Palaeontology
The Palaeontological Association
www.palass.org

[Palaeontology, 2023, e12680]

Inaccurate fossil placement does not compromise tip-dated divergence times

by NICOLÁS MONGIARDINO KOCH[1,*], RUSSELL J. GARWOOD[2,3] *and* LUKE A. PARRY[4,*]

[1]Scripps Institution of Oceanography, University of California San Diego, San Diego, CA, USA; nmongiardinokoch@ucsd.edu
[2]Department of Earth and Environmental Sciences, University of Manchester, Manchester, UK; russell.garwood@manchester.ac.uk
[3]Natural History Museum, London, UK; russell.garwood@manchester.ac.uk
[4]Department of Earth Sciences, University of Oxford, Oxford, UK; luke.parry@earth.ox.ac.uk
*Corresponding author

Typescript received 30 September 2022; accepted in revised form 15 August 2023

Abstract: Time-scaled phylogenies underpin the interrogation of evolutionary processes across deep timescales, as well as attempts to link these to Earth's history. By inferring the placement of fossils and using their ages as temporal constraints, tip dating under the fossilized birth–death (FBD) process provides a coherent prior on divergence times. At the same time, it also links topological and temporal accuracy, as incorrectly placed fossil terminals should misinform divergence times. This could pose serious issues for obtaining accurate node ages, yet the interaction between topological and temporal error has not been thoroughly explored. We simulate phylogenies and associated morphological datasets using methodologies that incorporate evolution under selection, and are benchmarked against empirical datasets. We find that datasets of 300 characters and realistic levels of missing data generally succeed in inferring the correct placement of fossils on a constrained extant backbone topology, and that true node ages are usually contained within Bayesian posterior distributions. While increased fossil sampling improves the accuracy of inferred ages, topological and temporal errors do not seem to be linked: analyses in which fossils resolve less accurately do not exhibit elevated errors in node age estimates. At the same time, inferred divergence times are biased, probably due to a mismatch between the FBD prior and the shape of our simulated trees. While these results are encouraging, suggesting that even fossils with uncertain affinities can provide useful temporal information, they also emphasize that palaeontological information cannot overturn discrepancies between model priors and the true diversification history.

Key words: tip dating, time calibration, morphological clock, fossilized birth–death model.

ESTABLISHING an accurate timeline for the diversification of life on Earth is integral to understanding major events in evolutionary history and linking these to the history of the planet. Phylogenies scaled to absolute time underpin our understanding of the interplay between biology and geology (Mills *et al.* 2022; Palazessi *et al.* 2022), the interactions between lineages across deep time (Strassert *et al.* 2021; Asar *et al.* 2022; Onstein *et al.* 2022), the movement of clades across the globe (O'Hara *et al.* 2019; Landis *et al.* 2022), and the tempo and mode by which ecological, molecular and morphological novelties have emerged (Jabłońska & Tawfik 2021; Coombs *et al.* 2022; Mongiardino Koch *et al.* 2022a; Suissa & Friedman 2022). Two distinct sources of information are used to infer divergence times: character datasets for the species under study, which reflect their relative degrees of divergence; and temporal constraints that help translate these into absolute times (Donoghue & Benton 2007). Combined with various modelling assumptions regarding patterns of character evolution, rate variation and lineage diversification (Warnock & Wright 2021), these data produce time-calibrated phylogenetic trees (i.e. chronograms or timetrees) whose node ages have been estimated using a molecular and/or morphological clock.

While various sources of temporal evidence can be used to calibrate a phylogenetic clock model, fossil occurrences are the most widely employed (Hipsley & Müller 2014). When applied to the nodes of a phylogenetic tree, fossil calibration points provide minimum bounds for divergence times, as the earliest occurrence of a lineage in the fossil record necessitates older origination times. The minimum ages of multiple nodes can be bracketed in such a way, providing local temporal information that can be combined with relaxed clock models that accommodate heterogeneities in evolutionary rates among branches (Kishino *et al.* 2001; Drummond

doi: 10.1111/pala.12680

et al. 2006). However, biological and geological processes can impair the degree to which fossil occurrences approximate the ages of cladogenetic events (Marshall 2019), as well as the overall accuracy of fossil-calibrated clocks (e.g. Budd & Mann 2020; Carruthers & Scotland 2020). While some of the limitations of node-based approaches can be overcome by using fossils to derive prior distributions for node ages (Yang & Rannala 2006), the (semi)arbitrary configuration of these calibration densities can also largely determine the results obtained (Ho & Phillips 2009; Warnock *et al.* 2011; Brown & Smith 2018; Moody *et al.* 2022). Furthermore, results can also be compromised by incorrect assumptions regarding the relationships between fossils and extant terminals (Near *et al.* 2005; Lee 2009; Parham *et al.* 2012), which are treated with absolute certainty in node dating analyses.

An alternative approach to extracting the temporal signal from the fossil record is to incorporate extinct taxa as terminals whose phylogenetic placement is inferred using morphological information (Ronquist *et al.* 2012a; O'Reilly *et al.* 2015; Wright *et al.* 2022). While methods to calibrate trees using fossil tips have existed for several decades (Fisher 1992; Wagner 1998), modern Bayesian implementations (often referred to as 'tip dating') have become dominant in recent years. Early tip-dating efforts (Pyron 2011; Ronquist *et al.* 2012a) made use of simplistic models of diversification (also known as tree priors), but subsequent studies have applied fossilized birth–death (FBD) processes (Heath *et al.* 2014) that explicitly model speciation, extinction and fossil recovery dynamics. The switch to these mechanistic models of diversification has generally resulted in improved estimates of divergence times (Matzke & Wright 2016; Zhang *et al.* 2016; Arcila & Tyler 2017; May *et al.* 2021). Simulations have also revealed that tip dating outperforms undated Bayesian and parsimony approaches in terms of topological accuracy (Mongiardino Koch *et al.* 2021), while being relatively robust to several potential sources of error (Klopfstein *et al.* 2019; Luo *et al.* 2020; O'Reilly & Donoghue 2020).

