University of Zurich

Year: 2019

Ignoring stratigraphic age uncertainty leads to erroneous estimates of species divergence times under the fossilized birth–death process

Barido-Sottani, Joëlle; Aguirre-Fernández, Gabriel; Hopkins, Melanie J; Stadler, Tanja; Warnock, Rachel

Abstract: Fossil information is essential for estimating species divergence times, and can be integrated into Bayesian phylogenetic inference using the fossilized birth–death (FBD) process. An important aspect of palaeontological data is the uncertainty surrounding specimen ages, which can be handled in different ways during inference. The most common approach is to fix fossil ages to a point estimate within the known age interval. Alternatively, age uncertainty can be incorporated by using priors, and fossil ages are then directly sampled as part of the inference. This study presents a comparison of alternative approaches for handling fossil age uncertainty in analysis using the FBD process. Based on simulations, we find that fixing fossil ages to the midpoint or a random point drawn from within the stratigraphic age range leads to biases in divergence time estimates, while sampling fossil ages leads to estimates that are similar to inferences that employ the correct ages of fossils. Second, we show a comparison using an empirical dataset of extant and fossil cetaceans, which confirms that different methods of handling fossil age uncertainty lead to large differences in estimated node ages. Stratigraphic age uncertainty should thus not be ignored in divergence time estimation and instead should be incorporated explicitly.

Authors for correspondence:
Joëlle Barido-Sottani
e-mail: joelle.barido-sottani@m4x.org
Rachel Warnock
e-mail: rachel.warnock@bsse.ethz.ch

THE ROYAL SOCIETY
PUBLISHING

Ignoring stratigraphic age uncertainty leads to erroneous estimates of species divergence times under the fossilized birth–death process

Joëlle Barido-Sottani[1,2,3], Gabriel Aguirre-Fernández[4], Melanie J. Hopkins[5], Tanja Stadler[1,2] and Rachel Warnock[1,2,4]

[1]Department of Biosystems Science and Engineering, ETH Zurich, Basel, Switzerland
[2]Swiss Institute of Bioinformatics (SIB), Lausanne, Switzerland
[3]Department of Ecology, Evolution and Organismal Biology, Iowa State University, Ames, IA, USA
[4]Palaeontological Institute and Museum, University of Zurich, Zurich, Switzerland
[5]Division of Paleontology, American Museum of Natural History, New York, NY, USA

JB-S, 0000-0002-5220-5468; GA-F, 0000-0002-9964-1646; TS, 0000-0001-6431-535X

Fossil information is essential for estimating species divergence times, and can be integrated into Bayesian phylogenetic inference using the fossilized birth–death (FBD) process. An important aspect of palaeontological data is the uncertainty surrounding specimen ages, which can be handled in different ways during inference. The most common approach is to fix fossil ages to a point estimate within the known age interval. Alternatively, age uncertainty can be incorporated by using priors, and fossil ages are then directly sampled as part of the inference. This study presents a comparison of alternative approaches for handling fossil age uncertainty in analysis using the FBD process. Based on simulations, we find that fixing fossil ages to the midpoint or a random point drawn from within the stratigraphic age range leads to biases in divergence time estimates, while sampling fossil ages leads to estimates that are similar to inferences that employ the correct ages of fossils. Second, we show a comparison using an empirical dataset of extant and fossil cetaceans, which confirms that different methods of handling fossil age uncertainty lead to large differences in estimated node ages. Stratigraphic age uncertainty should thus not be ignored in divergence time estimation and instead should be incorporated explicitly.

1. Introduction

The fossil record provides essential evidence for calibrating species trees to time, as molecular sequences from extant species are informative about the relative age of species but do not typically provide information about the absolute age. A common approach to calibration, referred to as node dating, is to assign a single fossil to a specific node in a phylogeny and to reflect the uncertainty in its age using a probability distribution, where the minimum bound of the distribution corresponds to the age of the specimen [1–3]. It has been shown that divergence time estimates are extremely sensitive to the choice of fossil(s), the age assigned to fossil specimens and the distribution chosen to model uncertainty [4–8]. However, regardless of specimen choice, node dating has additional drawbacks. For instance, this approach effectively uses one fossil per node and makes it extremely challenging to derive or implement explicit priors on divergence times [9,10]. The fossilized birth–death (FBD) process offers an alternative approach to calibration, which integrates fossil samples into the tree under the same diversification process that describes extant species [11,12]. This approach greatly increases the amount of fossil evidence that can

be used during inference, but the impact of the taxonomic, stratigraphic range and stratigraphic age uncertainty has not been fully explored using this framework.

Here, we explore the impact of stratigraphic age uncertainty. Fossils are rarely composed of material that can be directly dated and their age must be established with detailed reference to the geological record. This procedure leads to some uncertainty. First, the rock layer (or lithostratigraphic unit) from which a specimen was collected must be established. If layers directly above and below that unit have not been directly dated, the relative age (or biostratigraphic unit) of a specimen must be established using index fossils. Finally, the absolute minimum and maximum age of a specimen must be obtained with reference to a numeric time scale (or chronostratigraphic chart). The process of dating fossils is not always straightforward, because the link between litho-, bio- and chronostratigraphy can be challenging to establish, or the stratigraphic provenance of a specimen may be ambiguous, and there are limits to the level of resolution that can be obtained [13–18].

Current applications of the FBD model typically assign specimen ages using the midpoint of the known interval of age uncertainty (e.g. [19]) or a random age drawn from that interval (e.g. [20,21]). However, fossil age uncertainty can also be modelled explicitly by placing a prior on the fossil ages and co-estimating these along with other model parameters [22]. A previous study suggested that divergence estimates are not affected by the uncertainty surrounding specimen ages [23], but this work focused on ancient DNA samples, which are relatively young (less than 60 000 years old) and associated with much lower age uncertainty than most fossil occurrences. Other authors have shown that modelling age uncertainty versus fixing fossil occurrence ages can lead to substantial differences in node age estimates in tip-calibrated analyses [24]. However, this study used a uniform tree prior, rather than the FBD prior, and did not explicitly quantify model performance, as this work was based on empirical data. More recently, the performance of the FBD model has been examined using extensive simulations [25]. This work found that sampling fossil ages had a slight impact on estimates in total-evidence analyses, compared with fixing ages to the truth, but did not study the effects of fixing fossil ages to a random or midpoint value, which is commonly done in practice. In this paper, we explore fossil age uncertainty as a potential source of error in FBD analyses using simulated and empirical data, and we describe how various methods of handling age uncertainty can affect the results. Our simulations show that fixing specimen ages can lead to erroneous estimates of divergence times, but that incorporating stratigraphic age uncertainty explicitly using a hierarchical modelling approach substantially increases the chances of recovering the correct node ages. An analysis of Cetacea (the clade containing dolphins and whales) reveals that alternative approaches to handling specimen ages have major implications for dating speciation events.

