Palaeontology
The Palaeontological Association
www.palass.org

[Palaeontology, 2021, pp. 1–11]

FOSSILIZATION PROCESSES HAVE LITTLE IMPACT ON TIP-CALIBRATED DIVERGENCE TIME ANALYSES

by JOSEPH E. O'REILLY[1,2] *and* PHILIP C. J. DONOGHUE[1]

[1]School of Earth Sciences, University of Bristol, Life Sciences Building, Tyndall Avenue, Bristol, BS8 1TQ, UK; joseph.oreilly@igmm.ed.ac.uk; phil.donoghue@bristol.ac.uk
[2]MRC Institute of Genetics & Molecular Medicine, University of Edinburgh, Edinburgh, UK;

Typescript received 23 August 2020; accepted in revised form 17 May 2021

Abstract: The importance of palaeontological data in divergence time estimation has increased with the introduction of Bayesian total-evidence dating methods, which use fossil taxa directly for calibration, facilitated by the joint analysis of morphological and molecular data. Fossil taxa are invariably incompletely known as a consequence of taphonomic processes, resulting in the decidedly non-random distribution of missing data. The impact of non-random missing data on the accuracy and precision of clade age estimation is unknown. In an attempt to constrain the impact of taphonomy on tip-calibrated dating analyses, we compared clade ages estimated from a very complete morphological matrix to ages estimated from the same matrix permuted to simulate the progressive loss of anatomical information resulting from taphonomic processes. We demonstrate that systematically distributed missing data negatively influence clade age estimates, but that successive stages within the taphonomic process introduce greater differences in age estimates, when compared to estimates obtained from untreated data. Despite these effects, the general influence of missing data is weak, presumably due to the compensatory effect of extensive morphological data from extant taxa. We suggest that, in the absence of models that can explicitly account for taphonomic processes, morphological datasets should be constructed to minimize the impact of taphonomy on divergence time estimation.

Key words: morphology, divergence time estimation, missing data, taphonomy, Bayesian inference, phylogenetics.

EVOLUTIONARY timescales are essential to effect tests of hypotheses on the coevolution of Earth and life. Molecular clock methodology has effectively displaced direct interpretation of the geological and fossil record in this endeavour (Zuckerkandl & Pauling 1965; Thorne *et al.* 1998; Drummond *et al.* 2006; Yang & Rannala 2006; Donoghue & Yang 2016). Nevertheless, fossil data remain integral to divergence time estimation, in calibrating the rate of molecular evolution to time. This has traditionally been achieved indirectly through node calibration, using fossil and geological evidence to constrain probabilistically the minimum and maximum age of clades (Donoghue & Benton 2007; Parham *et al.* 2012). Tip-calibration overcomes challenges associated with the indirect interpretation of fossils to inform node-calibrations, allowing fossil taxa to inform molecular clock analyses directly, including them en par with their living relatives through the inclusion of a morphological dataset and model of evolution (Lewis 2001). This approach is particularly attractive since it allows all fossil species to be included in divergence time analyses, not just those informing the age of extant clades. Through co-estimation of time and topology in 'total evidence dating' and 'tip-calibration' (TED; Ronquist *et al.* 2012*a*; Pyron 2011) the phylogenetic uncertainty of fossil species can be controlled for. However, fossil taxa are invariably incompletely preserved and the extent to which these biases impact molecular clock analyses has not yet been explored (Sansom & Wills 2013; O'Reilly *et al.* 2015).

The impact of missing data on the accuracy of topology estimates has been studied extensively, but almost all such studies have assumed that missing data are randomly distributed (Sansom & Wills 2013; Guillerme & Cooper 2016*a*). This is unlikely even for living species because of research agenda resulting, for instance, from the study of organ systems. Random distribution of missing data is not expected in fossil taxa either, where biases in the processes of decay and preservation lead to the progressive, but non-random, loss of biological information. Long before the chemical processes of preservation and diagenesis begin, a prospective fossil may experience physical biostratinomic effects. These include decay, post-mortem collapse, disarticulation, fragmentation, erosion and differential transportation (Behrensmeyer & Kidwell 1985). The nature and degree of biological information loss is dependent on whether the effects of these processes

doi: 10.1111/pala.12564

can be diminished, such as through early burial and mineralization, and this is invariably dependent on the nature of the depositional environment (Behrensmeyer & Kidwell 1985). At the most general level, soft tissue anatomy quickly rots away; only biomineralized anatomical elements are routinely fossilized, but even these have differential preservation potential, with lighter and more fragile skeletal elements most likely to be transported away, fragmented and eroded (Behrensmeyer 1990). Hence, missing data are systematically distributed across fossil taxa, associated with specific classes of anatomical characters (Fig. 1). The non-random distribution of missing data has been shown to have a biasing effect on topology estimation within a parsimony framework (Sansom & Wills 2013) but its impact on tip-calibration and TED is unknown.

Here we explore the impact of non-random missing data on the accuracy and precision of divergence time analyses that employ tip-calibration. Traditionally, the effect of missing data has been investigated using simulation analyses in which the true tree is known and missing data are introduced randomly. However, designing simulations that approach the disparate fossilization process that vary both with intrinsic biology and the nature of the post-mortem environment, is challenging. Instead we use a large, complete, empirical morphological matrix as the basis for our analyses in which we simulate the progressive loss of biological information that results from decay and biostratinomic processes, informed by empirical evidence for the proportional loss of anatomical characters. The ages estimated from the complete, unfossilized, dataset are used as a benchmark against which the effects on divergence times estimated using datasets with artificial fossilization are measured.

MATERIAL AND METHOD

Given our goal, to determine the influence of non-random missing morphological data in divergence time analyses, we required a dataset with a sufficiently large quantity of characters, such that branch lengths could be estimated with an acceptable level of accuracy both before and after artificial taphonomic biases have been introduced. We used the 4541 morphological character dataset from O'Leary *et al.* (2013), which encompasses the diversity of placental mammals, as the basis for our experiments.

