000 fossil occurrences were generated for each replicate. The time intervals of all strata were equal and were obtained by dividing the time scale of the generating tree by the number of strata. Fossil occurrences were not simulated along the root edge of the tree.

After fossil occurrence data were simulated, a uniform random sample of 25 extant taxa was extracted from the 500 tip tree. This reduced tree represents a random 5% subsample of extant taxa. Fossil occurrences on lineages that were not present in the reduced tree were assigned to their most recent ancestral node present in this topology. If the most recent common ancestor (MRCA) of the 5% subsample was not the MRCA of the 500 tip tree, the origin of the birth–death process was considered to be the age of the node ancestral to the 5% subsample MRCA. Otherwise, the origin as simulated by TreeSim was maintained.

#### Simulating molecular data for extant tips

Molecular data for extant taxa were simulated on the 25 tip tree using the software package Seq-Gen (Rambaut and Grassly 1997). Sequences of 1000 sites

were simulated with the GTR+G model. The transition rate parameters were drawn from a Dirichlet distribution with all concentration parameters set to 2. The nucleotide frequencies were drawn from a separate Dirichlet distribution with all concentration parameters set to 5, and the single parameter of the four category discretized Gamma( $a = b$ ) distribution constraining among site rate variation was drawn from a Gamma(2, 2) distribution for each replicate. Branch rate heterogeneity was modeled with rates being drawn from a log-normal distribution with a standard deviation on the log scale of 0.2 and a mean of 0.05. As Seq-Gen offers no intrinsic method to simulate relaxed clock sequence data, the branch lengths of the simulated tree were rescaled by their sampled branch rates before being used by Seq-Gen.

#### *Subsampling fossil occurrence data*

Small subsamples of each simulated fossil occurrence data set were obtained to model the necessary subsampling of data that will be encountered in empirical analyses. For the case of uniform random subsampling, fossil samples within each slice had a uniform probability of being included in the final subsample, with a 5% subsample being taken from the complete set of occurrences.

A second subsampling framework was also employed in which a random subsample (obtained as above) was supplemented with the oldest fossil occurrences for each node in the tree of extant taxa. These extra fossil occurrences are analogous to the unequivocally oldest fossil members of clades used to define the hard minima of node calibrations. When the FBD process is applied to empirical data it is unlikely that it will be possible to identify the oldest unequivocal member of every crown group present in an analysis. To account for this phenomenon we also analyzed subsamples which included the oldest fossil occurrences for only five of the 24 crown groups in each simulated topology. We employed one further subsampling approach in which the final subsample consisted of either the oldest or youngest occurrences exclusively for each node in the tree as two separate cases of arbitrary subsampling.

#### *Maximum-likelihood estimates of fossil sampling*

To obtain a MLE of the rate of fossil sampling, the method of Solow and Smith (1997) was employed again, as with the empirical analyses. Unique values for fossil sampling rate within each time slice were estimated by binning fossil occurrences by strata and applying the Solow and Smith (1997) method to these bins individually. This approach to estimating sampling rate requires fossil occurrences to be collected into taxonomic groups. To achieve this, all fossils that occurred on a single branch of the complete 500 tip tree were considered to belong to a single species and were therefore placed together as a single group. The estimates of fossil sampling rate were inferred from the

total set of simulated fossil occurrences. The increased waiting times between occurrences was accounted for by multiplying the estimates of sampling rate by the proportion of sampled fossils (0.05).

#### *Divergence time estimation analysis in BEAST2*

All analyses were performed in BEAST2 (Bouckaert et al. 2014) using the Sampled Ancestors package (Gavryushkina et al. 2014) and the bdsky package (Stadler et al. 2013). For all analyses, the relationships between extant taxa were fixed to the 25 tip simulated topology, and the skyline FBD process was applied as the tree prior. All fossil occurrences were assigned their true age, without error. The GTR+G substitution model was applied to the sequence data. An exponential prior with a mean of 0.05 was placed on the mean of the clock rate, and a G (2, 0.1) distribution was placed on the standard deviation of the log-normal relaxed clock. Improper U(0, infinity) priors were placed on the birth and death parameters. In the BEAST2 implementation of the skyline FBD process, there is a single prior distribution applied for the rate of fossil sampling within each of the time slices. For the analysis of simulated data, three approaches to modeling the rate of fossil sampling were applied: 1) leaving fossil sampling rate as an estimated parameter of the skyline FBD process with a U(0, infinity) prior, 2) fixing the rate of fossil sampling in each time slice to the MLE values estimated *a priori*, or 3) using the mean of the MLE estimates as the mean of an exponential prior distribution of fossil sampling rate. The extant sampling proportion was fixed to its true value of 0.05 and the age of the origin was fixed to its true value. The timings of changes in sampling rate were also fixed to the true values the fossil occurrence data was generated with. The birth and death rates were modeled as time homogeneous, matching the simulation protocol. The skyline FBD process was conditioned on rho sampling and the age of the origin, paralleling our simulation protocol. All analyses were performed for 25,000,000 MCMC generations, sampling every 2500. A postrun burn-in of 25% was employed, resulting in 7500 samples per analysis.

#### *Postrun analysis*

Stationarity was assessed with the R package coda (Plummer et al. 2006), discarding any replicate analyses that did not achieve an effective sample size of 100 or more for the posterior probability. This filtering procedure results in a variable number of replicates across the different analytic frameworks. For each remaining replicate, the median estimate of each node age in the extant tree was obtained, in addition to the width of the 95% HPDs of these model parameters. To quantify error in the age estimates, the medians of the posterior age estimates were subtracted from the simulation values; the absolute value of this quantity was also obtained. For each node, accuracy was quantified

with the coverage metric, which is measured as the percentage of replicates that result in a 95% HPD that encompasses the simulated age. In a traditional node calibration framework, the oldest unequivocal member of a clade (the calibrating fossil) is applied as the hard minimum age constraint on the calibration distribution. As this value represents an age by which the clade must have appeared, there should be no positive posterior probability assigned to ages younger than this. For a subsampled FBD analysis, there is no requirement to include calibration fossils, and it is therefore possible that the posterior estimates of node age may be younger than the oldest descendant of the node. For this reason, we also determined the proportion of node age estimates with HPDs encompassing ages younger than their oldest descendant. As not all analyses reached the required threshold for stationarity the number of analyses that did manage to reach stationarity was also recorded.