Calibrating phylogenies using fossil terminals provides numerous benefits in addition to improvements in the accuracy of tree topology and divergence times. First, it does not require directly specifying fossil calibrations as prior distributions on node ages (except for the root node), avoiding a number of subjective (yet often impactful) decisions. Furthermore, the implementation of FBD processes provides a means of encompassing living and extinct terminals within a unified model of diversification and sampling, allowing the recovery of ancestor–descendant relationships that are not accommodated by other methods (Gavryushkina *et al.* 2014), and permitting the straightforward integration of morphological, stratigraphic and molecular datasets for total-evidence dated inference (Zhang *et al.* 2016; Heath *et al.* 2017). The resulting topologies not only provide an adequate basis for macroevolutionary studies (Slater & Harmon 2013; Arcila & Tyler 2017; Mongiardino Koch 2021), but macroevolutionary dynamics can even be incorporated into the process of phylogenetic reconstruction (Simões *et al.* 2020; May *et al.* 2021; Wright *et al.* 2021). Finally, tip-dated inference makes a much more thorough and objective use of palaeontological data by potentially allowing all fossil species to be incorporated, not just the oldest representatives of clades; and inferring, rather than enforcing, their phylogenetic positions (O'Reilly *et al.* 2015; Wright *et al.* 2022). Nevertheless, there are a number of types of temporal constraint that cannot be incorporated as tips in phylogenetic analyses, such as trace fossils (e.g. tetrapod trackways; Simões & Pierce 2021), relative constraints like horizontal gene transfer events (Wolfe & Fournier 2018) or interpretations of absence in the fossil record that are often used to construct constraints on the root, such as the absence of particular biomineralizing animal clades in the Precambrian (Vinther *et al.* 2012).

In summary, the simultaneous estimation of tree topology and divergence times in tip-dated analyses allows for a better integration of sources of uncertainty through the interaction of morphological and stratigraphic signals (Ronquist *et al.* 2012a; King *et al.* 2017; King 2021). Nonetheless, multiple challenges of tip dating remain to be thoroughly assessed, including the exploration of avenues for incorporating heterogenous temporal dynamics, the formulation of adequate models of morphological change and taxon sampling schemes, and the exploration of how these factors affect model parameters, node ages, and tree topology (e.g. Matschiner 2019; Barido-Sottani *et al.* 2020; Luo *et al.* 2020; O'Reilly & Donoghue 2020; Simões *et al.* 2020; May *et al.* 2021; Spasojevic *et al.* 2021; Barido-Sottani *et al.* 2023; Wright *et al.* 2022; Luo *et al.* 2023; May & Rothfels 2023; Simões *et al.* 2023). Another unique aspect of tip dating that has received little attention is the relationship between topological and temporal accuracy. Given that fossil terminals directly constrain divergence times only in the path that connects them to the root of the tree, inaccurate fossil placements should misinform node ages in a manner that depends on the magnitude of topological error (Barido-Sottani *et al.* 2023). This is problematic, as morphological phylogenetics suffers from many shortcomings, including the subjectivity associated with character selection and scoring, the limited size of datasets, the overall difficulty in modelling the process and rate of morphological evolution, and the high prevalence of selective processes that overprint historical signals (O'Reilly *et al.* 2015; dos Reis *et al.* 2015; Donoghue & Yang 2016; Simões *et al.* 2017; Keating *et al.* 2020; Mongiardino Koch & Parry 2020). Furthermore, fossils are affected by high levels of missing data, which can affect tree topology by reducing the precision with which they can be placed (Wiens & Moen 2008; Aberer *et al.* 2013; Guillerme & Cooper 2016; Mongiardino Koch *et al.* 2021), and sometimes even biasing their inferred phylogenetic position (Sansom *et al.* 2010; Sansom & Wills 2013). Several studies, both empirical (Turner *et al.* 2017; Spasojevic *et al.* 2021; Mongiardino Koch & Thompson 2021) and simulation-based (Luo *et al.* 2020; Mongiardino Koch *et al.* 2021) have revealed the difficulties associated with accurately placing terminals devoid of molecular data (such as fossils). Some have even suggested that this impacts the accuracy of inferred divergence times to such an extent that it potentially negates the other benefits of tip dating (O'Reilly *et al.* 2015; Donoghue & Yang 2016). To date, the nature and strength of the relationship between the accuracy of fossil placement and inferred divergence times in tip-dated analyses under morphological clocks has not been characterized. We do so here, using simulations to test the hypothesis that more accurate fossil placement improves estimates of divergence times.

METHOD

Simulation approach

Our simulation approach follows the procedures described in Mongiardino Koch *et al.* (2021). Evolutionary histories, encompassing both morphological datasets and phylogenetic trees, were generated using the individual-based simulation framework TREvoSim v2.0.0 (Keating *et al.* 2020; Garwood 2021). This approach differs in a number of ways from other commonly-used simulation methodologies, namely: it does not employ any of the models routinely used for phylogenetic inference (i.e. neither Markov substitution models nor birth–death tree models); it outputs paired character matrices and time-scaled phylogenetic trees that are concurrently generated by the simulation process, rather than producing these separately; and it incorporates important biological concepts such as an estimate of fitness and a species concept (Garwood *et al.* 2019; Keating *et al.* 2020). This introduces a level of model misspecification that is probably pervasive in morphological phylogenetics under probabilistic methodologies (Wright & Hillis 2014), as well as better resembling empirical morphological datasets that have been shaped by adaptive evolution (Zou & Zhang 2016). Briefly, TREvoSim simulations incorporate a set number of organisms comprising binary strings of a fixed size, which in this case, represent their phenotype. Organisms compete against each other within one of several playing fields, each of which is a complex environment composed of multiple sets of randomly-generated binary strings (each set being akin to a discrete ecological niche). A fitness is assigned to each organism based on a comparison between phenotypic and ecological strings, and used to determine its reproductive success. As lineages reproduce, random mutations are incorporated into their phenotypes, leading to increased divergence that ultimately results in speciation events. Although mutation rates are held constant through time, constancy in the rates of phenotypic change is not enforced. Ecological strings also mutate, simulating environmental change. As the simulation progresses, a phylogenetic tree representing the relationships and divergence times between all species is produced and stored, while phenotypes are recorded either upon extinction or when the simulation concludes (indicating that the present day has been reached). All aspects of the process (number of organisms and phenotypic characters, number of playing fields and environments, probability of both mutation and ecological change, phenotypic distance before speciation) are under control of the user (see Garwood *et al.* 2019; Keating *et al.* 2020; Mongiardino Koch *et al.* 2021; the software is also fully documented, see Garwood 2021).

Settings were identical to those used by Mongiardino Koch *et al.* (2021), as these were validated to replicate properties seen in 12 empirical morphological datasets that include fossil terminals, such as evolutionary rates, distributions of branch lengths, and levels of treeness (empirical datasets and a TREvoSim settings file can be found in the supplementary materials of Mongiardino Koch *et al.* 2021). Topologies incorporated a series of short internodes close to the root and followed by long unsplit branches giving rise to major extant clades. This pattern is reminiscent of ancient rapid radiations (Whitfield & Lockhart 2007), which pose numerous challenges for phylogenetic inference and dating. Settings were chosen to produce datasets containing 500 binary characters and 999 terminals, of which an average of 150 were extant.