To prepare the data for the simulation of the effects of the fossilization process, we removed fossil taxa that are not members of crown-Mammalia, converted polymorphisms and ambiguities to missing data, and removed both characters and taxa that have more than 25% missing data (inapplicable characters were not considered to be missing data for this purpose as knowledge of a conditional relationship, as implied by an inapplicability, would not be possible if the data defining the conditional relationship were missing due to taphonomic effects). This resulted in a dataset composed of 66 taxa, of which 20 taxa were extinct. Using this approximately complete dataset (95.52% of cells in the dataset did not contain missing data; 88.07% of cells belonging to extinct taxa did not contain missing data) we adopted two approaches to simulating the effects of fossilization: (1) biostratinomic processes (transport, abrasion, erosion, fragmentation, etc.); and (2) the loss of characters across all fossil taxa, reflecting the decay of soft tissues.

Simulating the effects of post-decay biostratinomic processes

Aslan & Behrensmeyer (1996) identified categories of mammalian osteological characters that are likely to be preserved in a fluvial deposit (the environment in which the majority of terrestrial vertebrates are preserved), with a probability of recovery reported for each category. The O'Leary dataset was reduced to only those characters relevant to the categories identified by Aslan & Behrensmeyer (1996), resulting in a dataset of 2454 characters. We randomly sampled a selection of extinct taxa and converted character codings to missing data randomly within each class of characters, such that the proportion of non-missing data matched the probability of recovery as reported by (Aslan & Behrensmeyer 1996). This approach resulted in ~10% of the total matrix of extinct and extant taxa being composed of missing data. To further investigate the influence of increasingly extreme levels of missing data we multiplied the proportion of missing data reported by Aslan & Behrensmeyer (1996) in each character category by 2, as a separate case, replacing observed character data with missing data to match these larger proportions. For example, if a class of characters exhibited 20% missing data in the original empirical analysis, for the scaled simulation we introduced missing data into this class such that a total of 40% of the data was missing. A scaling factor of 2 resulted in a matrix with ~95% missing data for fossil taxa. A survey of 286 mammalian morphological matrices (Guillerme & Cooper 2016*b*) showed that the mean percentage of missing data was 21% (range = 0–73%; Fig. S1), suggesting that the quantities of missing data we produced through simulation were generally empirically realistic. The only exception being the analyses with a scaling factor of 2, with such matrices representing an extreme level of missing data unlikely to be encountered regularly in empirical datasets. The process of randomly introducing missing data was repeated 30 times to produce 30 replicate matrices (with or without a scaling factor) in which a different random subsample of

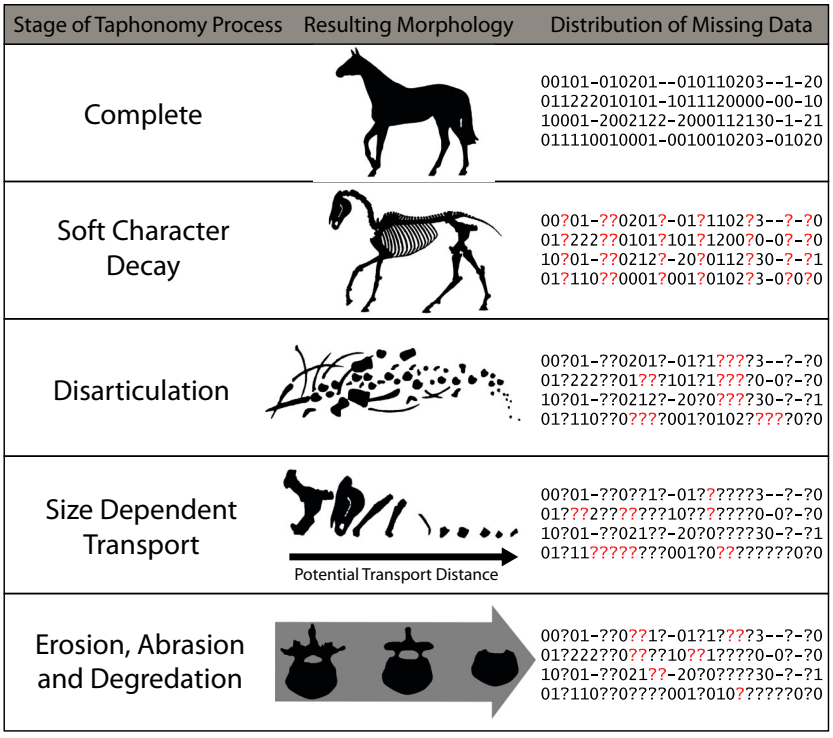

FIG. 1. The effect of different stages of the fossilization process on the distribution of missing data (characters in red font) in morphological matrices. Starting with a complete matrix before the influence of any stage of the fossilization process, the loss of soft characters through decay introduces a character-wise distribution of missing data as these characters are unlikely to preserve for any fossil taxon. Physical biostratinomic processes introduce further missing data in a taxon-wise manner as a number of characters from morphological structures are lost simultaneously. Disarticulation and size dependent transport leads to taxon specific loss of large numbers of characters associated with lost morphological structures. Erosion and abrasion lead to further loss of detailed characters that are worn away before deposition.

taxa was selected within each character class. This resulted in 60 simulated morphological matrices.

Simulating the loss of entire characters

To enable a straightforward comparison of the effect of different taphonomic effects, we used the 2454 character matrix produced for the simulation of biostratinomic processes when simulating the loss of entire characters. Of the 4541 characters in the original O'Leary *et al.* (2013) matrix, 19% are unlikely to be preserved through routine modes of fossilization. Thus, we selected at random 19% of the characters in the 2454 character matrix and changed their coding to missing data for all fossil taxa. This process was repeated to obtain 30 distinct matrices. The initial filter which reduced the dataset to 2454 characters removed many soft characters. Therefore, it is important to note that many of the characters selected to be missing in this simulation may be osteological or routinely preserved characters. Despite this, the patterns and quantities of missing data that our simulation procedure produced match those of the complete empirical matrix.

Exploring the effects of character matrix dimensions

Though our experimental dataset has just over half of the number of characters of the original (O'Leary *et al.* 2013), the remaining 2454 characters comprise an unusually large empirical phenotype dataset. Thus, we also explored the effects of taphonomic biases in datasets that approximate the reduced dimensions of the majority of phylogenetic datasets, which are typically composed of characters measured in the low hundreds. To achieve this, we produced five smaller matrices in which each character class is reduced to a random subsample of 10% from the original matrix, resulting in 245 characters. For each of the five reductions, 30 datasets were produced using the post-decay biostratinomic filter outlined above.