2. Methods

(a) Simulated datasets
Our simulation study is based on the class Mammalia, which is very well represented in both molecular and palaeontological databases. Mammals and their subgroups have also been the subject of a large number of molecular dating studies [26–28], as well as studies that rely on time-scaled phylogenies [29,30]. Simulated age uncertainty was based on the fossil record of North American mammals, which is relatively complete, has been studied in detail and is stratigraphically well constrained. Thus, the degree of age uncertainty incorporated into our simulations represents a best-case scenario for inferring the divergence times of mammal-like groups.

(i) Simulation of extant species phylogenies and fossil samples
Trees were simulated under a birth–death process using the R package TreeSim [31]. The speciation rate was set to $\lambda = 0.15$ per Myr and the extinction rate to $\mu = 0.1$ per Myr, in accordance with estimates from [29] for the mammalian phylogeny. The birth–death process was stopped once $n_{\text{extant}} = 25$ extant species had been reached. At this stage, we rejected trees whose origin was not between 40 Ma and 100 Ma. This interval broadly encompasses the estimated origin time for many major groups of mammals [27,30], but avoids simultaneously conditioning on tree age and tip number, which can be problematic [32].

Fossils were then sampled on the complete phylogeny using the R package FossilSim, following a Poisson process with a constant rate. The fossilization rate was set to $\psi = 0.2$ per Myr, based on estimates of fossil recovery rates among Cenozoic mammals [33]. Note that under this process, any number of fossils can be sampled for a given species. We did not filter the fossils further, as the current implementation of the FBD model in BEAST2 is designed for specimen-level data. We rejected phylogenies with less than 4 or more than 125 sampled fossils. A minimum of 4 was chosen to ensure convergence when analysing the simulated data, while a maximum of 125 was chosen to avoid generating trees with a large number of extinct samples relative to the number of extant samples ($n = 25$), which is not typical in divergence time studies. The number of extant samples was chosen to make a large number of simulations possible (under these settings our largest trees have up to 300 branches), however, the results should be similar those obtained for trees with $n > 25$. The sampled tree was then obtained by assuming complete extant species sampling ($\rho = 1$), which is present in many empirical mammal datasets (see [34]; electronic supplementary material, table S14), and discarding all unsampled lineages.

(ii) Simulation of sequences
Sequences were simulated for the extant species using seq-gen [35] via the R package phyclust [36]. We simulated sequences of length 2000 nucleotides under an HKY + Γ model with 5 rate categories, following the substitution model used in [37]. We used a lognormal uncorrelated clock, where the substitution rate for each branch of a given tree was sampled from a lognormal distribution with standard deviation 0.1. The mean of the lognormal distribution was drawn separately for each tree from a gamma distribution. A full list of all parameters used in the sequence simulation can be found in table 1.

(iii) Simulation of fossil age uncertainty
The minimum and maximum age of each fossil occurrence was simulated based on the sampling interval ages of North American mammals. These intervals were downloaded from the Paleobiology Database (PBDB) on 12 December 2017, using the following parameters: time intervals = Mesozoic and Cenozoic, region = North America, scientific name = Mammalia. If a fossil age could be assigned to multiple intervals, a single interval was selected at random by weighting all possible intervals by their frequency of appearance in the PBDB data. If no intervals appeared in the PBDB data for a simulated fossil, a random interval of fixed length was drawn. This length was fixed to the average length of all intervals present in the PBDB
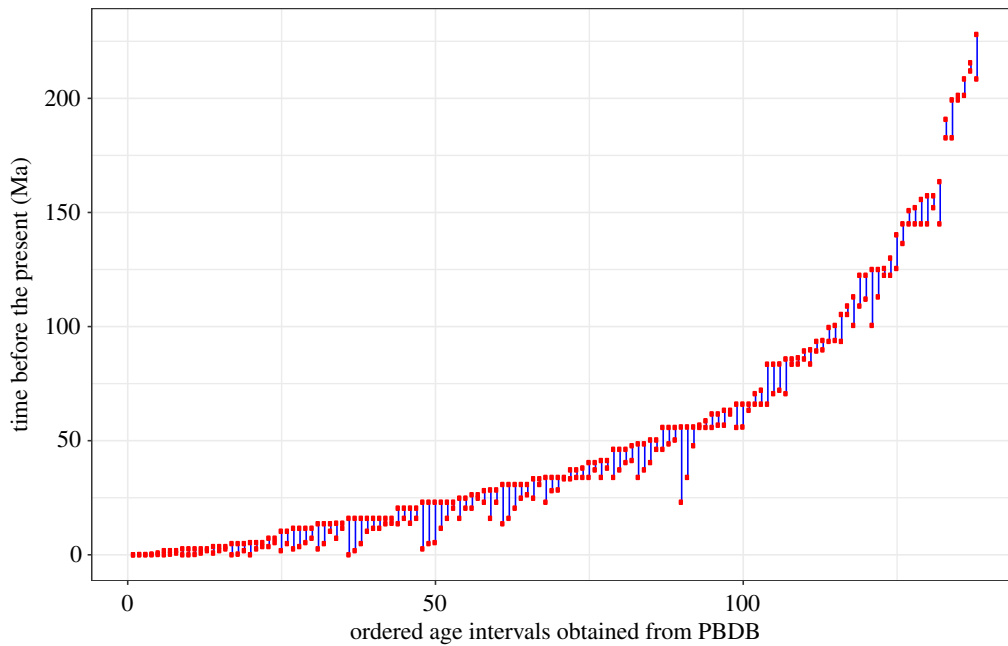
Figure 1. Representation of the age intervals obtained from PBDB for North American mammals sampled during the Mesozoic and Cenozoic. Intervals are ordered by the maximum age of the range, from youngest to oldest.

Table 1. List of all parameter values used to simulate sequences.

parameter	value
mean substitution rate	$\sim$ Gamma($\alpha = 2$, $\beta = 2$)/100
branch-specific substitution rate	$\sim$ LogNormal(mean rate, 0.1)
number of rate categories	5
shape of the gamma distribution on rates (α)	0.25
transition/transversion ratio (κ)	5
base frequencies	0.25, 0.25, 0.25, 0.25

data (i.e. 8 Myr). Thus, the simulated interval for each fossil always included the correct age of the fossil. This procedure was designed so that the pattern of uncertainty ranges in our simulated data would be similar to the patterns found in real data, where in particular younger fossils tend to be associated with less uncertainty than older fossils. A visual representation of all the intervals used for simulation is shown in figure 1.

(iv) Bayesian inference

We used the Sampled Ancestors package [38], which provides an implementation of the fossilized birth–death process for the Markov chain Monte Carlo (MCMC) software BEAST2 [39], to perform Bayesian phylogenetic inferences on our simulated datasets.