Data manipulation

Subsequent processing steps were undertaken in the R programming environment v4.2.2 (R Core Team 2013), using scripts provided in Mongiardino Koch *et al.* 2022b, and relying on packages ape (Paradis & Schliep 2019), Claddis (v0.6.3; Lloyd 2016), dplyr (Wickham *et al.* 2021) and phytools (v1.5.1; Revell 2012). Phylogenetic trees were checked for zero-length branches, and datasets with non-fully bifurcating topologies were discarded. Given that the chronograms simulated by TREvoSim are scaled to the number of iterations, we translated these to absolute time by assuming a 210.9 myr-old root, given the average timespan of the 12 reference empirical datasets, and rescaled branch lengths accordingly. Datasets were then subsampled by: (1) eliminating fossils at random to produce a dataset size of 300 terminals, a process that mimics the incompleteness of the fossil record; and (2) selecting 300 parsimony-informative characters at random, mimicking the process of character selection involved in matrix construction. Simulated datasets were further modified by imputing missing data. We counted the number of characters that were missing/inapplicable for each taxon in the empirical datasets employed, and combined these to derive realistic distributions of the proportion of missing data in both living and extinct terminals (Fig. S1). Values were then drawn from these distributions, and data was deleted at random from the simulated characters accordingly, producing datasets that varied widely in their levels of missing data (Fig. S2).

In order to evaluate the effect of increasing numbers of fossils on the accuracy of inferred divergence times, each dataset was subsampled one last time to 100 final terminals, varying the proportion of fossil and extant tips. Seven levels of fossil sampling were explored (5, 9, 16, 28, 44, 61 and 75 terminals), which were chosen as they represent approximately doubling numbers of fossil tips relative to internal nodes connecting extant taxa (0.05, 0.10, 0.19, 0.39, 0.80, 1.61 and 3.12, respectively). We generated 250 datasets for every level of fossil sampling (1750 datasets overall), using each TREvoSim simulation only once.

Phylogenetic inference

All phylogenetic analyses were performed in MrBayes v3.2.7 (Ronquist *et al.* 2012b). Given the difficulties of assessing the accuracy of divergence times when topology is being co-estimated, as nodes might only exist in a fraction of posterior trees, we used partial constraints to enforce a topological scaffold connecting all extant terminals. This relied on functions from paleotree (v3.44; Bapst 2012), and enforced the true (i.e. simulated) relationships among extant terminals, while leaving the position of fossils unconstrained. This approach has also been advocated by many to improve topological and/or temporal accuracy of tip-dated inference, and reflects the fact that the overall topological resolution of morphological matrices is generally lower than that of genome-scale datasets (Ronquist *et al.* 2012a; Slater 2013; Beck & Lee 2014; Crawford *et al.* 2015; Lee & Palci 2015; O'Reilly *et al.* 2015; Donoghue & Yang 2016; O'Reilly & Donoghue 2016; Brown & Smith 2018; Lee & Yates 2018; Mongiardino Koch & Thompson 2021; Darlim *et al.* 2022). We note, however, that despite being well resolved, phylogenomic trees inevitably often contain incorrect nodes, especially given ongoing controversies in the placement of particular clades where different dataset construction approaches, modelling choices and taxon sampling result in conflicting but maximally supported phylogenies (Ballesteros & Sharma 2019; Li *et al.* 2021; Mulhair *et al.* 2022; Mongiardino Koch *et al.* 2023). Our use of extensive extant scaffolding represents an ideal scenario, that could be problematic if applied in empirical studies, but allows us to assess the impact of fossil misplacement in isolation, rather than accounting for all of the sources of error in tip dating approaches to inferring evolutionary timescales.

Realistic levels of stratigraphic uncertainty were modelled using information from the Paleobiology Database (PBDB; https://paleobiodb.org/). The entire PBDB was downloaded on 27 March 2020 (using parameters: time interval = Cambrian through Quaternary) and used to build a distribution of species longevities. Occurrences were grouped by species, and longevities defined as the time spanned between the minimum age of their youngest occurrence and the maximum age of their oldest one. Longevities longer than 100 myr were filtered out, and the remaining data was found to approximate an exponential distribution with a rate parameter of 0.115 (estimated using 'fitdistr' in MASS; Venables & Ripley 2002), corresponding to a mean duration of 8.67 myr (Fig. S3). Time intervals were generated by sampling from this distribution and assigning to fossils in such a way that the true age (i.e. the absolute time upon extinction) was contained randomly within this interval. A minimum interval of 0.0117 myr (the shortest in the PBDB) was enforced, and ranges were checked not to cross the present. Temporal ranges for fossils were translated into uniform priors for their tip ages (Barido-Sottani *et al.* 2020). The age of extant terminals was fixed to the present.

The age of the root of the tree was treated as an unknown parameter, and priors for it were designed in a way that imitates the uncertainties of empirical phylogenetics. As routinely done, the earliest known fossil was used as a minimum age constraint. Note that this taxon is unlikely to be the true oldest member of the clade, as an average of 82% of extinct taxa were treated as unfossilized/undiscovered. If said taxon was not included in the final matrix (which only incorporated between 5 and 75 of the approximately 150 known fossil taxa, see above), then uncertainty in its age was generated as explained above and accounted for when establishing the minimum value for the root age prior. This parameter was modelled using both uniform and offset exponential distributions, allowing us to assess the effects of this decision. To find a justifiable width for these distributions, we scraped the online Fossil Calibration Database (https://fossilcalibrations.org/; Ksepka *et al.* 2015) using package rvest (Wickham 2021). We obtained 202 node calibrations with both minimum and maximum bounds, which showed an average difference of 107.7 myr. When implementing a uniform root age prior, values were set to span 107.7 myr immediately prior to the maximum value of the temporal range of the oldest known fossil. When implementing an offset exponential distribution, the mean value was set so that 95% prior probability was contained in the 107.7 myr preceding said value, thus establishing a less restrictive soft maximum (Yang & Rannala 2006; Warnock *et al.* 2011). These alternative root age priors are exemplified in Figure S4.

Diffuse priors were used for most parameters of the birth–death model (exponential with a rate parameter of 10 for the speciation probability, default flat beta distributions for both extinction and fossilization probabilities), as TREvoSim does not incorporate them, and their true values are therefore unknown. The probability of sampling extant taxa was set to the true value. Morphological evolution was modelled using an $Mk_{parsinf} + \Gamma$ model (Lewis 2001; Matzke & Irmis 2018), accommodating both among-character rate heterogeneity and the ascertainment bias introduced by sampling only parsimony-informative characters. An uncorrelated relaxed morphological clock (IGR) was implemented (Lepage *et al.* 2007), with a variance drawn from an exponential prior with a rate parameter of 10. A normal distribution (mean = 0.001, standard deviation = 0.01) was used for the base clock rate, and the species sampling strategy was set to 'fossiltip' to avoid the recovery of fossils as direct ancestors, given that TREvoSim outputs reflect only cladogenetic processes. Two independent runs of four chains were continued for either 50 million generations or until attaining an average standard deviation of split frequencies (ASDSF) of 0.005, which was taken to represent a thorough sampling of posterior topologies (Mongiardino Koch *et al.* 2021). Convergence of parameters was confirmed using potential scale reduction factors (PSRF; Brooks & Gelman 1997), and are reported in Table S1. Trees were sampled every 1000 generations, and a burn-in fraction of 0.5 was applied.