Divergence time estimation

Divergence time analyses based on these matrices were performed using MrBayes 3.2 (Ronquist *et al.* 2012*b*) In all analyses, we employed a fixed topology compatible with the results of the combined molecular and morphological analysis of O'Leary *et al.* (2013). We were not interested in the accuracy of the fixed topology but, rather, used it to isolate the effects of the introduction of non-random missing data, without the confounding effect of topology estimation. The root calibration was taken from Benton *et al.* (2015) and assigned an offset exponential distribution; tip calibrations were assigned uniform distributions, or point estimates, based on the ranges or single point ages reported by O'Leary *et al.*

(2013). A diffuse prior on the morphological rate was set as N(0.001, 0.01), based on the prior for morphological rate of Beck & Lee (2014). The independent gamma rate (IGR) clock model (Lepage *et al.* 2007) was used and assigned a variance parameter prior of *exp*(10). Morphological data were analysed with the Mkv model (Lewis 2001) and the uniform tree prior was applied. The mcmc approximation of the posterior distribution was performed for 40 000 000 generations, sampling every 4000 generations over 4 runs of 4 chains. Convergence was assumed when ESS scores of greater than 200 were observed and based on a qualitative assessment of the stationarity of the chain in Tracer (Rambaut *et al.* 2014). Consensus trees were constructed after a conservative burn-in of 25%, after which sampling was deemed to be from the posterior distribution.

Averaging over consensus trees

For each analysis, 30 replicate consensus trees were produced and, for each constituent node, we obtained the mean age estimate over the 30 replicates, producing a single tree that encompasses the results from 30 replicate analyses. These averaged trees were then compared to a consensus tree estimated from the untreated matrix from which the fossilized matrices were derived. We also considered aspects of the distribution of the 30 individual replicate trees before averaging over their node ages. In a small number of cases, the construction of consensus trees resulted in the introduction of polytomies when an internal branch length estimate was of negligible length; in these instances we discarded the results as a direct comparison between the divergence times in the treated and untreated consensus trees could only be made if the topology of the trees was identical.

Identifying patterns of age estimate bias

Our aim was to test whether distributions of missing data arising from different taphonomic biases result in distinct age estimates. For simulation-based divergence time analyses, it is common to use coverage as a measure of accuracy, the proportion of replicates that possess a 95% highest posterior density (HPD) that encompasses the generating age (Gavryushkina *et al.* 2014; Heath *et al.* 2014; Warnock *et al.* 2017). However, in this instance we do not know the generating ages, and so instead we assessed accuracy based on the percentage of the 95% HPD obtained from simulated data that resides within the bounds of the 95% HPD obtained when untreated data were analysed. These values provide insight into the general differences in marginal distributions of age estimates that are introduced by specific taphonomic processes, but if we wish to identify specific patterns of bias in age estimates we cannot rely on the 95% HPD alone.

To identify the impact of taphonomic processes on age estimates we used the median of the marginal posterior distribution of any given node age to identify where the bulk of posterior probability density is concentrated. The paired differences of median age estimates (PDMA) from untreated and treated datasets then provides an overview of the magnitude of the effect of taphonomic processes, but this measure is more informative about systematic effects on age estimates caused by such taphonomic processes. We also calculated the percentage error for each median node age estimate, by dividing the PDMA by the age estimate obtained from the untreated data and multiplying this value by 100. We also calculated the percentage absolute error by taking the absolute value of the PDMA when calculating percentage error. We aggregated PDMA, HPD overlap and percentage error across all nodes in all 30 replicate analyses, and across each node in all 30 replicate analyses (i.e. the average PDMA/HPD overlap/percentage error for each node across multiple replicate analyses), as separate measures. The distribution of PDMA across all nodes in all replicate analyses provides the best characterization of fundamental patterns of bias introduced by particular taphonomic processes. The alternative aggregation approach provides a better characterization of the pattern of PDMA and percentage error for specific nodes across replicate analyses.

The matrices used in this study are available in O'Reilly & Donoghue (2021).

RESULTS

Loss of data resulting from the loss of entire characters across all fossil taxa

Full matrix. The average percentage of missing data in each simulated matrix was 9.65%, with 28.99% the mean quantity for fossil taxa (Table 1). There is considerable overlap between the 95% HPDs for each node averaged over the 30 replicates and the respective 95% HPDs obtained with untreated data (mean HPD per-node overlap = 98.36%; range = 93.05–100.0; Fig. 2). The mean PDMA is −0.13 myr, with a p-value of 1.97×10^{-12} for the null hypothesis that the PDMA is 0. The distribution of PDMA shows a propensity for treated matrices to result in the estimation of node ages that are slightly older than their counterparts obtained from untreated data (Fig. 3). The distribution of PDMA on the tree topology shows that there is a tendency for node ages that are either under or over estimated to be close to other nodes exhibiting the same direction and magnitude of paired difference.

Reduced matrices. Mean PDMA for the five reductions ranges from 0.01 myr to 0.27 myr. As in the results obtained using the full matrix, there is considerable overlap between 95% HPDs averaged over the 30 replicates and the respective 95% HPDs obtained with untreated data (Fig. 2).

Loss of data resulting from the effects of later biostratinomic processes

Full matrix. When the scaling factor was 1, the average percentage of missing data in each simulated matrix was 10.57%, with 32.04% the mean quantity for fossil taxa (Table 1). The mean HPD overlap is large (mean HPD per-node overlap = 92.48%; range = 84.04–100.0; Fig. 2) but it is reduced when compared to the HPD overlap obtained with soft character fossilization. The mean PDMA was −0.36 myr. The distribution of PDMA shows a slight propensity for the underestimation of node ages from treated data, but in a small number of cases age estimates were greatly overestimated. The distribution of per-node PDMA on the tree topology shows that certain nodes were often underestimated and that these nodes were clustered together (Fig. 4).

When the scaling factor was doubled, the average percentage of missing data in each simulated matrix increased to 29.62%, with 94.90% the mean quantity for fossil taxa. The mean HPD overlap was further reduced (mean HPD per-node overlap = 85.89%, range = 67.14–97.72; Fig. 2) and the PDMA increased to 1.43 myr.