#### *Positive control*

To validate our simulation framework a set of positive control simulations were constructed. In these simulations, fossil occurrences were generated with a heterogeneous Poisson process and not subsampled, resulting in no model misspecification when analyzed with the FBD skyline model. The processes of simulating trees and sequence data were exactly the same as employed in the focal simulation framework. Whereas the focal simulations used 100 replicate data sets, the positive control analyses were conducted on 10 replicate data sets. These simulations should yield coverage values for node ages that approach 1, coupled with small errors normally distributed around the generating values for these parameters.

## RESULTS

### *Empirical Data*

The use of different approaches to constructing and constraining subsamples of fossil occurrence data produce broadly different estimates of clade age. When a uniform subsample of occurrences is analyzed and there is no enforcement of the constraint of crown group membership for the oldest unequivocal fossil members of clades, the resulting age estimates are often younger than if this constraint is applied (Fig. 3). Similarly, the confidence intervals on these age estimates often extend to ages that are younger than the calibration fossil, even when these fossils are included in analyses but there is no constraint on their membership of the correct crown clade. For Xyelidae, the age estimates for the fully constrained analyses are pushed up against their hard minimum bound as defined by the age of the oldest member of this clade. Conversely, for the more topologically relaxed analysis and when this calibrating fossil was not included, these age estimates take on far younger and more normally distributed ages.

A similar pattern is seen in the resulting age estimates when only the oldest fossil occurrences that can be assigned to each crown clade are analyzed. There is widespread violation of the hard minimum bound on clade age when no topological constraints are applied, and older ages estimated when such constraints are used. When the results obtained from the uniform subsample and the arbitrary subsample are compared across all clades, there is no consistent pattern in which analyses produce older or younger age estimates, but there are often significant differences in both median age estimates and the extent of the 95% HPDs (Supplementary Fig. S2 available on Dryad at doi:10.5061/dryad.g7s0hk3).

## SIMULATED DATA

### *Positive control*

For all positive control analyses, the expected behavior of the FBD process is observed, with coverage above 0.9 and a clear linear relationship between estimated and generating node ages (Supplementary Fig. S1 available on Dryad). These results demonstrate that the simulation procedure generates FBD data without misspecification before the subsampling of fossil occurrences is implemented.

### *Unconstrained placement of uniformly subsampled fossil occurrences*

When the fossil sampling rate was an estimated parameter of the FBD model with a  $U(0, \text{Infinity})$  prior, node ages tend to be underestimated when the placement of fossil taxa is unconstrained. This is reflected in the median estimates of node age deviating away from the expected linear relationship with the simulating ages (Fig. 4; Row 1). Coverage of node age estimates is 0.39 and there is a clear directionality of the distribution of error. When the sampling rate is fixed to the MLE values, the average coverage improves to a value of 0.91 and the directionality of error in age estimates is reduced. Despite this, the precision of age estimates is decreased as the mean HPD width increases when the sampling rate is fixed. Broadly the same performance is observed when the MLE of sampling rate is set as the mean of the prior on this parameter, as when fixed fossil sampling rates are used, albeit with slightly reduced accuracy (mean absolute error = 0.26 Myr). For all analyses, the HPDs of estimated node ages extended to ages younger than their respective calibration fossils. When the rate of fossil sampling was fixed to a random value not estimated from the data, there is widespread underestimation of clade age and an average coverage value of only 0.07 (Fig. 5). For these analyses, there was a smaller number of replicates that reached satisfactory levels of convergence ( $n = 80$ ).