In order to assess the effect on node ages that is introduced by the models employed, we ran a subset of 350 analyses (50 from each level of fossil sampling, randomly selected) in the absence of morphological information. To this effect, we fixed the entire topology to match that of the simulated tree, obtaining dates that are informed exclusively by the ages of fossil tips and the model priors (Warnock *et al.* 2015; Brown & Smith 2018).

Data analysis

Posterior distributions of topologies were imported into R and analysed using packages already mentioned, as well as functions from castor (Louca & Doebeli 2017), ggplot2 (v3.4.1; Wickham 2016), phangorn (v2.8.1; Schliep 2011), readr (Wickham & Hester 2020) and stringr (Wickham 2019). Two aspects of these topologies were explored through comparison with the true (i.e. simulated) tree: the placement of fossil taxa relative to the extant subtree, henceforth topological accuracy; and the inferred ages of the nodes in the extant subtree, henceforth temporal accuracy. In order to ensure that uncertainty was being meaningfully captured, analyses with posterior distributions containing fewer than 100 trees were excluded. This occurred only under low levels of fossil sampling, as the extensive topological constraints meant low ASDSF values were reached rapidly. Results of analyses including 5 and 9 terminals are therefore not based on 250 datasets, but on 174 and 248, respectively.

Topological accuracy was explored by iterating through every fossil, and pruning true and posterior trees to just the extant taxa plus the focal fossil terminal. True and inferred fossil positions were compared in a random sample of 100 posterior trees using two approaches. First, we counted the fraction of posterior trees in which the fossil was placed correctly. Second, we calculated the average distance between true and inferred placements, estimated using the number of nearest-neighbour interchange (NNI) moves between the topologies. Note that some types of topological errors are not captured by these estimates, such as those relating to the relationships among entirely extinct clades. This was purposefully done in order to focus on metrics of topological accuracy that only capture the extent to which fossil tips are misplaced relative to the extant subtree, and therefore potentially misinforming node ages within it. The relative age of fossil terminals (i.e. their true age divided by the root age) and their proportion of missing data were gathered as potential predictors. Values for all fossil terminals were averaged resulting in a single value that represents the overall topological error of each analysis.

Temporal accuracy was estimated in random samples of 1000 posterior trees pruned to just the extant taxa. For every node, we recorded the difference between the true age and the median posterior estimate, as well as whether the true age was contained within the 95% highest-posterior density (HPD) interval (known as the 95% HPD coverage). Relative errors were further obtained by dividing age differences by the true age value. Depending on the analysis, errors were expressed as signed deviations, or transformed into absolute values. We considered median posterior estimates to be the expected value of inferred parameters, and thus refer to systematic differences between median estimates and true values as biases. As with topological accuracy, values for all nodes were averaged, resulting in a single value that characterizes the overall temporal accuracy of each analysis.

The relationships between different estimates of topological accuracy and temporal accuracy were explored

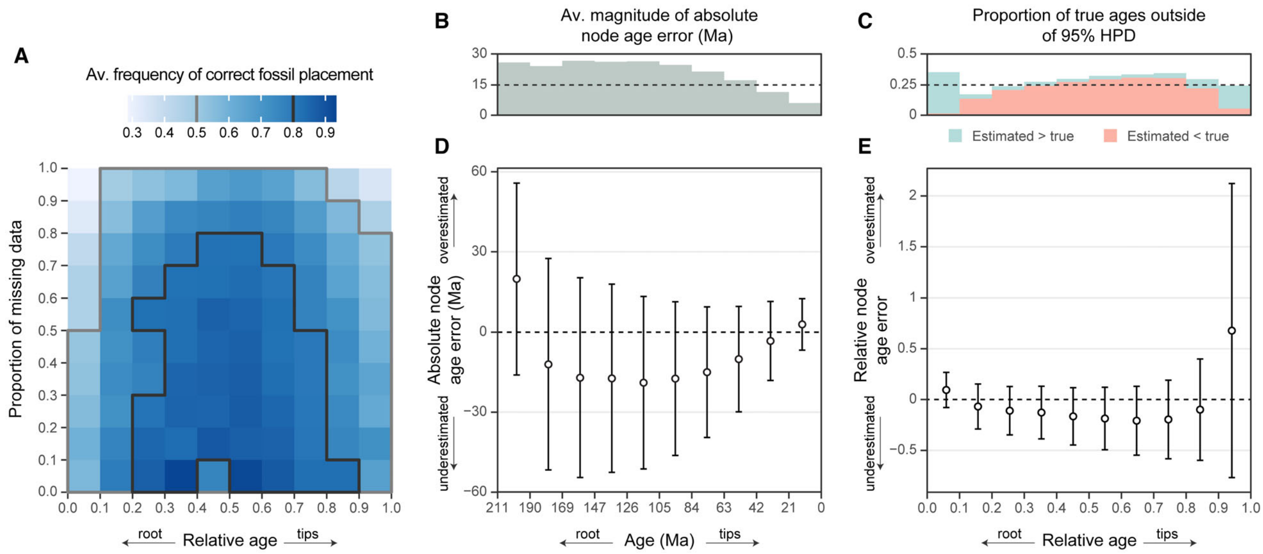

FIG. 1. Determinants of accuracy in fossil placement and node age estimates. A, the frequency of correct fossil placement for various relative ages and proportions of missing data; fossils with small amounts of missing data, and occurring further from the root and tips of the tree, are placed with higher levels of accuracy. B, average magnitude of the absolute error in age estimates for nodes at different depths; error increases with node depth before stabilizing halfway between the root and tips of the tree. C, proportions of nodes with true ages outside the 95% HPD; values oscillate around 25%, with the highest fraction seen close to the root of the tree. D, average error in absolute node age estimates; values correspond to those shown in B, but averaged without removing the sign of the deviation, showing biases in the direction of the estimation error (also evident in C); circles denote the mean difference between true ages and median posterior estimates, error bars show the average width of the 95% HPD; ages for the oldest and youngest nodes (i.e. those towards the root and tips) are generally overestimated, whilst those towards the middle of the tree tend to be underestimated. E, relative error in age estimates for nodes at different depths; values were computed as in D, but error was standardized by dividing by the true ages. Values correspond to analyses under an offset exponential prior for the root age; results under a uniform prior can be found in Figure S8.

using non-parametric (Spearman's rank) correlation coefficients. Generalized linear models incorporating the number of fossil terminals as an additional discrete predictor were also fitted. Before these analyses, multivariate outliers whose squared Mahalanobis distance exceeded the 95% quantile value of a chi-square distribution were excluded; p-values were corrected for multiple comparisons using the Benjamini & Hochberg (1995) approach.