The fossil ages were handled using five different methods, detailed here:

Correct ages: the fossil ages are fixed to the true ages as simulated.
Interval ages: the fossil ages are not fixed, but are sampled along with the other parameters within the simulated age range.

Median ages: the fossil ages are fixed to the midpoint of their simulated age range.
Random ages: the fossil ages are fixed to an age sampled uniformly at random inside of their simulated age range.
Symmetric interval ages: the fossil ages are not fixed, but are sampled along with the other parameters. Each fossil age is sampled within a symmetric interval around the true age of the fossil. The purpose of this setting was to evaluate the performance and accuracy in the situation where the midpoint of each prior interval was equal to the correct age.

A schematic of these different methods can be seen in figure 2. Note that for the interval age methods, we sample trees as in [22]. In particular, we set the probability density of the proposed tree to the FBD probability density if all fossil ages are within their intervals, and 0 otherwise. The effective prior on fossil ages (i.e. the fossil age distribution when using all information excluding sequence data) is thus not a uniform prior, as the FBD model already induces a distribution on fossil ages. Indeed, sampling the effective priors using BEAST2 shows a wide range of priors on fossil ages, from distributions which are very close to uniform over the entire age interval, to distributions where most of the density weight is concentrated on a fraction of the age interval. There is no clear correlation between the shape of the effective prior and the age of the fossil.

In all inferences, the tree topology was inferred but constrained, as fossils were assigned to clades according to the correct tree topology. No sequence data were included for the fossil samples. The substitution model and clock model were set to the models used during simulation, and the priors used can be found in the electronic supplementary material. The value of ρ was fixed to the truth during the analyses. The inference was run for 100 000 000 iterations, sampled every 10 000. Convergence was assessed in the software Tracer v. 1.7 [40] and considered satisfactory if the effective sample sizes were more than 200.

(b) Empirical dataset

To explore the impact of stratigraphic age uncertainty on empirical estimates of divergence times, we compiled a dataset of Cetacea containing both fossil occurrences and an alignment of sequences for extant species. This group was chosen based on
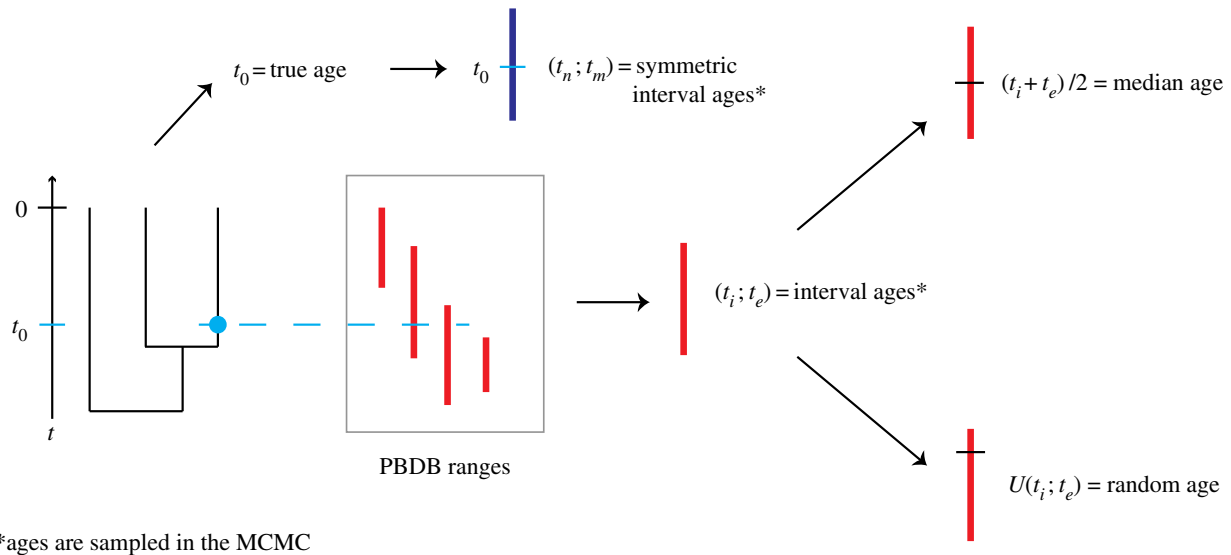
Figure 2. Representation of the age uncertainty simulation process. Phylogenies with fossils are simulated according to a birth–death–fossilization process. The correct age of each fossil is used to draw an age interval for that fossil from the set obtained from PBDB. This age interval is then used as the basis for the median and random age assignment. A symmetric age interval is also drawn from the correct age.

the availability of a large molecular alignment representing almost all extant species, in combination with well-curated and comprehensive stratigraphic occurrence data. This group has also been the focus of a large number of molecular dating studies [41–45].

(i) Fossil occurrence data

We obtained data on 4473 fossil occurrences from the PBDB on 5 April 2018, using the parameter 'scientific name = Cetacea'. The full dataset could not be used due to mixing issues, so we sub-sampled 10% and 5% of the fossils at random, obtaining, respectively, 448 and 224 fossil occurrences. The classification of taxa into suprageneric ranks was largely based on [46]. A list of genera and their taxonomic membership as used in the subsample is provided in the electronic supplementary material. We used the minimum and maximum age for each fossil occurrence as recorded in the PBDB.

(ii) Alignment

We used the alignment provided by Steeman et al. [41], which contains sequences for six mitochondrial and nine nuclear genes for 87 of 89 extant cetacean species. We excluded from our analysis the three outgroup taxa which were present in the original alignment, as our dataset contains no fossils for these families. The full alignment thus contains 87 sequences with 16 175 characters each.

Following [41], we divided this alignment into 28 partitions. The best substitution models for each partition were selected using bModelTest [47]. A complete list of the substitution models used can be found in the electronic supplementary material.

(iii) Bayesian inference

Bayesian phylogenetic inference was performed using the same model parameters and priors as for the simulated data, with the exception of the substitution model, which was set as specified in the previous section. As the correct ages of the fossils in this dataset are unknown, we limited our comparison to median ages, random ages and interval ages.

Similar to the simulation study, the topology was inferred but constrained. Topological constraints were set at both the genus and the family level, following the classification from PBDB, so that each genus or family formed a monophyletic clade in the tree. Samples whose position could not be determined were not included in any clade constraint, and thus could appear anywhere outside of the determined clades.

Following the model described in [48] for the analysis of empirical data, we unlinked the substitution models and among-site rate variation across partitions but linked the clock model and applied partition-specific rate multipliers to account for variation in evolutionary rates.

The inference was run for 300 000 000 iterations sampled every 10 000. Convergence was assessed in the software Tracer v. 1.7 [40] and by running two chains with different starting values.