Reduced matrices. For one of the reduced matrix analyses the untreated matrix produced a consensus tree containing a polytomy, for this analysis all replicate simulated results had to be discarded. With a scaling factor of 1, over 5 separate subsamples of the full matrix, the mean PDMA ranged from 0.32 to 2.47 myr. The mean HPD overlap per node ranged from 87.84% to 93.52% (Fig. 2), with a maximum value of 100% and a minimum of 68.64% across all 5 subsamples. When the scaling factor was doubled, the mean PDMA ranged from 0.33 to 1.09 myr, and the mean HPD overlap ranged from 88.20% to 89.51%, with a maximum value of 100% and a minimum of 31.15% across all four replicate subsamples (Fig. 2).

DISCUSSION

Complex taphonomic and biostratinomic processes control the distribution of missing fossil morphological data. Our results show the relative influence of two key stages in these processes on the accuracy of clade age estimation: (1) the loss of entire characters due to decay and

TABLE 1. Aggregate results obtained with different approaches to simulating fossil data.

Simulation strategy	Mean percentage error (absolute error) of median age estimates (%)	Mean paired difference of median age estimates (PDMA) (myr)	Mean standard deviation of paired difference of median age estimates (PDMA)	Two-tailed paired *t*-test p-value (H_0 = PDMA has a mean of 0)	Mean missing data across all fossil taxa (%)	Mean HPD overlap across nodes and replicates (%)
Soft character loss	−0.22 (1.04)	−0.13	0.79	1.97×10^{-12}	28.99	97.57
Soft character loss – R 1	0.22 (2.39)	0.11	1.65	4.95×10^{-3}	29.34	96.21
Soft character loss – R 2	0.02 (1.86)	0.01	1.52	7.25×10^{-1}	29.23	96.66
Soft character loss – R 3	0.65 (2.28)	0.24	1.58	9.59×10^{-11}	28.28	96.46
Soft character loss – R 4	0.44 (1.88)	0.27	1.63	$1.40E \times 10^{-12}$	28.84	96.51
Soft character loss – R 5	0.41 (2.00)	0.21	1.46	6.30×10^{-9}	28.95	96.41
Biostratinomy – sc = 1.0	−0.64 (2.66)	−0.36	2.33	9.74×10^{-12}	32.04	92.02
Biostratinomy – sc = 2.0	2.53 (5.94)	1.43	4.19	1.75×10^{-48}	94.90	84.72
Biostratinomy – sc = 1.0 R 1	1.31 (4.89)	0.69	3.63	4.74×10^{-16}	75.85	91.94
Biostratinomy – sc = 1.0 R 2	0.96 (4.62)	0.32	3.74	1.66×10^{-4}	73.62	92.04
Biostratinomy – sc = 1.0 R 3	1.37 (5.37)	0.72	3.75	1.08×10^{-16}	76.08	90.73
Biostratinomy – sc = 1.0 R 4	4.29 (7.27)	2.47	4.90	3.16×10^{-23}	76.36	87.13
Biostratinomy – sc = 1.0 R 5	0.71 (5.65)	0.43	4.22	5.68×10^{-6}	75.54	91.48
Biostratinomy – sc = 2.0 R 1	2.14 (5.71)	1.09	4.42	7.55×10^{-27}	94.58	88.33
Biostratinomy – sc = 2.0 R 2	Comparison not possible					
Biostratinomy – sc = 2.0 R 3	0.66 (6.21)	0.33	5.11	7.45×10^{-4}	94.65	88.77
Biostratinomy – sc = 2.0 R 4	0.50 (6.09)	0.43	5.49	7.21×10^{-4}	94.86	87.33
Biostratinomy – sc = 2.0 R 5	0.87 (5.15)	0.57	4.40	1.53×10^{-8}	94.73	88.84

sc, scaling factor; R, reduction.