RESULTS

The accuracy with which fossils were placed in the extant scaffold was affected by both their age and proportion of missing data (Fig. 1A). The position of fossils was correctly inferred more frequently for fossils lying approximately half-way between the root of the tree and the present, but decreased for fossils that were younger and older. However, accuracy remained high overall; only highly fragmentary fossils occupying both temporal edges of the tree (e.g. those within the earliest and latest decile) were incorrectly placed relative to their extant counterparts more often than not.

On average, error in the estimation of divergence times increased with node age up until half-way between the root and tips of the tree, and remained constant thereafter, attaining absolute errors between 24.3 and 27 myr (Fig. 1B). This corresponds to between 11.5 and 12.8% of total tree depth. When the sign of these deviations is considered, a convex relationship with age emerges, showing a complex interplay between accuracy and bias (Fig. 1D, E; a somewhat similar pattern can also be seen for analyses run under the joint prior, see Fig. S5). On average, divergence times for nodes lying close to both the root and tips of the tree were systematically overestimated, while all others were underestimated. Absolute error was highest for the deepest nodes, whose inferred ages exceeded true ages by an average of 19.8 myr (Figs 1D, S6); relative error was highest for the youngest nodes, whose inferred ages showed an excess of 68% of the true value (despite an average overestimation of only 2.9 myr; Fig. 1E). While these two regions of the tree also showed the highest levels of error in fossil placement (Fig. 1A), we note that temporal windows with high topological accuracy (i.e. those towards the centre of the time spanned by the phylogeny) still had relatively large absolute levels of temporal error. Despite this, coverage was also relatively high, with 95% HPD intervals including true ages 65.2–83.5% of the time, with the lowest values seen among the nodes closest to the root (Fig. 1C). While Figure 1 shows only results using offset exponential distributions for the root age prior, nearly identical results were obtained when implementing uniform root age priors (Figs S7, S8). The only noticeable difference was a slightly higher proportion of true ages outside the 95% HPD (i.e. lower coverage) for nodes closer to the root of the tree (Fig. S8).

The overall temporal and topological accuracy of analyses (obtained by averaging the data for all node ages and fossil placements) increased with the number of fossils sampled (or, conversely, their overall temporal and topological error decreased; see Fig. S9). While the increase in temporal accuracy likely reflects the positive influence of a larger number of fossil tips informing node ages, the increase in topological accuracy is probably related to the reduced number of branches in the extant subtree to which fossils can attach, as well as their increased length (Fig. S10), that result from a sparser sampling of extant taxa.

If not accounted for, the effect of the level of fossil sampling on both topological and temporal accuracies confounds a direct exploration of the relationship between the two variables. We took two approaches to decouple these effects. First, the data were subdivided by the different levels of fossil sampling, and correlations between temporal and topological accuracy were explored within each subset. Overall, correlations were weak (absolute ρ values ≤0.29), inconsistent in sign (with 5 instances of positive and 9 of negative correlations), and non-significant in 12 out of 14 subsets (Fig. 2). The only two examples of significant correlations between these variables involved negative correlation coefficients, contrary to our expectations (Fig. 2). A second approach involved fitting generalized linear models to predict temporal accuracy using the level of fossil sampling, either exclusively or in combination with topological accuracy. Once again, while likelihood ratio tests favoured the more complex models including both predictors ($p < 0.01$), and the effect of topological accuracy attained significance ($p < 0.01$), its effects were relatively small and negative in sign (estimates = −0.11 and −0.07, depending on the proxies of accuracy employed). The small impact of topological accuracy on temporal accuracy is further reflected by the fact that including this predictor alongside the level of fossil sampling in the linear model only increased the adjusted R^2 values by less than 0.01 (relative to R^2 values between 0.20 and 0.29 when using the level of fossil sampling as sole predictor). Once again, these results are not affected by the root age prior implemented (see Fig. S11 for results under a uniform prior). Further statistical details can be found in Table S2.

DISCUSSION

Phylogenetic trees with optimal levels of fossil sampling are key to reconstructing evolutionary history (Quental & Marshall 2010; Slater & Harmon 2013; Rothwell *et al.* 2018) and have illuminated the diversification of clades in ways unattainable by studies focusing only on living representatives (Slater 2013; Finarelli & Goswami 2013; Garwood *et al.* 2014; Betancur-R *et al.* 2015; Mitchell 2015; Arcila & Tyler 2017; Vinther *et al.* 2017; Norrell *et al.* 2020; Lloyd & Slater 2021; Mongiardino Koch & Thompson 2021; Wisniewski *et al.* 2022). The macroevolutionary potential of the fossil record is best unlocked through the use of tip-dated methods of inference that place fossil and living taxa in a common time-calibrated phylogenetic framework, built using mechanistic models of diversification and sampling, and informed by all available sources of information (Ronquist *et al.* 2012a; Gavryushkina *et al.* 2014; Zhang *et al.* 2016; Heath *et al.* 2017; Warnock & Wright 2021). One of the major benefits of this approach is that it facilitates the inclusion and extensive use of palaeontological data (O'Reilly *et al.* 2015; Donoghue & Yang 2016; O'Reilly & Donoghue 2020). Nonetheless, for these benefits to be realized, morphology must retain some degree of temporal signal (i.e. 'morphological clocks' need to exist, at least to some degree; Polly 2001; Beck & Lee 2014; Lee *et al.* 2014; King *et al.* 2017; Parins-Fukuchi & Brown 2017; Barba-Montoya *et al.* 2021), and morphological data need to allow accurate inference of the phylogenetic position of fossil terminals (Sansom *et al.* 2010; Sansom & Wills 2013; Guillerme & Cooper 2016; Luo *et al.* 2020; Mongiardino Koch *et al.* 2021; Barido-Sottani *et al.* 2023).