3. Results

(a) Simulated datasets

(i) Accuracy

We measured the relative error of the median estimates for the divergence times and FBD model parameters obtained using different approaches to handling stratigraphic age uncertainty. We also calculated the coverage, which is the number of analyses (out of 100 trees) in which the true parameter value was included in the 95% highest posterior density (HPD) interval. The error and coverage of the divergence times for each tree were averaged across all nodes of the extant tree (i.e. all nodes that were the most recent common ancestors of extant tips). Results for the divergence times, diversification rate and turnover estimates are shown in figure 3.

Using either interval ages or symmetric interval ages leads to error and coverage values that are very close to the results obtained using the correct age of the fossils. However, median and random ages did much worse than other methods. This is particularly apparent for estimates of divergence times, where error values range from 0.15 to 0.20 for correct, interval and symmetric interval ages, versus 0.57 and 0.74 for median and random ages, respectively. Similarly, the coverage of divergence times is 0.89 and 0.86 for correct and interval ages, respectively, versus 0.34 and 0.29 for
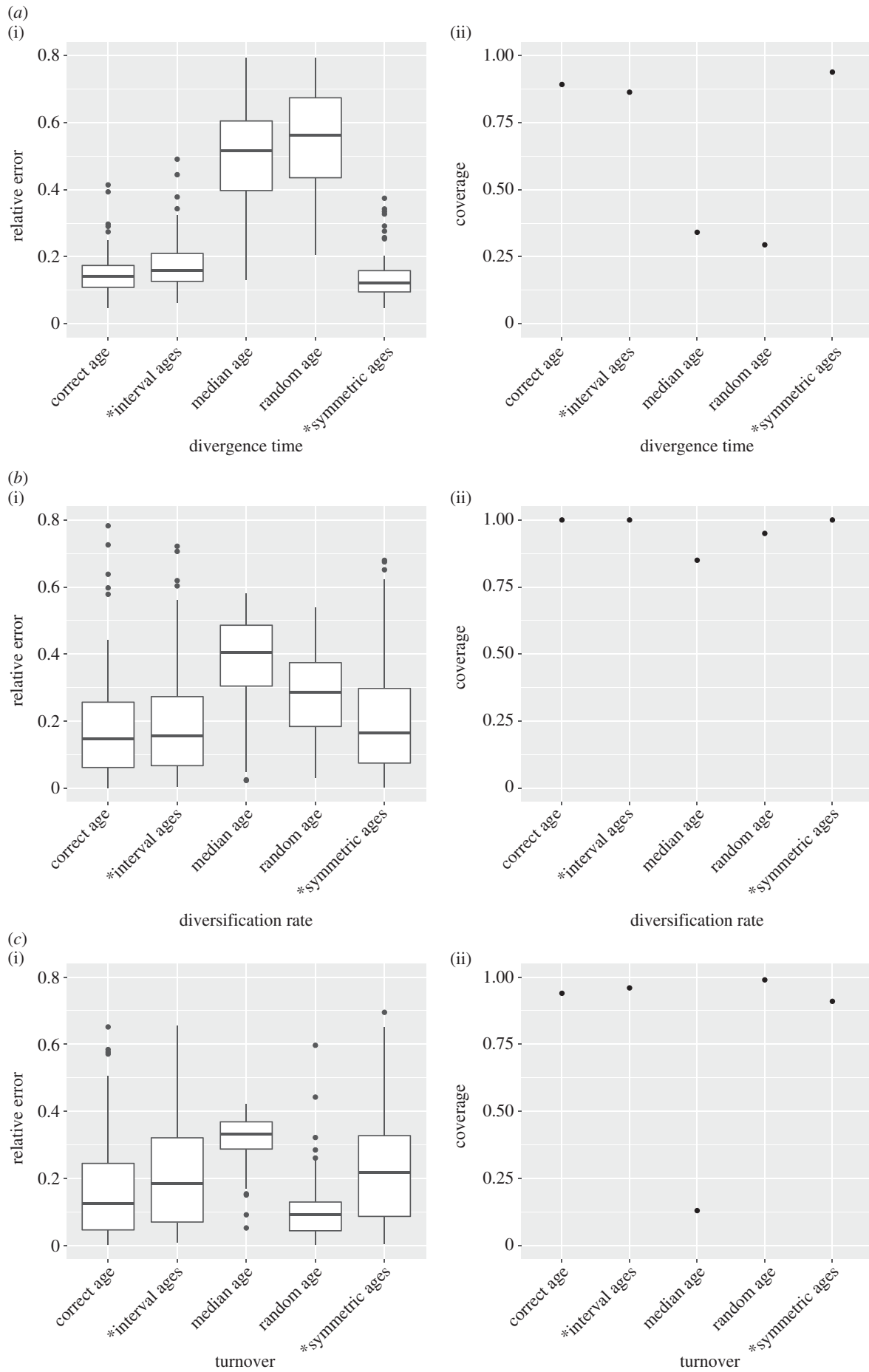
Figure 3. Average relative error of the median estimate (column 1) and 95% HPD coverage (column 2) achieved by different age handling methods for the following parameters: (a) divergence times of extant species, (b) diversification rate and (c) turnover. Ages sampled as part of the MCMC are marked by (*).

Table 2. Average relative width of the 95% HPD intervals obtained for different estimates using different methods. The relative width is calculated as the width of the HPD interval divided by the true value.

estimate	correct age	interval ages	median age	random age	symmetric ages
divergence times	0.69	0.76	1.05	1.29	0.72
diversification rate	1.60	1.58	1.18	1.16	1.67
turnover	0.87	0.91	0.36	0.78	0.92

Table 3. Average performance of the MCMC using different age handling methods. The performance is calculated as the CPU time (in seconds) per effective sample of the posterior distribution and of the total length of the tree.

time/ESS	correct age	median age	interval ages	random age	symmetric ages
posterior	188.2	36.8	163.9	131.7	150.4
total height	12.7	15.8	22.1	26.0	26.2

median and random ages, respectively. The diversification rate is less sensitive to the choice of fossil ages, but still shows a relative error of 0.42 and a coverage of 0.85 for median ages, compared with 0.18 and 1.0 for interval ages. Median ages result in inaccurate estimates of turnover, with a coverage of 0.13, compared with values above 0.9 for all other methods. Overall, there are important discrepancies between the results obtained using correct ages, interval ages and symmetric interval ages, versus median ages and random ages.

(ii) Precision

The relative 95% HPD interval widths are shown in table 2. Sampling fossil ages along with other model parameters (based on the PBDB or symmetric age intervals) did not result in wider HPD intervals than fixing the fossil ages to the truth. Fixing fossil ages to the wrong values (i.e. using median ages or random ages) did not have a consistent effect on HPD interval width. For example, HPD intervals were wider for the divergence time estimates but narrower for the diversification rate. This reveals that the better coverage obtained for interval ages compared to median or random ages were not obtained at the expense of precision in the case of divergence times (i.e. higher coverage is not simply due to wider HPD intervals).