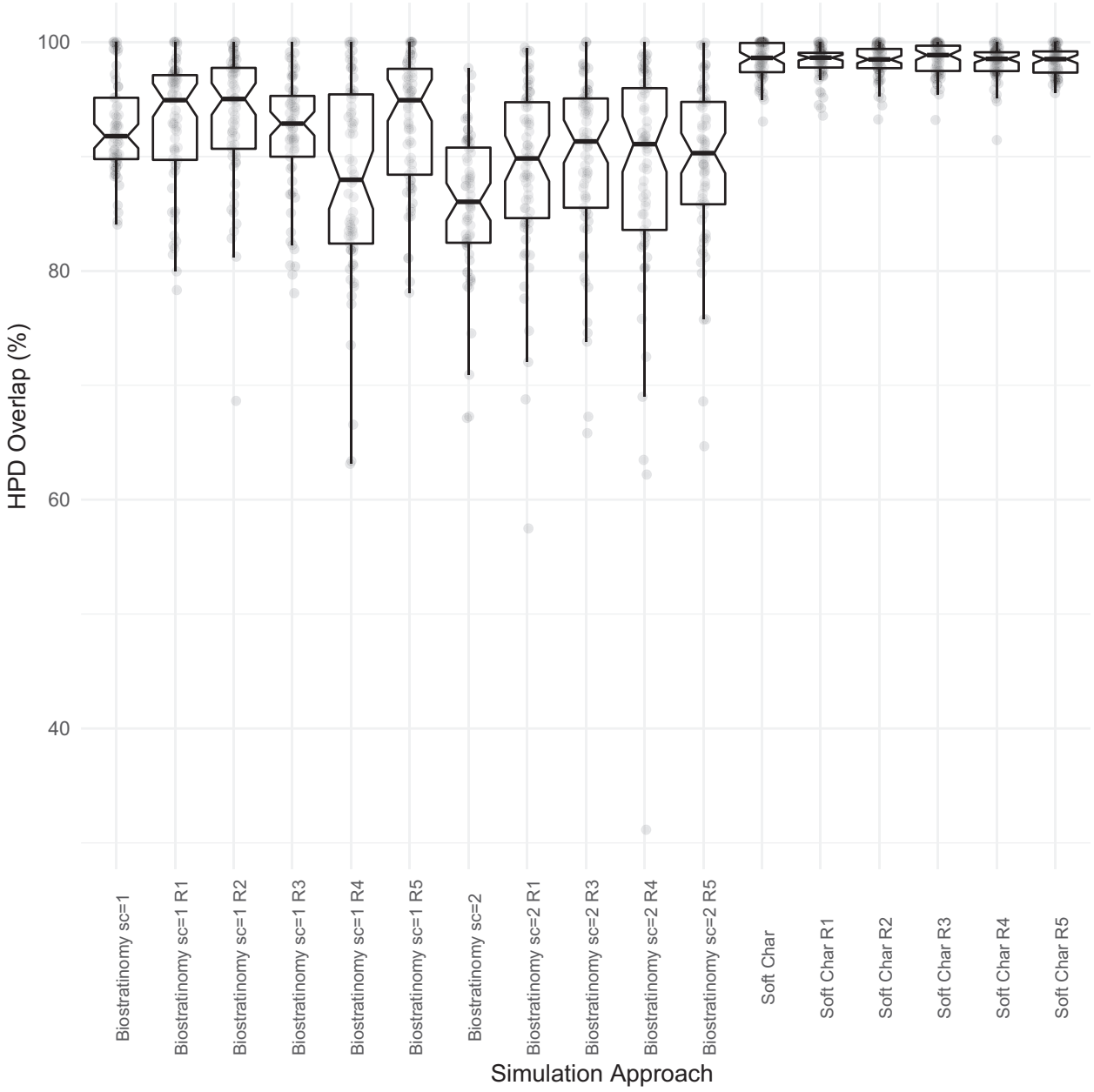

FIG. 2. Boxplots of overlap between 95% HPDs obtained with untreated data and different data treated with different distributions of missing data. Individual node age 95% HPD overlaps are represented by each point, with box plots covering the distribution for each taphonomic bias treatment approach. *Abbreviations*: sc, scaling factor; R, reduction.

degradation; and (2) the combined effect of physical biostratinomic processes such as disarticulation and transport.

Impact of data loss resulting from the loss of entire characters across all fossil taxa

Despite paired differences between median age estimates being significantly different from zero in a number of tests, it is evident from our results that the effect of the loss of entire characters for fossil taxa on divergence time estimation is small in absolute terms. Both percentage error and the magnitude of paired differences between median age estimates obtained with untreated and treated data sets are relatively small. Similarly, overlap between 95% HPDs obtained with untreated and treated data sets is extensive (Fig. 2). This level of performance is also seen in datasets of reduced size, with small paired differences between median age estimates obtained with and without treatment. This resilience to missing data is

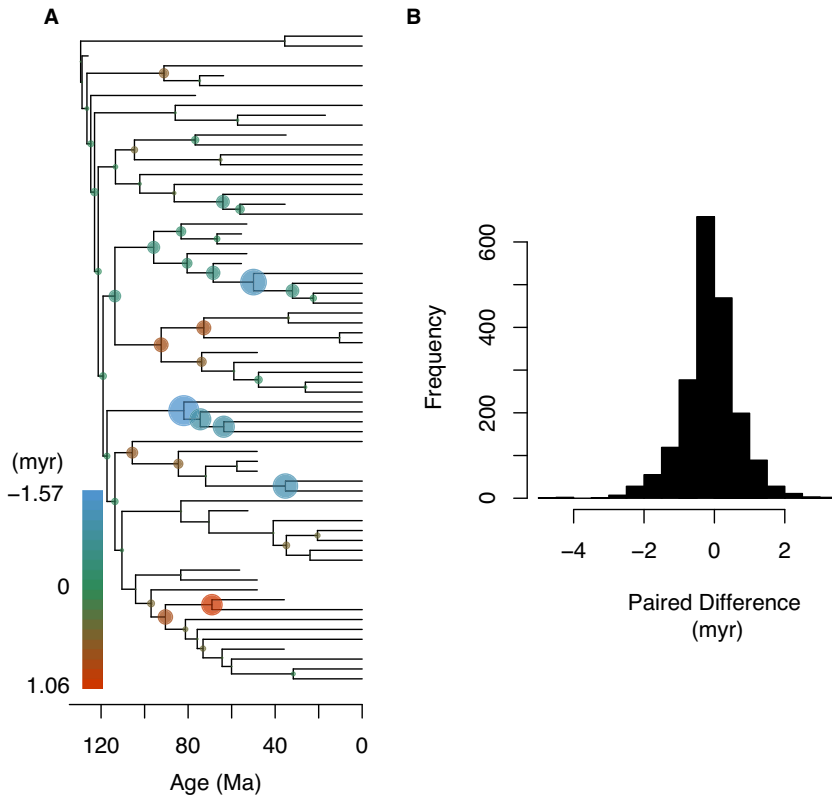

FIG. 3. A, mammalian time scaled phylogeny with node colour indicating the difference between median estimated ages obtained from the full 2454 character untreated matrix and median ages obtained from 30 replicate treated matrices when the distribution of missing data approximates that which is expected when soft characters are lost due to degradation and decay; the size of the node label is proportional to the magnitude of the difference associated with that node age estimate. B, histogram of node age estimate difference (PDMA) across all 30 replicate analyses.

doubtless a consequence of Bayesian analysis down-weighting characters with large amounts of missing data (King 2019). However, it must also be impacted by the inclusion of extant taxa that remain scored for these characters, influencing the interpretation of missing data in fossil taxa in the posterior distribution. The ability of observed data to influence the interpretation of unobserved data is obviously reliant on the quality, distribution, and proportion of observed data relative to missing data. The topological distribution of fossil taxa in the analysed matrix is fairly even, allowing interspersed extant taxa to inform parameter estimates across the tree. Therefore, a topologically uneven distribution of fossil taxa is likely to amplify the deleterious effects of data missing from all fossil taxa, such as soft character data (Guillerme & Cooper 2016*b*). Tip-calibrated analyses that exclude extant taxa are likely to be impacted most severely.