Our results show that morphological datasets of 300 characters and with realistic levels of missing data generally succeed in inferring both the position of fossil terminals and the divergence times of extant lineages, at least when combined with topological constraints for living terminals. The correct position of fossils relative to fixed extant scaffolds is generally attained with high confidence, although this is impaired by high levels of missing data, and is particularly difficult for both very ancient and very recent fossil taxa (Fig. 1A). These patterns are also seen in empirical morphological datasets, which generally target divergent taxa to avoid issues of poor resolution, and often exhibit poor topological accuracy for deep splits (Scotland *et al.* 2003; Ronquist *et al.* 2012a; Pyron 2015; Halanych 2016; Ronquist *et al.* 2016; Zhang *et al.* 2016). Similarly, despite an increase in the error of inferred node

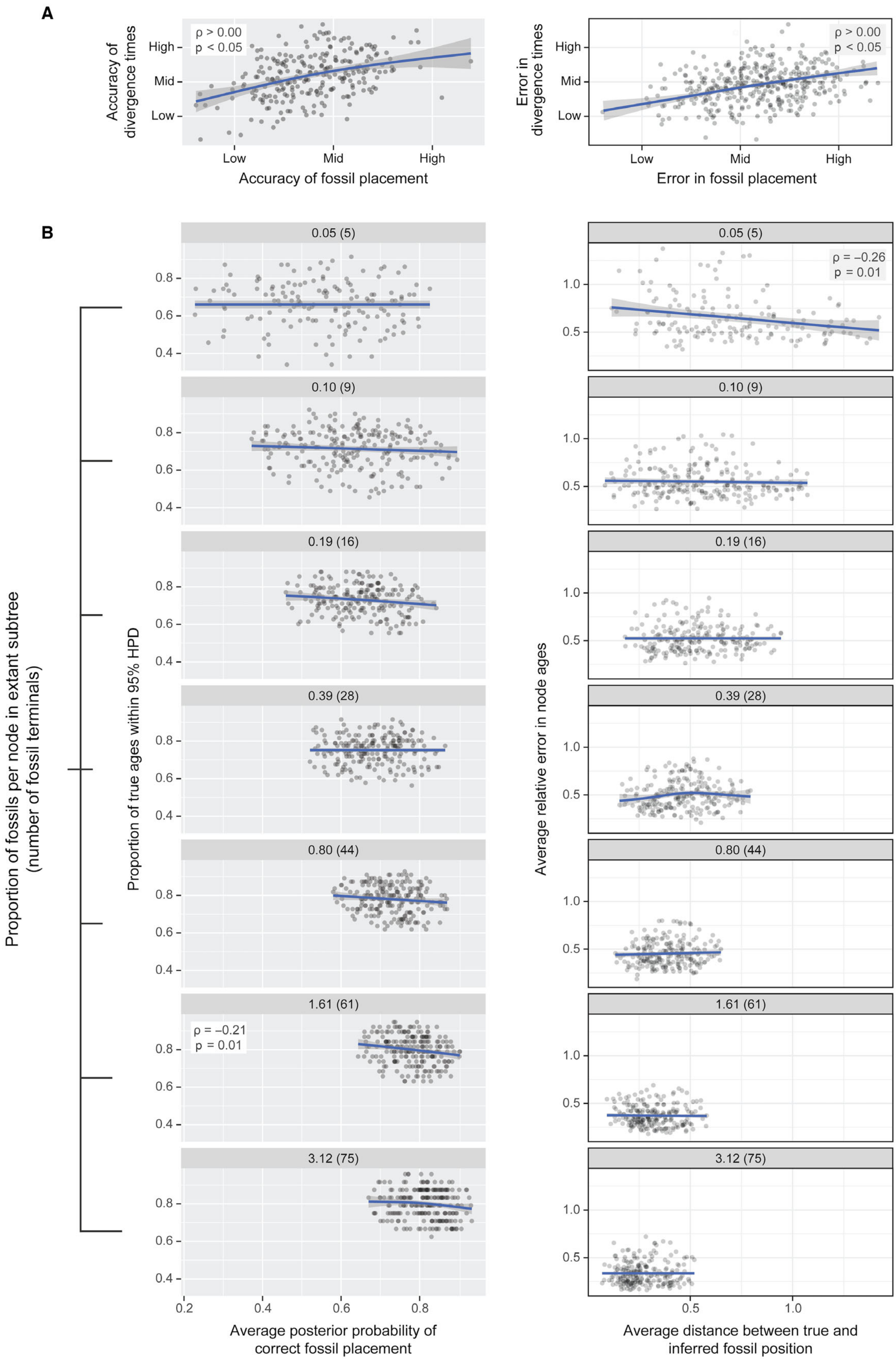
A
Accuracy of divergence times
High
Mid
Low
ρ > 0.00
p < 0.05
Accuracy of fossil placement
Error in divergence times
Error in fossil placement
B
Proportion of fossils per node in extant subtree (number of fossil terminals)
Proportion of true ages within 95% HPD
Average relative error in node ages
0.05 (5)
0.10 (9)
0.19 (16)
0.39 (28)
0.80 (44)
1.61 (61)
3.12 (75)
ρ = −0.26
p = 0.01
ρ = −0.21
p = 0.01
Average posterior probability of correct fossil placement
Average distance between true and inferred fossil position

FIG. 2. Higher topological accuracy does not translate into higher temporal accuracy. Left and right sets of graphs show alternative proxies to estimate the success in inferring overall fossil placement and divergence times. A, within each plot, a positive and significant relationship is expected under the hypothesis that accurate fossil placement improves divergence time estimates. B, no such relationship is obtained, with topological accuracy being either uncorrelated or negatively correlated with temporal accuracy (ρ and p values correspond to Spearman's rank correlations). Each dot represents a tree-wide summary of the accuracy/error obtained for a given dataset. Trend lines were fitted using general additive models, and are shown for visualization purposes only. Values correspond to analyses under an offset exponential prior for the root age; results under a uniform prior can be found in Figure S11.

ages towards the root of the tree (Fig. 1B), true ages are generally included within the 95% HPD intervals that are commonly used to summarize posterior distributions of trees (Fig. 1C).