(iii) Performance

We evaluated the performance of the different inference methods by calculating the processing time required per effective sample. We used the effective sample sizes for the posterior distribution and for the total height of the tree. The results are shown in table 3. We expected that sampling fossil ages would lead to an increase in computational cost; however, we observed no clear correlation between sampling or fixing fossil ages and the performance in this dataset. One possible explanation is that while sampling fossil ages does add a number of parameters to the model, these parameters are real numbers, which are less costly to sample than complex objects like the tree topology, and are constrained to a fairly narrow range. Some of these parameters are also constrained further by the effective prior distributions, which

(as mentioned in the Methods) were very narrow for some ages. This would contribute to a lower additional computational cost than expected. In summary, sampling the fossil ages did not significantly slow down analyses in our study.

(b) Empirical dataset

We discuss here the results obtained with 10% subsampling, the results with 5% subsampling are presented in the electronic supplementary material. Figure 4 shows the MCC trees obtained using different methods of fixing ages, restricted to extant tips. There are few differences in topology, which is expected as we applied strong topological constraints in this analysis. However, divergence time estimates vary considerably between different approaches to handling age uncertainty, and is most apparent for older nodes in the tree. For instance, the most recent common ancestor (MRCA) of all extant cetaceans is 44 Ma using interval ages, in contrast to 50 and 61 Ma using random and median ages, respectively. The relative difference between the median node ages inferred with interval ages versus median ages, averaged across all nodes, is 15%. However, there is wide variability; some nodes show almost no difference (less than 1%), while in other cases the median age estimate obtained using median ages is double the estimate obtained using interval ages. The relative length of the 95% HPD intervals for the divergence time estimates is 52% of the median estimate for interval ages, 54% for median ages and 26% for random ages, also averaged across all nodes. Thus using random ages lead to much narrower posterior distributions for the divergence time estimates.

An example of the strong influence exerted by fixing fossil ages on estimated node ages is shown in figure 5. We can see that the posterior distribution obtained using interval ages is much wider. However, when using median or random ages, the age of the node is constrained to within a much narrower interval. The fossil specimen imposes a lower bound on the distribution that is potentially in conflict with the phylogenetic data and/or other age constraints, resulting in a posterior distribution with a strong peak at the lower bound. For this node, the 95% HPD interval is of length 2.92 for the interval ages, 3.91 for median ages and 2.65 for random ages.
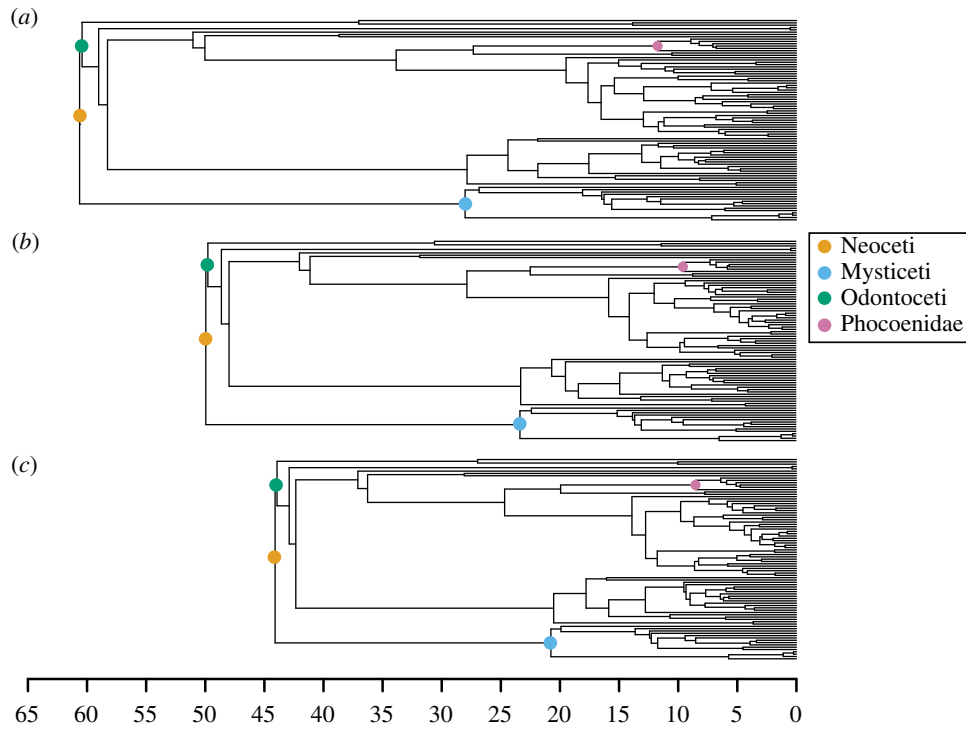
Figure 4. MCC trees inferred for the Cetacea dataset using the FBD process with fossil ages fixed to (*a*) median ages, (*b*) random ages or (*c*) sampled within the known interval of uncertainty. The major clades and the clade shown in figure 5 are highlighted.
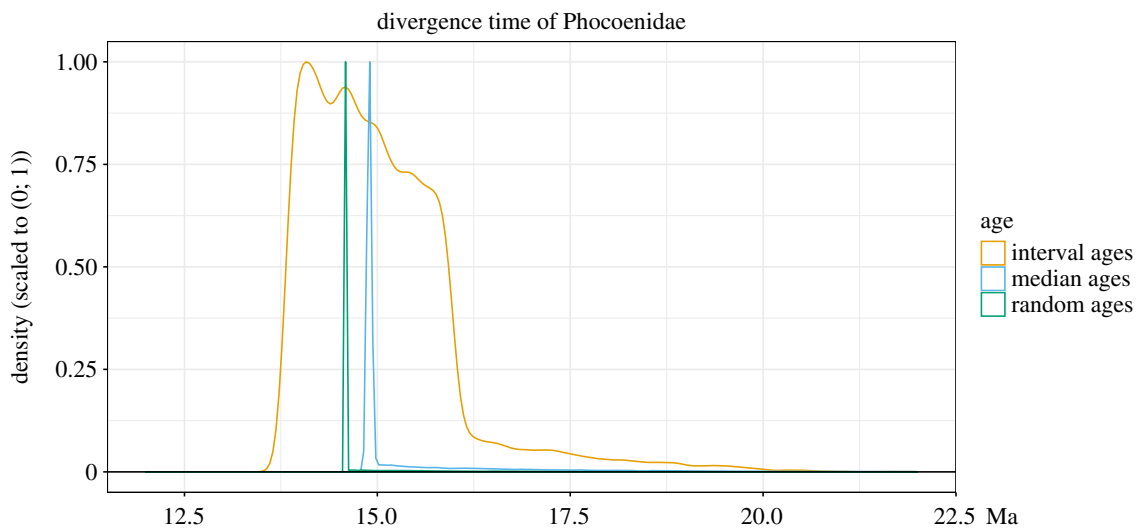

Figure 5. Posterior density obtained for the most recent common ancestor of the family Phocoenidae in the Cetacea dataset using the FBD process with fossil ages fixed to median ages, random ages or sampled within the known interval of uncertainty. The densities were scaled to the interval (0; 1).