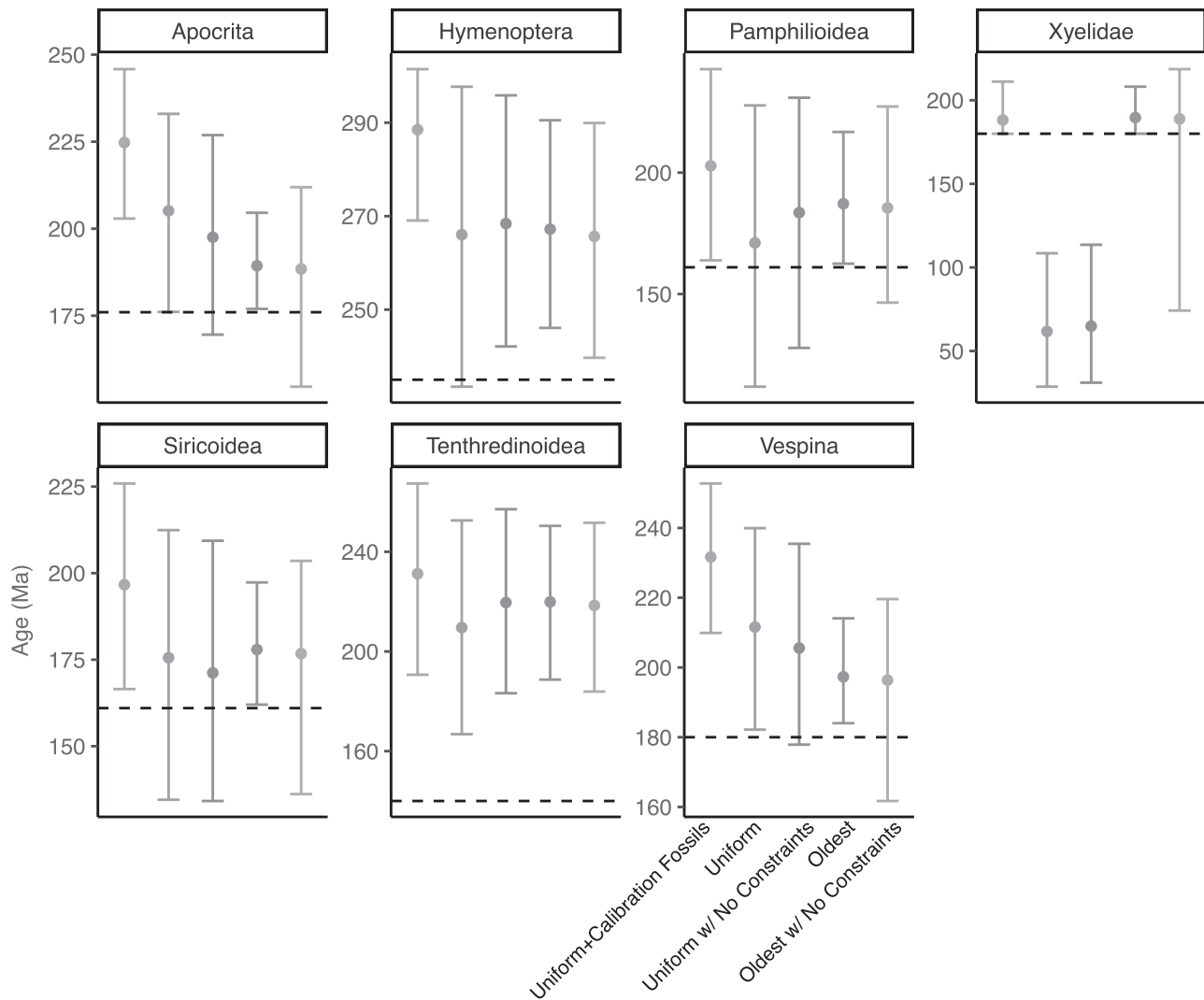


FIGURE 3. Median age estimates and 95% HPDs for key clades in Hymenoptera obtained using a range of approaches to constructing a subsample of fossil occurrences and constraining their placement. These approaches consist of: a uniform subsample of occurrences with and without topological constraints, a uniform subsample of occurrences supplemented with the oldest unequivocal members of clades which were then constrained to their respective crown groups, a subsample consisting of only the oldest unequivocal members of clades with and without topological constraints applied to constrain them to their respective crown groups. Dashed lines indicate the age of the oldest unequivocal member of each clade as determined by Ronquist et al. (2012). When no topological constraints are applied to the placement of fossil occurrences there is often positive probability assigned to ages that are in disagreement with the fossil record. Median age estimates and 95% HPDs for all other clades are presented in Supplementary Figure 2 available on Dryad.

#### *Uniformly subsampled fossil occurrences placed at the direct ancestral node*

If fossil occurrences are constrained to the correct clades without error then the estimates of node age are accurate (Fig. 4; Row 2). Both the mean width of the HPDs (1.09 Myr) and mean absolute error (0.24 Myr) are relatively small for the case in which a diffuse prior is applied to the rate of fossil sampling, and coverage is 0.94. When these data sets are analyzed with fixed rates of fossil sampling, coverage drops to 0.7 and ages are overestimated, coupled with an increase in mean absolute error (0.54 Myr) and a decrease in precision (mean HPD width = 1.42 Myr). When the MLE estimates of fossil sampling rate are applied as the prior on this

parameter, the same pattern in age estimates is seen as when fossil sampling rates are fixed, with slightly increased accuracy (absolute error = 0.42 Myr). For all analyses, the HPDs of estimated node ages extended to ages younger than their respective calibration fossils except for the case in which fixed fossil sampling rate is applied.

#### *Uniformly subsampled fossil occurrences supplemented with calibration fossils and placed at the direct ancestral node*

Constraining the placement of fossil occurrences and forcing the subsample to incorporate the oldest