Despite the overall high levels of accuracy, we also uncover biases in the inference of divergence times (Fig. 1D, E and Fig. S7D, E). These manifest in the overestimation of node ages close to both the root and the tips of the tree, while the ages of all remaining nodes tend to be underestimated. Excessively old ages for deep nodes were also a commonly reported artefact of early tip-dating analyses (e.g. Ronquist *et al.* 2012a; Slater 2013; Arcila *et al.* 2015). A number of drivers were proposed for this 'deep root attraction' phenomenon, such as conflicts between morphological and molecular datasets, the use of inappropriate uniform tree priors, and a failure to accommodate for diversified sampling, or to account for morphological integration (Arcila *et al.* 2015; Ronquist *et al.* 2016; Zhang *et al.* 2016), but none of these apply to our simulations. Recently, Luo *et al.* (2020) suggested that instances of overestimation of deep node ages might be linked to fossil misplacements. Our analysis also finds node ages to be overestimated in regions of the tree where fossil placement is most inaccurate. Even so, regions of the tree where fossils tend to be placed with high accuracy still show relatively large absolute errors as well as biased estimates, although in this case the tendency is towards recovering systematically younger dates. Therefore, these effects seem unlikely to be driven by a failure to correctly place fossils, but rather by a mismatch between the constant-rate FBD process implemented and the overall shape of our phylogenetic trees (Figs S5, S6). In fact, previous studies that simulated topologies under birth–death processes found little to no overestimation of the age of deep nodes (Zhang *et al.* 2016; Luo *et al.* 2020), which further highlights the importance of simulation approaches that incorporate some degree of model misspecification, as is expected of empirical datasets (O'Reilly *et al.* 2016; Ronquist *et al.* 2016; Goloboff *et al.* 2018; Mongiardino Koch *et al.* 2021). If these biases do arise from model misspecification, they might not affect tip dating in general, and only impact clades that have undergone ancient and rapid radiations (see also Budd & Mann 2020). For such clades, it remains possible that the addition of molecular data would succeed in overcoming these biases, although divergence times derived from molecular data can also be highly sensitive to model priors (see e.g. Warnock *et al.* 2015; Brown & Smith 2018; Mongiardino Koch *et al.* 2022a; Sauquet *et al.* 2022). Alternatively, skyline FBD models that accommodate changes in rate parameters through time should also be considered when there is sufficient *a priori* information that diversification or sampling rates have been variable (Zhang *et al.* 2016; Simões *et al.* 2020; May *et al.* 2021; Wright *et al.* 2021; Barido-Sottani *et al.* 2023). Still, studies inferring tip-dated trees under a morphological clock should be aware that node ages might be strongly determined by the choice of tree prior, and that any given FBD model can still impose biases if it fails to capture relevant aspects of the history of diversification (Ronquist *et al.* 2016; Matschiner 2019).

CONCLUSION

While previous studies have cautioned that correct fossil placement was probably necessary to infer accurate divergence times (O'Reilly *et al.* 2015; Donoghue & Yang 2016; Luo *et al.* 2020), we find no evidence for this hypothesized effect. Although extremely inaccurate fossil placements have been shown to distort divergence time analyses (Near *et al.* 2005; Lee 2009; Barido-Sottani *et al.* 2023), tip-dating using morphological datasets on the scale of 300 traits, and implementing extant scaffolds, is sufficiently robust to result in broadly accurate time-trees. This finding is particularly noteworthy as our simulation pipeline incorporated all major caveats thought to complicate the use of morphological clocks: we incorporated adaptive phenotypic variation, used overly simplistic models of character change, and did not enforce constancy of evolutionary rates (Sanderson 1998; Beck & Lee 2014; Parins-Fukuchi & Brown 2017; Barba-Montoya *et al.* 2021).

Furthermore, divergence times significantly improve with increased fossil sampling, but are not strongly impacted by the success with which fossil tips are placed relative to extant taxa (Fig. 2 and Fig. S11). This result is encouraging as it demonstrates that a thorough sampling of fossils is beneficial for tip dating, even when many of these taxa prove unstable, or even resolve in slightly incorrect positions. Palaeontologists should be aware of the general difficulties faced when attempting to infer the relationships of ancient, young, and highly incomplete fossil terminals, but their *a priori* exclusion from tip-dated inference under morphological clocks is not justified. When coupled with molecular scaffolds for extant lineages, tip dating benefits from the advantages of co-estimating tree topology and divergence times, without suffering from its potential drawbacks.

Acknowledgements. Analyses were run using the HTCondor service provided by the IT Services Research Infrastructure team at the University of Manchester, which runs the HTCondor software developed by the CHTC Team at UW-Madison, Wisconsin. The HTCondor service also uses the AWS Marketplace Community AMI HTCondor image used for 'bursting into the (AWS) cloud'.

This work was supported by the Natural Environment Research Council (NERC, grant no. NE/T000813/1 awarded to RJG) and by the National Science Foundation (grant no. DEB-2036186 awarded to Greg W. Rouse). LAP was supported by an early career fellowship from St Edmund Hall, University of Oxford and a NERC independent research fellowship (grant no. NE/W007878/1), NMK through the aforementioned NSF DEB grant.

This manuscript benefitted from discussions with Dahiana Arcila, and was substantially improved by the comments of two anonymous referees.

Author contributions. NMK, LAP and RJG conceived and designed the study, wrote the manuscript, and approved final publication. RJG performed the simulations and coded TREvoSim. NMK and RJG performed the phylogenetic analyses. NMK wrote the analytical pipeline in R, and designed the downstream statistical analyses.

DATA ARCHIVING STATEMENT

Data for this study including supplementary figures, analytical scripts, software settings, and simulated datasets, are available in the Zenodo Repository: https://doi.org/10.5281/zenodo.7125201. This is Paleobiology Database official publication number 469.

Editor. Gustavo Ballen

SUPPORTING INFORMATION

Additional Supporting Information can be found online (https://doi.org/10.1111/pala.12680):

Fig. S1. Histograms showing the number of extant and fossil taxa displaying different proportions of missing data, gathered from the same 12 empirical datasets employed to validate other aspects of the evolutionary simulations.

Fig. S2. Average amounts of missing data across empirical and simulated datasets.

Fig. S3. Empirical and fitted distributions of fossil longevities in millions of years.

Fig. S4. Alternative prior probabilities used to constrain the root age of the FBD analyses, exemplified with a dataset whose earliest fossil representative had an older bound at 199.13 Ma.

Fig. S5. Average error in absolute node age estimates obtained under the joint prior.

Fig. S6. Examples of mismatches between inferred and true extant subtrees.

Fig. S7. Determinants of accuracy in fossil placement and node age estimates.

Fig. S8. Comparison of node age coverage (i.e. fraction of true ages contained within 95% HPD inference intervals) for analyses run under uniform and offset exponential priors for the age of the root.

Fig. S9. Increased fossil sampling results in higher topological accuracy (i.e. increased probability of inferring the correct position of fossil terminals, decreased distance between true and inferred positions) and higher temporal accuracy (increased proportion of true node ages contained within 95% HPD, decreased error in node age estimates).

Fig. S10. The average length of branches in the extant subtree increases as fossils represent a larger fraction of sampled terminals.

Fig. S11. Higher topological accuracy does not translate into higher temporal accuracy.

Table S1. Convergence statistics for all parameters inferred.

Table S2. Detailed results of the generalized linear models' fit.

REFERENCES

ABERER, A. J., KROMPASS, D. and STAMATAKIS, A. 2013. Pruning rogue taxa improves phylogenetic accuracy: an efficient algorithm and webservice. *Systematic Biology*, **62** (1), 162–166.