Impact of data loss resulting from post-decay biostratinomic processes

Missing data resulting from physical biostratinomic processes introduce a greater effect on age estimates than missing data distributed amongst soft characters. Distributions of 95% HPD overlap show that node age estimates obtained in the presence of missing data distributed according to biostratinomic processes are often markedly different to ages obtained with untreated data (Fig. 2). For all analyses in which the distribution of missing data is constrained by biostratinomic processes, the absolute value of the mean PDMA is greater than any analysis in which missing data is constrained to (proxy) soft tissue characters (Table 1). This further suggests that the effect of biostratinomic processes on age estimates is far greater than for soft tissue character loss. Despite a relatively small mean PDMA, when the scaling factor is 1, the standard deviation of the PDMA shows that there is strong variation in the differences of paired median age estimate differences between matrices that are untreated and those that simulate biostratinomic loss. This increase in PDMA was not caused by differing quantities of missing data in matrices exhibiting soft tissue character or biostratinomic loss, since they both contained ~10% missing data in total. Therefore, the increase in PDMA is likely to be caused by the distribution of missing data. Increasing the scaling factor further increases PDMA. The magnitude and direction of these paired median differences suggests that biostratinomic loss generally results in an underestimation of node age. The distribution of PDMA obtained with data simulated with a scaling factor of 1 shows that the mean PDMA was negative. However, this

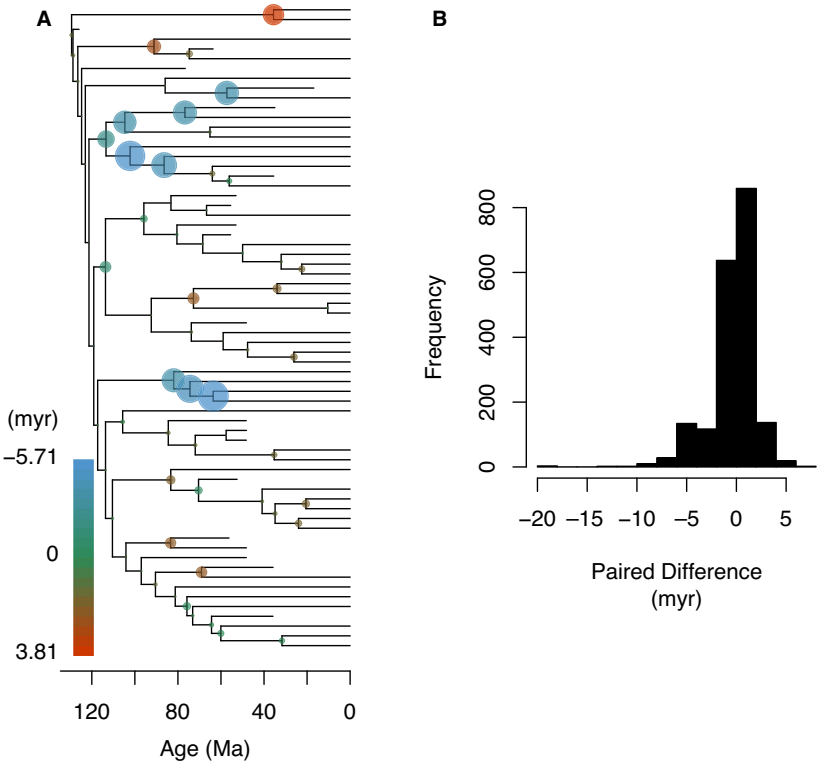

FIG. 4. A, mammalian time scaled phylogeny with node colour indicating the difference between median estimated ages obtained from the full 2454 character untreated matrix and average ages obtained from 30 replicate treated matrices when the distribution of missing data approximates that which is expected when characters are lost in blocks due to physical biostratinomic processes (scale factor = 1); the size of the node label is proportional to the magnitude of the difference associated with that node age estimate. B, histogram of node age difference (PDMA) across all 30 replicate analyses.

value may be misleading since a small number of severely overestimated ages mask the presence of many moderately underestimated node ages. When smaller datasets exhibit distributions of missing data constrained by biostratinomic processes, the mean PDMA is consistently positive and greater in magnitude than the PDMA obtained for matrices exhibiting soft character loss, further suggesting that, on average, biostratinomic processes result in an underestimation of node age. A potential cause of the increased difference in PDMA and reduction of HPD overlap is the non-independence of morphological characters. While relationships between characters are not explicitly accounted for in this Bayesian framework, the loss of an entire suite of characters may still exert a strong influence on the accuracy of divergence time estimates.

The role of extant taxa in mitigating against the effects of missing data

The amount of missing data seems to exert little influence on clade age estimates. For example, when missing data is distributed according to the effects of biostratinomic processes, with a scale factor of 1 there is ~32% missing data for fossil taxa in the complete matrix, but ~75% missing data for fossil taxa in the reduced matrix. Despite this marked difference in the quantity of missing fossil data, the distributions of 95% HPD overlap are comparable (Fig. 2) and the PDMA values for these analyses are relatively small. This suggests that the distribution of missing data within fossil taxa is routinely compensated for by the morphological data present for the extant taxa, as demonstrated in previous simulation-based studies (Guillerme & Cooper 2016*a*). Furthermore, when the widths of the 95% HPD intervals are considered, it is clear that age estimate precision remains approximately consistent across all analyses, irrespective of the quantity of missing fossil morphological data (Fig. 5). This further demonstrates the importance of the presence of including morphological data for extant taxa in these analyses. The tip-calibration framework now facilitates divergence time analyses consisting entirely of fossil taxa (e.g. Cau 2017); for such analyses the beneficial effects of well scored extant taxa will be missing and, in such cases, the negative influence of missing morphological data in fossil taxa is likely to be exacerbated. Our simulation framework was based on a fixed topology, reflecting a situation in which there is little uncertainty regarding the relationships between extant and fossil taxa. Tip-calibration is often applied in a co-estimation framework, in which uncertainty regarding topology is accounted for through joint estimation with divergence times. The effect of taphonomic bias on divergence times in the joint estimation framework requires further investigation. This will be