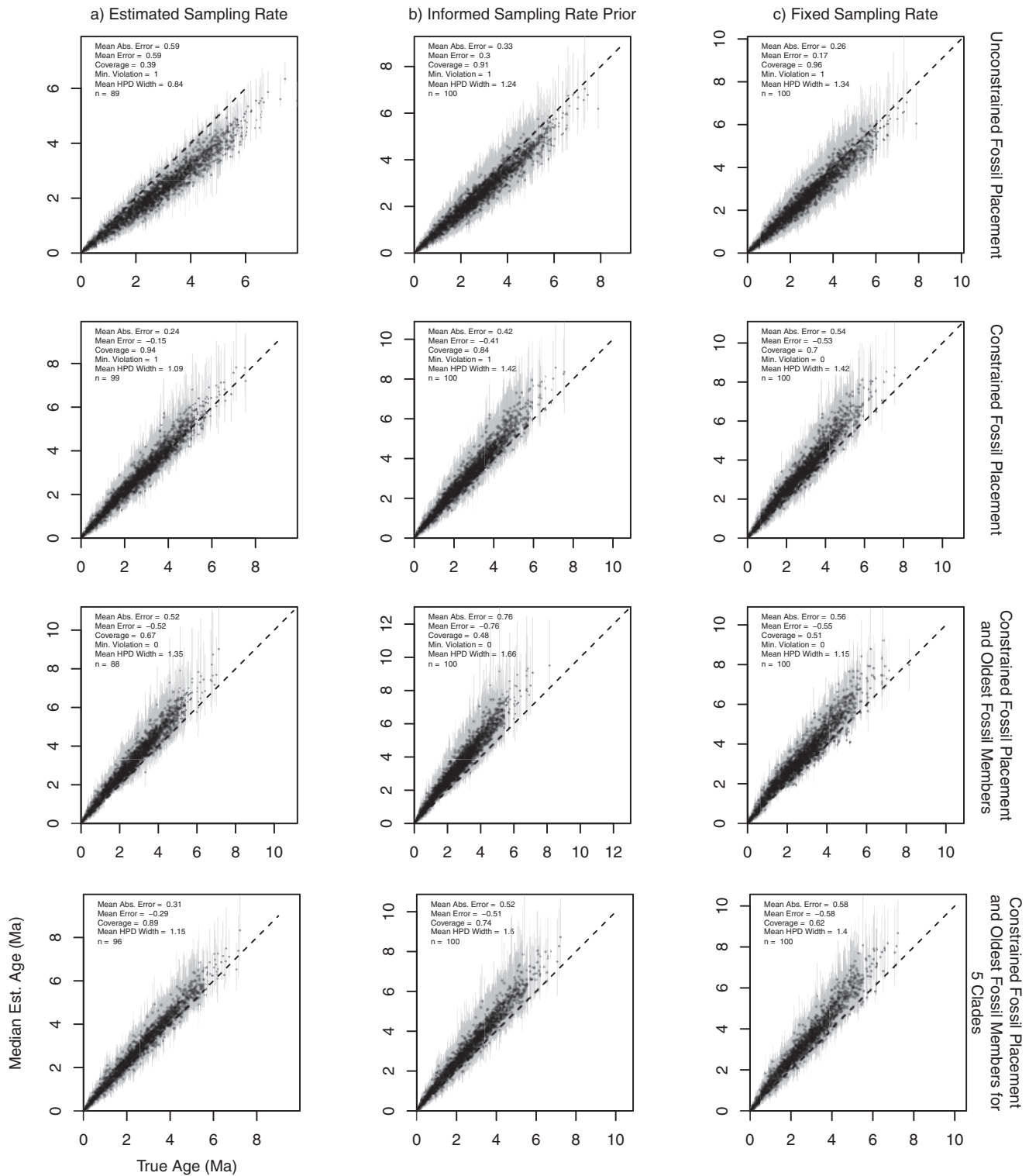


FIGURE 4. The accuracy and precision of estimated node ages obtained from subsamples of 100 replicate fossil occurrence data sets. Each point represents the median posterior age estimate of one clade, with grey bars representing the 95% HPD for that node age estimate. The three columns present the results when the rate of fossil sampling is accounted for in a different manner. In column a) the rate of fossil sampling is assigned a diffuse prior, in column b) the rate of fossil sampling is assigned a prior informed by an *a priori* estimate of fossil sampling rate, and in column c) the rate of fossil sampling is fixed to an *a priori* estimate of fossil sampling rate. Each row presents the results of the three approaches to accounting for fossil sampling rate when the topological constraints on the placement of fossils are varied, or the fossil subsample is supplemented with the oldest fossils that could be assigned to each clade or a subset of all clades. For the case in which only a subset of crown clades are provided with their oldest member the age estimates are only reported for the remaining clades for which the oldest fossil occurrences were not included. The minimum violation is measured as the proportion of nodes across all replicates for which an oldest fossil occurrence is available which possess an estimated HPD that extends to ages that are younger than the oldest fossil member of that same clade in the complete set of simulated fossil occurrences.

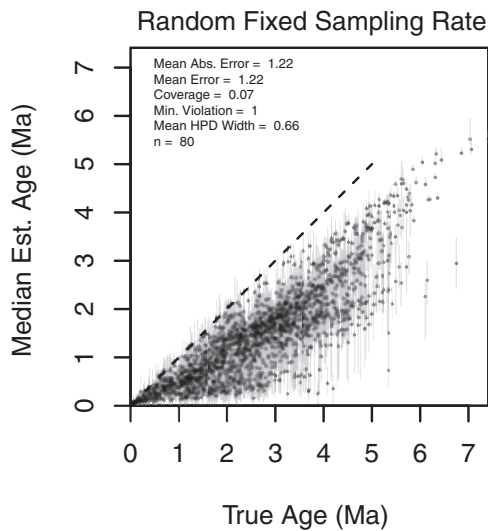


FIGURE 5. The accuracy and precision of node age estimates obtained from a uniform subsample of fossil occurrences when no topological constraints are applied to these occurrences and the rate of fossil sampling is fixed to a random value. There is a widespread underestimation of age when the rate of fossil sampling is not estimated from the data itself. This underestimation is likely to be caused by the random values of fossil sampling rate being far larger than the rate of fossil sampling with which the data were generated.

members of each clade should result in no violation of the true minimum age of each clade, as found in the complete simulated fossil data before subsampling. When this approach is taken the coverage is generally poor, irrespective of the approach to incorporating the rate of fossil sampling (estimated sampling rate = 0.67, fixed sampling rate = 0.51, MLE informed prior = 0.48) (Fig. 4; Row 3). Median node ages are systematically overestimated across all three approaches, but HPD ranges are not younger than the oldest fossil member of their respective node and, therefore, there is no posterior probability placed on impossible age estimates. Some node ages estimates are considerably younger than the generating ages (Fig. 4, third row: points below the dashed line); these are likely to be clades unrepresented by fossil occurrences in the full simulated data set and, therefore, had no available minimum constraint.

If a uniform subsample is supplemented with the oldest occurrences for only a subset of clades, there is improved accuracy in age estimates for the clades that do not have their oldest members included when the rate of fossil sampling is estimated from this reduced data set. Despite the improved accuracy there is still a tendency to overestimate clade ages. If the rate of fossil sampling is fixed or assigned an informative prior this results in a reduction in accuracy and increases the overestimation of clade age.

#### *Uniformly subsampled fossil occurrences supplemented with calibration fossils and placed with reduced precision*

In analyses where sampling rate was an estimated parameter, the accuracy of age estimates decreased as the constraints on the placement of fossil occurrences were relaxed (Supplementary Fig. S2 available on Dryad). This can be seen in the decrease in coverage as the precision of fossil placement decreases (1 Node = 0.96, 2 Nodes = 0.81, 4 Nodes = 0.5). Conversely, coverage improved when fixing the rate of fossil sampling (1 Node = 0.87, 2 Nodes = 0.93, 4 Nodes = 0.93) or placing an informed prior on this parameter (1 Node = 0.93, 2 Nodes = 0.96, 4 Nodes = 0.93). For all analyses, there was widespread extension of node age HPDs to ages younger than their respective calibration fossils, despite the constrained framework for fossil placement and inclusion of calibration fossils alongside the random sample of fossil occurrences.