Figure 6 shows a comparison of the estimates of the FBD process parameters obtained using different methods of age handling. For these parameters, all estimates show a trend going from median to random to interval ages. The diversification rate is most robust to the choice of fossil age, as all the HPD intervals overlap. However, we see a trend for an increasing diversification rate estimate, from median to random to interval ages. The turnover is estimated to be higher and the sampling proportion much lower with median ages than with interval ages. These trends in parameter estimates are likely to be in part correlated to the observed differences in divergence time estimates. For example, as the number of fossil samples increases with the

age of the process and with the sampling rate, estimating a greater height for the tree leads to lower estimates of sampling proportion and vice-versa. Similarly, as the number of extant species increases with the age of the process and with the diversification rate, estimating a greater height for the tree leads to lower estimates of diversification rates and vice-versa.

4. Discussion

The age of fossils is not known precisely and instead is associated with a range of uncertainty that results from the nature of the geological record. Our simulations demonstrate that
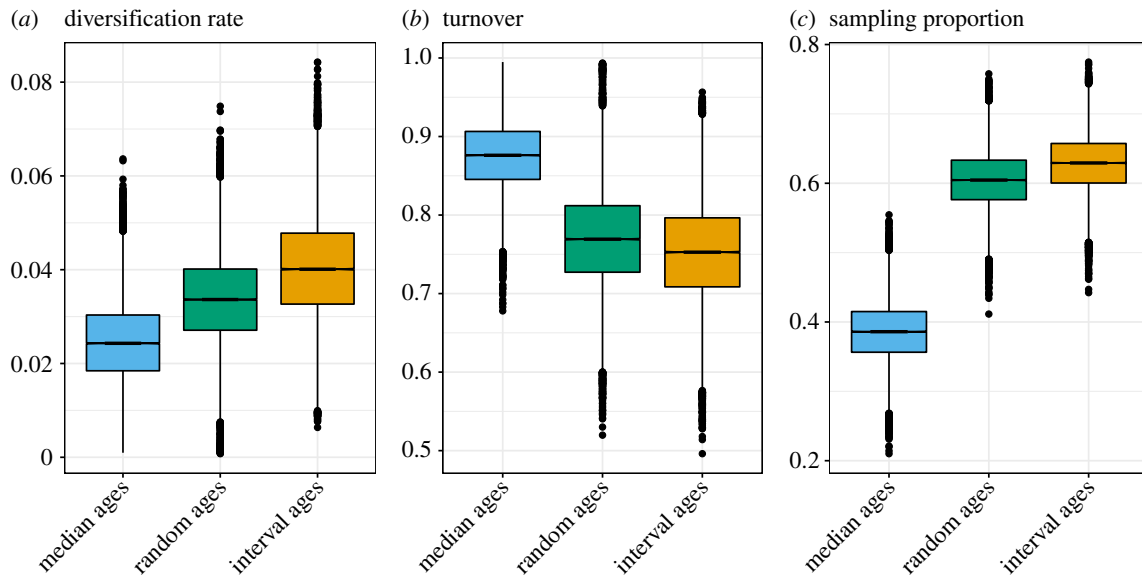
Figure 6. Estimates of the (a) diversification rate, (b) turnover and (c) sampling proportion obtained for the Cetacea dataset using the FBD process with fossil ages fixed to median ages, random ages or sampled within the known interval of uncertainty.

the popular approach of fixing fossil ages to point estimates within the known interval of uncertainty in molecular dating analysis can led to erroneous estimates of node ages. Although our simulated datasets were designed to represent a best-case scenario, with complete extant sampling and high rates of fossil recovery, less than 35% of node ages were recovered within the Bayesian 95% HPD intervals when fixing fossil ages. Conversely, explicitly incorporating fossil age uncertainty led to substantially more accurate estimates of node ages and other model parameters. It may be tempting to fix the age of fossils to reduce the computational burden. However, our results based on both the simulated and empirical datasets show that co-estimating fossil ages does not necessarily incur additional computational cost. Unless researchers have extremely good motivation for doing so, fossil ages should not be fixed in divergence time analysis that directly incorporate fossils into the tree. Emphasis on establishing rigorous and transparent calibration information that evolved in the context of node dating should also apply to other approaches to calibration, including inference under the FBD process [17,24,49].

The analysis of the cetacean dataset illustrates how strongly estimates of both divergence times and biologically relevant parameters, such as diversification rate, can be affected by the choices made in handling fossil age uncertainty. The difference between the median posterior estimate obtained for the age of Neoceti (crown cetaceans) using fixed median ages versus interval ages is nearly 20 Ma. Furthermore, analysing this dataset with median ages shows a strong discrepancy between the divergence time estimates obtained using the FBD model and the estimates obtained using fossil calibrations in the original node dating analysis [41], which estimated the origin of the Neoceti to be 36 Ma. However, accounting for stratigraphic age uncertainty shows that this is not the case: the MRCA age obtained using interval ages matches more closely both with the original analysis and with more recent studies such as [44], which estimated the MRCA age to be 39 Ma. It is not possible to definitively determine which outcome is closer to the truth in our empirical analysis, but our simulations clearly indicate

that estimates obtained using interval ages should be considered the most reliable.

It is worth noting that the average range of age uncertainty associated with fossils included in our simulations and empirical analysis is relatively small (8 and 4 Myr, respectively). This reflects our decision to focus on well-studied Cenozoic fauna with extant representatives, but the fossil record of many taxonomic groups and time periods will be associated with much greater uncertainty. For example, the age of many pre-Cenozoic deposits are poorly constrained. Thus, the discrepancies obtained using different age handling methods have the potential to be much larger than those demonstrated in this study. This may be especially important to consider in the context of FBD analyses for groups that have no extant representatives.

Our simulations focused on the scenario in which some prior information about the phylogenetic position of fossils is known: no character data were included for extinct samples, meaning fossil recovery times informed the FBD model parameters, but the phylogenetic position of these samples could not be estimated. If morphological character data are available for fossils, their phylogenetic position can also be inferred along with divergence times [48,50], meaning no taxonomic decisions have to be made *a priori* by assigning fossils to clades as done here. The inclusion of morphological data has been shown to improve estimates of divergence times, assuming that the clock and substitution models are appropriate and the age uncertainty is properly handled [22,38,50]. Investigating the impact of age uncertainty when using morphological data will be an interesting avenue for future simulation and empirical studies.