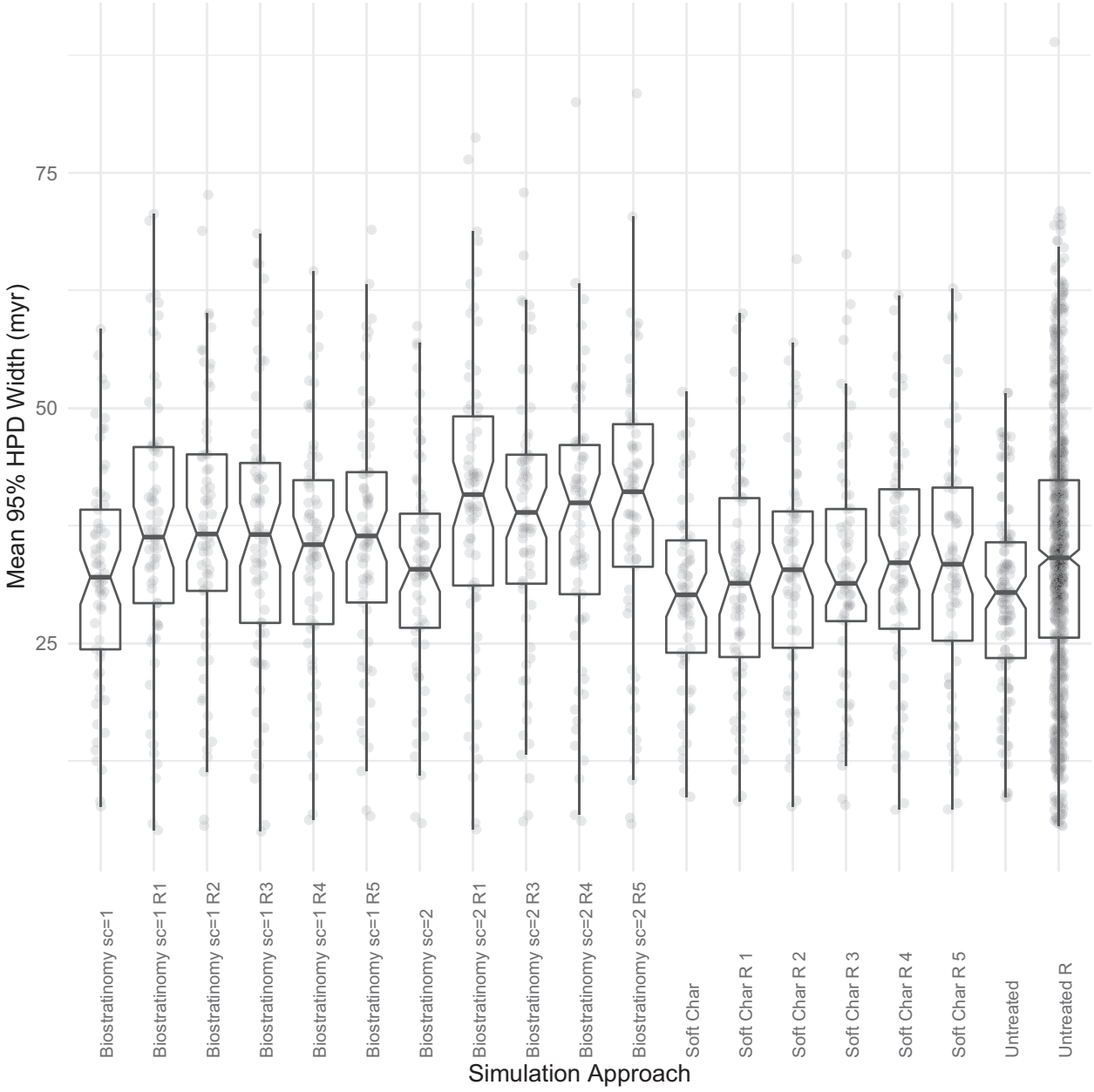

FIG. 5. Boxplots of mean width of 95% HPDs obtained with data treated with different distributions of missing data. Individual node age 95% widths averaged over 30 replicates are represented by each point, with box plots covering the distribution for each taphonomic bias treatment approach. *Abbreviations*: sc, scaling factor; R, reduction.

particularly important for situations in which the posterior relationships between fossil and extant taxa remain uncertain, as the mitigating effect of data from extant taxa, as demonstrated here, may be different in such cases.

The negative effects of taphonomic processes on matrix composition could potentially be mitigated by modifying the phylogenetic model to explicitly account for taphonomic processes, or by altering the qualities of the data exposed to the model. The development of such a model would be decidedly non-trivial due to the disparity of fossilization pathways that vary with intrinsic biology and the extrinsic environment of fossilization. However, our results suggest that the simplest solution may be to subsample datasets to minimize the number of characters that are coded only for a small subset of fossil taxa. The positive influence of this approach on the accuracy of divergence time estimation is effectively demonstrated in our simulations of the effects of soft tissue decay, analogous to an

empirical dataset subsampled to minimize physical biostratinomic effects. Theoretically, eliminating characters could strongly limit the statistical power of the remaining phenotypic data, but phylogenetic characters with large amounts of missing data have low power in contemporary Bayesian phylogenetic inference (King 2019). Thus, in practice, there may be nothing to lose and much to gain from the exclusion of missing data.

CONCLUSION

Missing data in morphological matrices are distributed systematically, with certain character types more likely to be missing than others in fossil taxa. Using a large palaeontological dataset and empirically derived distributions of simulated missing data, we have demonstrated the relative influence of missing morphological data distributed according to different biostratinomic and taphonomic processes. The decay and loss of entire characters appears to introduce little error into age estimates, whereas the loss of characters due to physical biostratinomic processes is likely to have a more significant impact on estimated clade ages. Mitigation against the effects of biostratinomic processes may be achieved by subsampling matrices to minimize the quantity of missing data introduced by these processes. Despite this, the magnitude of differences in age estimates obtained before and after the simulation of all taphonomic processes are generally small, and subsampling may be unnecessary when an abundance of data from extant taxa are available.

Acknowledgements. We thank Robert Sansom (Manchester) and our colleagues in the Bristol Palaeobiology Research Group, for discussion; Thomas Guillerme (Sheffield) and an anonymous referee provided constructive critical review that helped to improve our manuscript. This work was supported by BBSRC (BB/N000919/1 and BB/T012773/1 to PCJD) and NERC (NE/L501554/1 to JEO'R; NE/P013678/1 to PCJD) including funding from the Biosphere Evolution, Transitions and Resilience (BETR) programme, which is co-funded by the Natural Science Foundation of China (NSFC).

DATA ARCHIVING STATEMENT

Data available from the Dryad Digital Repository: https://doi.org/10.5061/dryad.b2rbnzsd4

Editor. Imran Rahman

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article:

Figure S1. The percentage of missing data in the 286 morphological matrices analysed by Guillerme & Cooper (2016*b*).