#### *Arbitrary subsamples of fossil occurrences*

The accuracy of age estimates decreases when an arbitrary subsample of fossil occurrences is analyzed (Fig. 6). If only the youngest members of each clade are analyzed there is widespread underestimation of clade age (coverage is never greater than 0.81), particularly if there is no constraint on the placement of fossil occurrences (maximum coverage is 0.36). For all approaches to analysis in which the youngest fossils only are included, fixing the rate of fossil sampling or placing an informed prior on this parameter provides a modest improvement in the accuracy of age estimates. When only the oldest member of each clade is included in analyses and the rate of fossil sampling is estimated then there is widespread over- or underestimation of clade ages when the position of fossil occurrences are constrained or unconstrained, respectively. If the empirical Bayes approach is applied then this improves accuracy for the unconstrained fossil placement case but, as with a uniform subsample of fossil occurrences, provides no meaningful improvement to accuracy when topological constraints on fossil placement are enforced.

## DISCUSSION

### *Different Approaches to Incorporating Empirical Fossil Occurrence Data Result in Different Estimates of Clade Age*

Through the analysis of an empirical data set of fossil occurrence data and extant molecular data we have demonstrated that different approaches to accounting for the constraining of fossil occurrences can greatly influence posterior estimates of clade age, and different approaches to accounting for the rate of fossil sampling can influence the ability of MCMC to effectively sample from the posterior distribution. In particular, our results demonstrate the importance of including and constraining the placement of occurrences that would traditionally be used to construct the minima of node

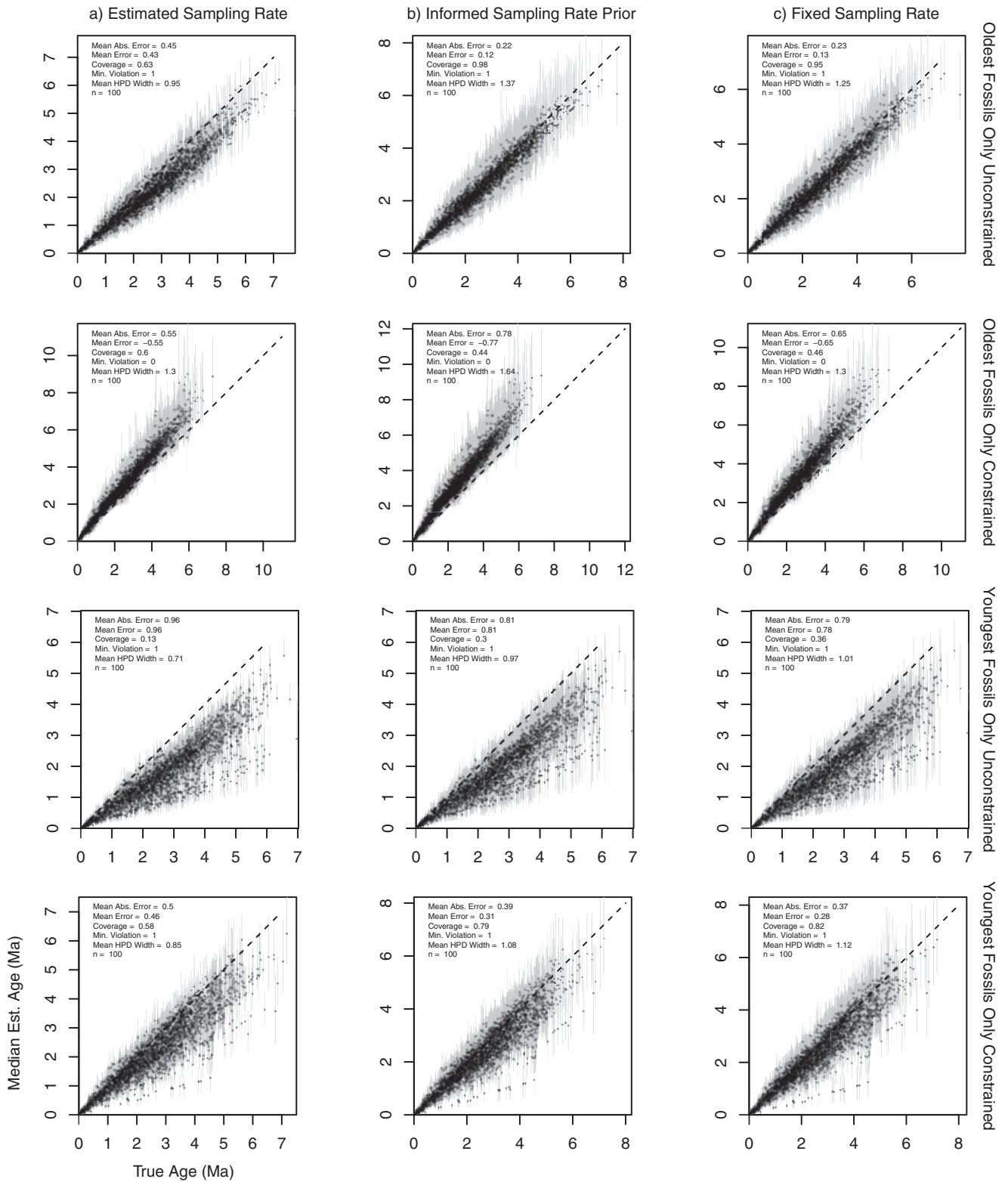


FIGURE 6. The accuracy and precision of node age estimates obtained from arbitrary subsamples of simulated fossil occurrence data.