Finally, it is worth emphasizing the distinction between the age range associated with fossil specimens and fossil species [18]. The latter is known as the stratigraphic range of a species, and represents the interval between the first and last appearance times. Here, we implemented the specimen-based FBD process, meaning all available specimens were incorporated into the analysis. Although we note some studies have applied this model to the analysis of stratigraphic range data, this is not technically appropriate.

Instead, stratigraphic ranges should be analysed under the FBD range process [51], however no implementation in BEAST2 is yet available. We note that when this model does become available, the uncertainty associated with specimens representing the ends of stratigraphic ranges should be incorporated into the analysis, rather than being fixed, otherwise we anticipate similar performance issues to those demonstrated in this study.

5. Conclusion

In this study, we demonstrate that the choice of method for handling fossil age uncertainty can have important effects on estimates of species divergence times obtained under the FBD process. Our simulation results clearly favour a Bayesian hierarchical approach to handling fossil age uncertainty based on the actual age intervals, as opposed to fixing the ages to an arbitrary value inside that interval. In addition, our empirical dataset demonstrates that the rigid age constraints given by fixed fossil ages can lead to age estimates that are very different from those obtained using a traditional node dating approach, whereas a more flexible approach to handling fossil ages recovers similar estimates. Thus we strongly recommend against fixing fossil ages in FBD analyses. Overall, this work illustrates the critical importance of accurately reflecting the available information regarding uncertainty in Bayesian phylogenetic analyses. As we demonstrate, discarding this information can have detrimental impacts on the accuracy of the results.

References

1. Kishino H, Thorne JL, Bruno WJ. 2001 Performance of a divergence time estimation method under a probabilistic model of rate evolution. *Mol. Biol. Evol.* **18**, 352–361. (doi:10.1093/oxfordjournals.molbev.a003811)

2. Yang Z, Rannala B. 2005 Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. *Mol. Biol. Evol.* **23**, 212–226. (doi:10.1093/molbev/msj024)

3. Drummond AJ, Ho SY, Phillips MJ, Rambaut A. 2006 Relaxed phylogenetics and dating with confidence. *PLoS Biol.* **4**, e88. (doi:10.1371/journal.pbio.0040088)

4. Ho SY, Phillips MJ. 2009 Accounting for calibration uncertainty in phylogenetic estimation of evolutionary divergence times. *Syst. Biol.* **58**, 367–380. (doi:10.1093/sysbio/syp035)

5. Clarke JT, Warnock RC, Donoghue PC. 2011 Establishing a time-scale for plant evolution. *New Phytol.* **192**, 266–301. (doi:10.1111/j.1469-8137.2011.03794.x)

6. Warnock RC, Yang Z, Donoghue PC. 2012 Exploring uncertainty in the calibration of the molecular clock. *Biol. Lett.* **8**, 156–159. (doi:10.1098/rsbl.2011.0710)

7. Heath TA. 2012 A hierarchical Bayesian model for calibrating estimates of species divergence times. *Syst. Biol.* **61**, 793–809. (doi:10.1093/sysbio/sys032)

8. Joyce WG, Parham JF, Lyson TR, Warnock RC, Donoghue PC. 2013 A divergence dating analysis of turtles using fossil calibrations: an example of best practices. *J. Paleontol.* **87**, 612–634. (doi:10.1666/12-149)

9. Heled J, Drummond AJ. 2011 Calibrated tree priors for relaxed phylogenetics and divergence time estimation. *Syst. Biol.* **61**, 138–149. (doi:10.1093/sysbio/syr087)

10. Warnock RC, Parham JF, Joyce WG, Lyson TR, Donoghue PC. 2015 Calibration uncertainty in molecular dating analyses: there is no substitute for the prior evaluation of time priors. *Proc. R. Soc. B* **282**, 20141013. (doi:10.1098/rspb.2014.1013)

11. Stadler T. 2010 Sampling-through-time in birth–death trees. *J. Theor. Biol.* **267**, 396–404. (doi:10.1016/j.jtbi.2010.09.010)

12. Heath TA, Huelsenbeck JP, Stadler T. 2014 The fossilized birth–death process for coherent calibration of divergence-time estimates. *Proc. Natl Acad. Sci. USA* **111**, E2957–E2966. (doi:10.1073/pnas.1319091111)

13. Erwin DH. 2006 Dates and rates: temporal resolution in the deep time stratigraphic record. *Annu. Rev. Earth Planet. Sci.* **34**, 569–590. (doi:10.1146/annurev.earth.34.031405.125141)

14. Reisz RR, Müller J. 2004 Molecular timescales and the fossil record: a paleontological perspective. *Trends Genet.* **20**, 237–241. (doi:10.1016/j.tig.2004.03.007)

15. Benton MJ, Donoghue PC. 2006 Paleontological evidence to date the tree of life. *Mol. Biol. Evol.* **24**, 26–53. (doi:10.1093/molbev/msl150)

16. Benton M, Donoghue P, Asher R. 2009 Calibrating and constraining molecular clocks. In *The timetree of life* (eds SB Hedges, S Kumar), pp. 35–86. Oxford, UK: Oxford University Press.

17. Parham JF, *et al.* 2011 Best practices for justifying fossil calibrations. *Syst. Biol.* **61**, 346–359. (doi:10.1093/sysbio/syr107)

18. Hopkins MJ, Bapst DW, Simpson C, Warnock RC. 2018 The inseparability of sampling and time and its influence on attempts to unify the molecular and fossil records. *Paleobiology* **44**, 561–574. (doi:10.1017/pab.2018.27)

19. Larabee FJ, Fisher BL, Schmidt CA, Matos-Maraví P, Janda M, Suarez AV. 2016 Molecular phylogenetics and diversification of trap-jaw ants in the genera Anochetus and Odontomachus (Hymenoptera: Formicidae). *Mol. Phylogenet. Evol.* **103**, 143–154. (doi:10.1016/j.ympev.2016.07.024)

20. Close RA, Johanson Z, Tyler JC, Harrington RC, Friedman M. 2016 Mosaicism in a new Eocene pufferfish highlights rapid morphological innovation near the origin of crown tetraodontiforms. *Palaeontology* **59**, 499–514. (doi:10.1111/pala.2016.59.issue-4)

21. Grimm GW, Kapli P, Bomfleur B, McLoughlin S, Renner SS. 2015 Using more than the oldest fossils: dating Osmundaceae with three Bayesian clock approaches. *Syst. Biol.* **64**, 396–405. (doi:10.1093/sysbio/syu108)

22. Drummond AJ, Stadler T. 2016 Bayesian phylogenetic estimation of fossil ages. *Phil. Trans. R. Soc. B* **371**, 20150129. (doi:10.1098/rstb.2015.0129)