REFERENCES

ASLAN, A. and BEHRENSMEYER, A. K. 1996. Taphonomy and time resolution of bone assemblages in a contemporary fluvial system: the East Fork River, Wyoming. *Palaios*, **11**, 411–421.

BECK, R. M. D. and LEE, M. S. Y. 2014. Ancient dates or accelerated rates? Morphological clocks and the antiquity of placental mammals. *Proceedings of the Royal Society B*, **281**, 10.

BEHRENSMEYER, A. K. 1990. Bones. 232–235. *In* BRIGGS, D. E. G. and CROWTHER, P. R. (eds). *Palaeobiology: A synthesis*. Blackwell Scientific Publications.

—— and KIDWELL, S. M. 1985. Taphonomy contributions to paleobiology. *Paleobiology*, **11**, 105–119.

BENTON, M. J., DONOGHUE, P. C. J., ASHER, R. J., FRIEDMAN, M., NEAR, T. J. and VINTHER, J. 2015. Constraints on the timescale of animal evolutionary history. *Palaeontologia Electronica*, **18**, 107.

CAU, A. 2017. Specimen-level phylogenetics in paleontology using the Fossilized Birth-Death model with sampled ancestors. *PeerJ*, **5**, e3055.

DONOGHUE, P. C. J. and BENTON, M. J. 2007. Rocks and clocks: calibrating the Tree of Life using fossils and molecules. *Trends in Ecology & Evolution*, **22**, 424–431.

—— and YANG, Z. H. 2016. The evolution of methods for establishing evolutionary timescales. *Philosophical Transactions of the Royal Society*, **371**, 20160020.

DRUMMOND, A. J., HO, S. Y., PHILLIPS, M. J. and RAMBAUT, A. 2006. Relaxed phylogenetics and dating with confidence. *PLoS Biology*, **4**, e88.

GAVRYUSHKINA, A., WELCH, D., STADLER, T. and DRUMMOND, A. J. 2014. Bayesian inference of sampled ancestor trees for epidemiology and fossil calibration. *PLoS Computational Biology*, **10**, e1003919.

GUILLERME, T. and COOPER, N. 2016*a*. Effects of missing data on topological inference using a total evidence approach. *Molecular Phylogenetics & Evolution*, **94**, 146–158.

—— —— 2016*b*. Assessment of available anatomical characters for linking living mammals to fossil taxa in phylogenetic analyses. *Biology Letters*, **12**, 20151003.

HEATH, T. A., HUELSENBECK, J. P. and STADLER, T. 2014. The fossilized birth-death process for coherent calibration of divergence-time estimates. *Proceedings of the National Academy of Sciences*, **111**, E2957–E2966.

KING, B. 2019. Which morphological characters are influential in a Bayesian phylogenetic analysis? Examples from the earliest osteichthyans. *Biology Letters*, **15**, 20190288.

LEPAGE, T., BRYANT, D., PHILIPPE, H. and LARTILLOT, N. 2007. A general comparison of relaxed molecular clock models. *Molecular Biology & Evolution*, **24**, 2669–2680.

LEWIS, P. O. 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. *Systematic Biology*, **50**, 913–925.

O'LEARY, M. A., BLOCH, J. I., FLYNN, J. J., GAUDIN, T. J., GIALLOMBARDO, A., GIANNINI, N. P., GOLDBERG, S. L., KRAATZ, B. P., LUO, Z. X., MENG, J., NI,

X. J., NOVACEK, M. J., PERINI, F. A., RANDALL, Z. S., ROUGIER, G. W., SARGIS, E. J., SILCOX, M. T., SIMMONS, N. B., SPAULDING, M., VELAZCO, P. M., WEKSLER, M., WIBLE, J. R. and CIRRANELLO, A. L. 2013. The placental mammal ancestor and the post–K-Pg radiation of placentals. *Science*, **339**, 662–667.

O'REILLY, J. E. and DONOGHUE, P. C. J. 2021. Fossilization processes have little impact on tip-calibrated divergence time analyses. *Dryad Digital Repository*. https://doi.org/10.5061/dryad.b2rbnzsd4

—— DOS REIS, M. and DONOGHUE, P. C. J. 2015. Dating tips for divergence-time estimation. *Trends in Genetics*, **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., 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. and BENTON, M. J. 2012. Best practices for justifying fossil calibrations. *Systematic Biology*, **61**, 346–359.

PYRON, R. A. 2011. Divergence time estimation using fossils as terminal taxa and the origins of Lissamphibia. *Systematic Biology*, **60**, 466–481.

RAMBAUT, A., SUCHARD, M., XIE, D. and DRUMMOND, A. 2014. Tracer v1.6. http://tree.bio.ed.ac.uk/software/tracer

RONQUIST, F., KLOPFSTEIN, S., VILHELMSEN, L., SCHULMEISTER, S., MURRAY, D. L. and RASNITSYN, A. P. 2012*a*. A total-evidence approach to dating with fossils, applied to the early radiation of the hymenoptera. *Systematic Biology*, **61**, 973–999.

—— TESLENKO, M., VAN DER MARK, P., AYRES, D. L., DARLING, A., HÖHNA, S., LARGET, B., LIU, L., SUCHARD, M. A. and HUELSENBECK, J. P. 2012*b*. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, **61**, 539–542.

SANSOM, R. S. and WILLS, M. A. 2013. Fossilization causes organisms to appear erroneously primitive by distorting evolutionary trees. *Scientific Reports*, **3**, 5.

THORNE, J. L., KISHINO, H. and PAINTER, I. S. 1998. Estimating the rate of evolution of the rate of molecular evolution. *Molecular Biology & Evolution*, **15**, 1647–1657.

WARNOCK, R. C. M., YANG, Z. and DONOGHUE, P. C. J. 2017. Testing the molecular clock using mechanistic models of fossil preservation and molecular evolution. *Proceedings of the Royal Society B*, **284**, 20170227.

YANG, Z. and RANNALA, B. 2006. Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. *Molecular Biology & Evolution*, **23**, 212–226.

ZUCKERKANDL, E. and PAULING, L. 1965. Molecules as documents of evolutionary history. *Journal of Theoretical Biology*, **8**, 357–366.