calibrations. Without the inclusion of these fossils in a subsample, clade age estimates commonly extend to times that are incongruent with the fossil record. The lack of constraint on the minimum ages of clades leads to the estimation of much more recent clade ages than when the oldest fossil members of clades are included in analyses. When fossil subsamples are constructed with the oldest fossils topologically constrained, age estimates are considerably older but they are compatible with fossil evidence. When only a subset of all clades are provided with their oldest unequivocal occurrences constrained as crown group members there is still widespread overestimation of the age of other clades. This shows that the introduction of these fossil occurrences results in overestimation in general, not just in the clades assigned the oldest unequivocal fossil occurrences. If the oldest unequivocal fossils are included in analyses then great care must be taken when assigning fossil occurrences to nodes if they are to be constrained as crown group members. As first demonstrated by [Ronquist et al. \(2012\)](#), the fossil occurrence used to constrain the minimum age of crown Xyelidae is likely to actually be stem-Xyelidae. This is also seen in the results presented here, with the posterior distribution on this age pushed up against the hard minimum if this fossil occurrence is included in analysis and topologically constrained to crown-Xyelidae. If the occurrence is not included, or its membership to the crown is not enforced, age estimates are much younger and normally distributed.

Obtaining a sufficient sample of the posterior distribution for these types of analyses appears difficult when the majority of the parameters of the FBD process are estimated together. This was evident in our preliminary empirical analyses in which fossil sampling rate was treated as an estimated model parameter. In these analyses, the MCMC sampler was not capable of converging on the stationary distribution for the parameters of the FBD process. Despite this, the other model parameters, including node ages and clock rate, and the likelihood and prior were well sampled with independent runs converging on similar age estimates, suggesting that it may not be necessary for the FBD model parameters to be thoroughly sampled to obtain meaningful age estimates from the posterior distribution. As the FBD process defines the prior probability over node ages it seems that the effect of poor sampling of its constituent parameters on the accuracy of age estimates requires further investigation.

*Accuracy of Age Estimates Obtained from the FBD Process Depends on Topological Constraints, Fossil Subsample Structure and Treatment of the Rate of Fossil Sampling*

The analysis of empirical data allowed us to investigate the relative differences in age estimates obtained under a range of conditions but simulations are required to test the absolute accuracy and precision of age estimates. Through the use of simulated data we have demonstrated that when a subsample of fossil

occurrence data is analyzed, the accuracy of estimated node ages is dependent on both the structure of the subsample and the constraints applied to the topological placement of these occurrences. We have also demonstrated that an empirical Bayes approach to incorporating fossil occurrence data can lead to greatly improved precision and accuracy of age estimates when a fossil subsample is analyzed. Our results demonstrate that the greatest accuracy is often achieved in two cases: 1) when there are no constraints on the placement of fossil occurrences and an empirical Bayes approach to modeling the sampling rate is applied, and 2) when accurate constraints on fossil occurrence placement are applied and the rate of fossil sampling is estimated from the data *a priori*. We have also demonstrated that when there is uncertainty regarding the phylogenetic affinity of fossil occurrences and constraints on their placement that reflect this uncertainty are applied, employing an empirical Bayes approach improves accuracy. Given that there is often considerable uncertainty associated with the phylogenetic affinity of fossil taxa because of fragmentary and incomplete preservation or conflicting interpretation of characters, it seems that employing an empirical Bayes approach may generally be preferable.

It is perhaps unexpected that a random subsample of fossil occurrences will not result in accurate age estimates when the rate of fossil sampling is an estimated model parameter. This may be caused by the subsampling of data, as the positive control analyses that did not employ subsampling managed to recover accurate age estimates. Reduced accuracy when subsampling may be caused by disparity in the number of lineages in the simulated trees between the root and the extant tips. Tree simulation under the time homogeneous birth–death process results in relatively few lineages toward the root of the tree as this model of growth naturally results in more lineages at the end of the process than the start. A consequence of this was that the oldest time slice often contained few fossil occurrences as there were few lineages upon which occurrences could be generated. Conversely, toward the extant tips of the tree there were large numbers of lineages upon which occurrences could be generated. Consequently, a uniform subsample of occurrences generated on this tree often included few fossils from the older lineages and this may mislead analyses if the rate of fossil sampling within these older lineages is estimated with this potentially insufficient subsample. It may also be the case that some constraint on the placement of fossil occurrences is necessary, as analyses in which topological constraints were applied to fossil occurrences resulted in greater accuracy than those that did not (Fig. 4). Despite this, including the oldest unequivocal fossil occurrences for clades results in overly ancient estimates when topological constraints are employed. When constructing a subsample, the structure of the fossil occurrences through time should be preserved. An estimate of sampling rate made from this subsample is expected to be equal to the sampling rate in the complete set of occurrences multiplied by

the proportion of occurrences in the subsample. Our results suggest that the FBD model often requires a large unbiased sample of fossil occurrences to produce accurate estimates of node age. The empirical Bayes approach avoids this problem by separating the issue of estimating the fossil sampling rate from the estimation of other parameters, resulting in two simple analyses instead of one potentially impractical analysis. Here, we show that the effect of accounting for fossil sampling rate in this manner improves age estimate accuracy, even when an arbitrary subsample of fossil occurrences is analyzed (Figs. 4 and 6). It is important to note that when the empirical Bayes approach is taken, the alternative parametrization of the FBD process (Heath et al. 2014; Gavryushkina et al. 2017) cannot be used as the fossil sampling rate is inextricably linked to the death rate in the sampling proportion parameter.

One of the most surprising results from our subsampled FBD simulation analyses is the positive probability placed on node ages that are younger than their oldest unsampled descendants. To avoid these artifacts, the placement of fossil occurrences in the tree must be both highly constrained and accurate. We have demonstrated that if fossil occurrences are resolved conservatively to more inclusive clades (1–4 nodes lower than their correct, more exclusive, clade) then the underestimation of clade age remains widespread. This occurs because fossil occurrences conservatively allied to more inclusive clades effectively remove the constraint that a more exclusive clade is older than the oldest fossil occurrence. Based on all of these results, determining which application of fossil data should be preferred involves a choice between (1) accepting accurate median estimates of clade age coupled with HPDs extending to impossible ages, or (2) inaccurate median estimates with HPDs that do not extend to impossible ages.