23. Ho SYW, Molak M, Lorenzen ED, Shapiro B. 2012 Phylogenetic estimation of timescales using ancient DNA: the effects of temporal sampling scheme and uncertainty in sample ages. *Mol. Biol. Evol.* **30**, 253–262. (doi:10.1093/molbev/mss232)

24. O'Reilly JE, dos Reis M, Donoghue PC. 2015 Dating tips for divergence-time estimation. *Trends Genet.* **31**, 637–650. (doi:10.1016/j.tig.2015.08.001)

25. Luo A, Duchene D, Zhang C, Zhu CD, Ho S. 2018 A simulation-based evaluation of total-evidence

dating under the fossilized birth–death process. *bioRxiv* 436303. (doi:10.1101/436303)

26. Bininda-Emonds OR, *et al.* 2007 The delayed rise of present-day mammals. *Nature* **446**, 507. (doi:10.1038/nature05634)

27. dos Reis M, Inoue J, Hasegawa M, Asher RJ, Donoghue PC, Yang Z. 2012 Phylogenomic datasets provide both precision and accuracy in estimating the timescale of placental mammal phylogeny. *Proc. R. Soc. B* **279**, 3491–3500. (doi:10.1098/rspb.2012.0683)

28. Close RA, Friedman M, Lloyd GT, Benson RB. 2015 Evidence for a mid-Jurassic adaptive radiation in mammals. *Curr. Biol.* **25**, 2137–2142. (doi:10.1016/j.cub.2015.06.047)

29. Stadler T. 2011 Mammalian phylogeny reveals recent diversification rate shifts. *Proc. Natl Acad. Sci. USA* **108**, 6187–6192. (doi:10.1073/pnas.1016876108)

30. Meredith RW, *et al.* 2011 Impacts of the cretaceous terrestrial revolution and KPg extinction on mammal diversification. *Science* **334**, 521–524. (doi:10.1126/science.1211028)

31. Stadler T. 2011 Simulating trees with a fixed number of extant species. *Syst. Biol.* **60**, 676–684. (doi:10.1093/sysbio/syr029)

32. Pennell MW, Sarver BA, Harmon LJ. 2012 Trees of unusual size: biased inference of early bursts from large molecular phylogenies. *PLoS ONE* **7**, e43348. (doi:10.1371/journal.pone.0043348)

33. Wagner PJ, Marcot JD. 2013 Modelling distributions of fossil sampling rates over time, space and taxa: assessment and implications for macroevolutionary studies. *Methods Ecol. Evol.* **4**, 703–713. (doi:10.1111/mee3.2013.4.issue-8)

34. Silvestro D, Warnock RCM, Gavryushkina A, Stadler T. 2018 Closing the gap between palaeontological and neontological speciation and extinction rate estimates. *Nat. Commun.* **9**, 5237. (doi:10.1038/s41467-018-07622-y)

35. Rambaut A, Grass NC. 1997 Seq-Gen: an application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. *Bioinformatics* **13**, 235–238. (doi:10.1093/bioinformatics/13.3.235)

36. Chen WC. 2011 Overlapping codon model, phylogenetic clustering, and alternative partial expectation conditional maximization algorithm. PhD thesis, Iowa State University, Ames, IA.

37. Warnock RCM, Yang Z, Donoghue PCJ. 2017 Testing the molecular clock using mechanistic models of fossil preservation and molecular evolution. *Proc. R. Soc. B* **284**, 20170227. (doi:10.1098/rspb.2017.0227)

38. Gavryushkina A, Welch D, Stadler T, Drummond AJ. 2014 Bayesian inference of sampled ancestor trees for epidemiology and fossil calibration. *PLoS Comput. Biol.* **10**, e1003919. (doi:10.1371/journal.pcbi.1003919)

39. Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu CH, Xie D, Suchard MA, Rambaut A, Drummond AJ. 2014 BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* **10**, 1–6. (doi:10.1371/journal.pcbi.1003537)

40. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. 2018 Posterior summarization in Bayesian phylogenetics using tracer 1.7. *Syst. Biol.* **67**, 901–904. (doi:10.1093/sysbio/syy032)

41. Steeman ME, *et al.* 2009 Radiation of extant cetaceans driven by restructuring of the oceans. *Syst. Biol.* **58**, 573–585. (doi:10.1093/sysbio/syp060)

42. McGowen MR, Spaulding M, Gatesy J. 2009 Divergence date estimation and a comprehensive molecular tree of extant cetaceans. *Mol. Phylogenet. Evol.* **53**, 891–906. (doi:10.1016/j.ympev.2009.08.018)

43. Slater GJ, Price SA, Santini F, Alfaro ME. 2010 Diversity versus disparity and the radiation of modern cetaceans. *Proc. R. Soc. B* **277**, 3097–3104. (doi:10.1098/rspb.2010.0408)

44. Marx FG, Fordyce RE. 2015 Baleen boom and bust: a synthesis of mysticete phylogeny, diversity and disparity. *R. Soc. open sci.* **2**, 140434. (doi:10.1098/rsos.140434)

45. Slater GJ, Goldbogen JA, Pyenson ND. 2017 Independent evolution of baleen whale gigantism linked to Plio-Pleistocene ocean dynamics. *Proc. R. Soc. B* **284**, 20170546. (doi:10.1098/rspb.2017.0546)

46. Marx F, Lambert O, Uhen MD. 2016 *Cetacean paleobiology*. New York, NY: Wiley-Blackwell.

47. Bouckaert RR, Drummond AJ. 2017 bModelTest: Bayesian phylogenetic site model averaging and model comparison. *BMC Evol. Biol.* **17**, 42. (doi:10.1186/s12862-017-0890-6)

48. Zhang C, Stadler T, Klopfstein S, Heath TA, Ronquist F. 2015 Total-evidence dating under the fossilized birth–death process. *Syst. Biol.* **65**, 228–249. (doi:10.1093/sysbio/syv080)

49. Ksepka DT, *et al.* 2015 The fossil calibration database – a new resource for divergence dating. *Syst. Biol.* **64**, 853–859. (doi:10.1093/sysbio/syv025)

50. Gavryushkina A, Heath TA, Ksepka DT, Stadler T, Welch D, Drummond AJ. 2017 Bayesian total-evidence dating reveals the recent crown radiation of penguins. *Syst. Biol.* **66**, 57–73. (doi:10.1093/sysbio/syw060)

51. Stadler T, Gavryushkina A, Warnock RCM, Drummond AJ, Heath TA. 2018 The fossilized birth–death model for the analysis of stratigraphic range data under different speciation modes. *J. Theor. Biol.* **447**, 41–55. (doi:10.1016/j.jtbi.2018.03.005)