#### *Fixing the Rate of Fossil Sampling with Subsampled Occurrence Data*

The results presented here demonstrate the viability and potential benefits of applying an empirical Bayes approach to accounting for quantities of fossil occurrence data that are too large to be applied together in a single FBD analysis. Such an approach is naturally dependent on the ability to obtain accurate estimates of fossil sampling rate using the full complement of available data, and this is demonstrated in the reduced accuracy of age estimates when fossil sampling is fixed to a random value (Fig. 5). The model we have used to obtain estimates of sampling rate was chosen because it provides continuous estimates of sampling rate. However, this model is dependent on both the availability of high precision data for the ages of specific fossil occurrences (Solow and Smith 1997; Bapst 2012) and the correct assignment of the fossil occurrences to specific taxa. This is often not the case for empirical data and presents a challenge for the application of this approach. Nevertheless, the age estimates obtained with fixed fossil sampling rates in the simulation analyses

are generally as accurate, and often more accurate, than those obtained with sampling rate estimated alongside the other model parameters. Similarly, divergence time estimates for Hymenoptera are not very different from previous estimates in which fossil data are analyzed alongside molecular data (Ronquist et al. 2012; Zhang et al. 2016). Even though these results suggest that this approach works in principle, the accuracy of this framework will likely benefit from further investigation into the application of alternative models for the *a priori* estimation of fossil sampling rate (Foote and Raup 1996; Silvestro et al. 2014). Here, we have considered the case where fossil sampling rates are regularly overestimated and fixed to these values, but further sensitivity tests will be required to thoroughly characterize the effect of inaccurate fixed fossil sampling rates. Another benefit of this framework is that fixing some parameter values to their MLEs results in better MCMC mixing for the other parameters, and in some cases enables convergence to the posterior distribution when a full coestimation of all parameters appears incapable of doing so.

#### *Implications for the Resolved FBD Process*

Our analyses have focused on the unresolved FBD process, however, our results also have implications for the resolved FBD process. As the distribution of fossils through time is integral to the FBD process it seems that the potentially nonrandom distribution of fossil occurrences induced by a dependence on the preservation of morphological data may influence the accuracy of age estimates obtained with the resolved framework. Our results show that if a fossil occurrence is not assigned to the correct node then it is possible to estimate ages which extend to impossibly young values. This highlights the importance of the accurate placement of fossil taxa in both unresolved and resolved FBD analyses. Furthermore, the oldest unequivocal member of a clade may be poorly preserved and possess only a handful of characters available for clade diagnosis, with no further morphological data suitable for quantitative phylogenetic analysis. If such fossils cannot be included among their relatives represented by occurrences preserving meaningful quantities of morphological data suitable for phylogenetic analysis, the minimum age constraint on their respective clade will be incorrectly constrained. Therefore, it seems that to accurately constrain the minimum ages of clades, analyses will be reliant on both the quality and availability of fossil morphological data or an approach that combines aspects of both the resolved and unresolved FBD frameworks. Nevertheless, our analyses did not explicitly address the situation in which fossil morphological data are included in the analysis and further simulation analyses are required to assess the performance of this class of FBD analyses.

#### CONCLUSION

The FBD process provides a framework in which more fossil biostratigraphic data can be included than

with node calibration methods. The benefits of this framework have been well-characterized, and the ability to model fossil sampling holds great promise in the goal of obtaining accurate and precise evolutionary timescales. When applying the FBD process to a group with an abundance of available fossil occurrence data a choice must be made regarding which occurrences to include. Here, we have demonstrated that relying on the FBD estimate of fossil sampling rate from a random sample of occurrence data may lead to inaccurate age estimates if the topological affinities of fossils are not correctly accounted for. To avoid this issue, and to obtain accurate age estimates, an empirical Bayes approach to incorporating occurrence data can be applied, which allows all available fossil occurrence data to influence divergence time estimates. We have also demonstrated that when analyzing a subsample of occurrence data the FBD process has a tendency to estimate node ages with probability density extending to impossibly young ages and controlling for this effect results in inaccurate median posterior estimates of node age.

#### SUPPLEMENTARY MATERIAL

Data available from the Dryad Digital Repository: <http://dx.doi.org/10.5061/dryad.g7s0hk3>.

#### AUTHORS' CONTRIBUTIONS

J.E.O.'R. conceived the simulation framework and carried out analyses, both J.E.O.'R. and P.C.J.D. contributed to the writing of the manuscript.

#### FUNDING

J.E.O.'R. and P.C.J.D. are funded NERC (NE/P013678/1), P.C.J.D. is also funded by NERC (NE/N002067/1) and BBSRC (BB/N000919/1).

#### ACKNOWLEDGMENTS

With thanks to Rachel Warnock, Joëlle Barido-Sottani, and Will Pett for discussion.

#### REFERENCES

- Bapst D.W. 2012. paleotree: an R package for paleontological and phylogenetic analyses of evolution. *Methods Ecol. Evol.* 3:803–807.
- 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., O'Reilly J.E., Pett W., Warnock R.C.M. 2018. FossilSim: An R package for simulating fossil occurrence data under mechanistic models of preservation and recovery. *Methods Ecol. Evol.* 2019; 00:1–6. <https://doi.org/10.1111/2041-210X.13170>.
- 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.
- Casella G. 1985. An introduction to empirical Bayes data-analysis. *Am. Stat.* 39:83–87.
- 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.
- Eguchi S., Tamura M.N. 2016. Evolutionary timescale of monocots determined by the fossilized birth-death model using a large number of fossil records. *Evolution.* 70:1136–1144.
- Footo M., Raup D.M. 1996. Fossil preservation and the stratigraphic ranges of taxa. *Paleobiology.* 22:121–140.
- 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(12):p.e1003919.
- Gernhard T. 2008. The conditioned reconstructed process. *J. Theor. Biol.* 253:769–778.
- Grimm G.W., Kapli P., Bomfleur B., McLoughlin S., Renner S.S. 2015. Using more than the oldest fossils: dating osmundaceae with three Bayesian clock approaches. *Syst. Biol.* 64:396–405.
- 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.
- Ho S.Y., Phillips M.J. 2009. Accounting for calibration uncertainty in phylogenetic estimation of evolutionary divergence times. *Syst. Biol.* 58:367–380.
- Hohna S., Stadler T., Ronquist F., Britton T. 2011. Inferring speciation and extinction rates under different sampling schemes. *Mol. Biol. Evol.* 28:2577–2589.
- Holland S.M. 1995. The stratigraphic distribution of fossils. *Paleobiology.* 21:92–109.
- Holland S.M. 2016. The non-uniformity of fossil preservation. *Philos. Trans. R. Soc. B Biol. Sci.* 371(1699):p.20150130.
- Kendall D.G. 1948. On the generalized birth-and-death process. *Ann. Math. Stat.* 19:1–15.
- Kolaczkowski B., Thornton J.W. 2007. Effects of branch length uncertainty on Bayesian posterior probabilities for phylogenetic hypotheses. *Mol. Biol. Evol.* 24:2108–2118.
- Lewis P.O. 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. *Syst. Biol.* 50:913–925.
- Matschiner M., Musilova Z., Barth J.M.I., Starostova 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., Wright A. 2016. Inferring node dates from tip dates in fossil Canidae: the importance of tree priors. *Biol. Lett.* 12:4.
- O'Reilly J.E., Donoghue P.C.J. 2016. Tips and nodes are complementary not competing approaches to the calibration of molecular clocks. *Biol. Lett.* 12(4):p.20150975.
- O'Reilly J.E., dos Reis M., Donoghue P.C.J. 2015. Dating tips for divergence-time estimation. *Trends Genet.* 31:637–650.
- Parham J.F., Donoghue P.C., Bell C.J., Calway T.D., Head J.J., Holroyd P.A., Inoue J.G., Irmis R.B., Joyce W.G., Ksepka D.T., Patané J.S.L., Smith N.D., Tarver J.E., van Tuinen M., Yang Z., Angielczyk K.D., Greenwood J.M., Hipsley C.A., Jacobs L., Makovicky P.J., Müller J., Smith K.T., Theodor J.M., Warnock R.C.M., Benton M.J. 2012. Best practices for justifying fossil calibrations. *Syst. Biol.* 61: 346–359.
- Peters S.E., McClennen M. 2016. The Paleobiology Database application programming interface. *Paleobiology.* 42:1–7.
- Plummer M., Best N., Cowles K., Vines K. 2006. CODA: convergence diagnosis and output analysis for MCMC. *R News.* 6:7–11.
- Pyron R.A. 2011. Divergence time estimation using fossils as terminal taxa and the origins of Lissamphibia. *Syst. Biol.* 60:466–481.
- Pyron R.A. 2017. Novel approaches for phylogenetic inference from morphological data and total-evidence dating in squamate reptiles (Lizards, Snakes, and Amphisbaenians). *Syst. Biol.* 66:38–56.
- 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.

- Ronquist F., Klopstein S., Vilhelmsen L., Schulmeister S., Murray D.L., Rasnitsyn A.P. 2012. A total-evidence approach to dating with fossils, applied to the early radiation of the hymenoptera. *Syst. Biol.* 61:973–999.
- Saladin B., Leslie A.B., Wuest R.O., Litsios G., Conti E., Salamin N., Zimmermann N.E. 2017. Fossils matter: improved estimates of divergence times in *Pinus* reveal older diversification. *BMC Evol. Biol.* 17(1):95.
- Sansom R.S., Wills M.A. 2013. Fossilization causes organisms to appear erroneously primitive by distorting evolutionary trees. *Sci. Rep.* 3:5.
- Silvestro D., Salamin N., Schnitzler J. 2014. PyRate: a new program to estimate speciation and extinction rates from incomplete fossil data. *Methods Ecol. Evol.* 5:1126–1131.
- Slater G.J., Goldbogen J.A., Pyenson N.D. 2017. Independent evolution of baleen whale gigantism linked to Plio-Pleistocene ocean dynamics. *Proc. R. Soc. B Biol. Sci.* 284(1855): p.20170546.
- Solow A.R., Smith W. 1997. On fossil preservation and the stratigraphic ranges of taxa. *Paleobiology.* 23:271–277.
- 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(5):676–684, <https://doi.org/10.1093/sysbio/syr029>.
- Stadler T., Kuhnert D., Bonhoeffer S., Drummond A.J. 2013. Birth-death skyline plot reveals temporal changes of epidemic spread in HIV and hepatitis C virus (HCV). *Proc. Natl. Acad. Sci. USA.* 110: 228–233.
- Warnock R.C., Yang Z., Donoghue P.C. 2012. Exploring uncertainty in the calibration of the molecular clock. *Biol. Lett.* 8:156–159.
- Warnock R.C.M., Parham J.F., Joyce W.G., Lyson T.R., Donoghue P.C.J. 2015. Calibration uncertainty in molecular dating analyses: there is no substitute for the prior evaluation of time priors. *Proc. R. Soc. B Biol. Sci.* 282(1798): p.20141013.
- Wright D.F. 2017. Bayesian estimation of fossil phylogenies and the evolution of early to middle Paleozoic crinoids (Echinodermata). *J. Paleontol.* 91:799–814.
- Yang Z.H., Rannala B. 1997. Bayesian phylogenetic inference using DNA sequences: a Markov Chain Monte Carlo method. *Mol. Biol. Evol.* 14:717–724.
- Zhang C., Stadler T., Klopstein S., Heath T.A., Ronquist F. 2016. Total-evidence dating under the fossilized birth-death process. *Syst. Biol.* 65:228–249.