ARCILA, D. and TYLER, J. C. 2017. Mass extinction in tetraodontiform fishes linked to the Palaeocene–Eocene thermal maximum. *Proceedings of the Royal Society B*, **284**, 20171771.

ARCILA, D., PYRON, R. A., TYLER, J. C., ORTÍ, G. and BETANCUR-R, R. 2015. An evaluation of fossil tip-dating versus node-age calibrations in tetraodontiform fishes (Teleostei: Percomorphaceae). *Molecular Phylogenetics & Evolution*, **82**, 131–145.

ASAR, Y., HO, S. Y. W. and SAUQUET, H. 2022. Early diversifications of angiosperms and their insect pollinators: were they unlinked? *Trends in Plant Science*, **27** (9), 858–869.

BALLESTEROS, J. A. and SHARMA, P. P. 2019. A critical appraisal of the placement of Xiphosura (Chelicerata) with account of known sources of phylogenetic error. *Systematic Biology*, **68** (6), 896–917.

BAPST, D. W. 2012. paleotree: an R package for paleontological and phylogenetic analyses of evolution. *Methods in Ecology & Evolution*, **3**, 803–807.

BARBA-MONTOYA, J., TAO, Q. and KUMAR, S. 2021. Molecular and morphological clocks for estimating evolutionary divergence times. *BMC Ecology & Evolution*, **21**, 83.

BARIDO-SOTTANI, J., van TIEL, N., HOPKINS, M. J., WRIGHT, D. F., STADLER, T. and WARNOCK, R. 2020. Ignoring fossil age uncertainty leads to inaccurate topology and divergence time estimates in time calibrated tree inference. *Frontiers in Ecology & Evolution*, **8**, 183.

BARIDO-SOTTANI, J., POHLE, A., DE BAETS, K., MURDOCK, D. and WARNOCK, R. C. M. 2023. Putting the F in FBD analyses: tree constraints or morphological data? *Palaeontology*, **66**, e12679.

BECK, R. M. and LEE, M. S. 2014. Ancient dates or accelerated rates? Morphological clocks and the antiquity of placental mammals. *Proceedings of the Royal Society B*, **281**, 1–10.

BENJAMINI, Y. and HOCHBERG, Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B*, **57**, 289–300.

BETANCUR-R, R., ORTÍ, G. and PYRON, R. A. 2015. Fossil-based comparative analyses reveal ancient marine ancestry erased by extinction in ray-finned fishes. *Ecology Letters*, **18**, 441–450.

BROOKS, S. P. and GELMAN, A. 1997. General methods for monitoring convergence of iterative simulations. *Journal of Computational & Graphical Statistics*, **7**, 434–455.

BROWN, J. W. and SMITH, S. A. 2018. The past sure is tense: on interpreting phylogenetic divergence time estimates. *Systematic Biology*, **67** (2), 340–353.

BUDD, G. E. and MANN, R. P. 2020. Survival and selection biases in early animal evolution and a source of systematic overestimation in molecular clocks. *Interface Focus*, **10**, 20190110.

CARRUTHERS, T. and SCOTLAND, R. W. 2020. Insights from empirical analyses and simulations on using multiple fossil calibrations with relaxed clocks to estimate divergence times. *Molecular Biology & Evolution*, **37** (5), 1508–1529.

COOMBS, E. J., FELICE, R. N., CLAVEL, J., PARK, T., BENNION, R. F., CHURCHILL, M., GEISLER, J. H., BEATTY, B. and GOSWAMI, A. 2022. The tempo of cetacean cranial evolution. *Current Biology*, **32**, 2233–2247.

CRAWFORD, N. G., PARHAM, J. F., SELLAS, A. B., FAIRCLOTH, B. C., GLENN, T. C., PAPENFUSS, T. J., HENDERSON, J. B., HANSEN, M. H. and SIMISON, W. B. 2015. A phylogenomic analysis of turtles. *Molecular Phylogenetics & Evolution*, **83**, 250–257.

DARLIM, G., LEE, M. S. Y., WALTER, J. and RABI, M. 2022. The impact of molecular data on the phylogenetic position of the putative oldest crown crocodilian and the age of the clade. *Biology Letters*, **18**, 20210603.

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.

DONOGHUE, P. C. J. and YANG, Z. 2016. The evolution of methods for establishing evolutionary timescales. *Philosophical Transactions of the Royal Society B*, **B371**, 20160020.

dos REIS, M., DONOGHUE, P. C. and YANG, Z. 2015. Bayesian molecular clock dating of species divergences in the genomics era. *Nature Reviews Genetics*, **17**, 71–80.

DRUMMOND, A. J., HO, S. Y. W., PHILLIPS, M. J. and RAMBAUT, A. 2006. Relaxed phylogenetics and dating with confidence. *PLoS Biology*, **4** (5), e88.

FINARELLI, J. A. and GOSWAMI, A. 2013. Potential pitfalls of reconstructing deep time evolutionary history with only extant data, a case study using the Canidae (Mammalia, Carnivora). *Evolution*, **67**, 3678–3685.

FISHER, D. C. 1992. Stratigraphic parsimony. 124–129. *In* MADDISON, W. P. and MADDISON, D. R. (eds) *MacClade: Analysis of phylogeny and character evolution. Version 3.0*. Sinauer Associates.

GARWOOD, R. J. 2021. TREvoSim v2.0.0. https://github.com/palaeoware/trevosim

GARWOOD, R. J., SHARMA, P. P., DUNLOP, J. A. and GIRIBET, G. 2014. A Paleozoic stem croup to mite harvestmen revealed through integration of phylogenetics and development. *Current Biology*, **24** (9), 1017–1023.

GARWOOD, R. J., SPENCER, A. R. T. and SUTTON, M. D. 2019. REvoSim: organism-level simulation of macro and microevolution. *Palaeontology*, **62**, 339–355.

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.

GOLOBOFF, P. A., TORRES, A. and ARIAS, J. S. 2018. Weighted parsimony outperforms other methods of phylogenetic inference under models appropriate for morphology. *Cladistics*, **34** (4), 407–437.

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

HALANYCH, K. M. 2016. How our view of animal phylogeny was reshaped by molecular approaches: lessons learned. *Organisms Diversity & Evolution*, **16**, 319–328.

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.

HEATH, T. A., KSEPKA, D. T., STADLER, T., WELCH, D. and DRUMMOND, A. J. 2017. Bayesian total evidence dating reveals the recent crown radiation of penguins. *Systematic Biology*, **66**, 57–73.

HIPSLEY, C. A. and MÜLLER, J. 2014. Beyond fossil calibrations: realities of molecular clock practices in evolutionary biology. *Frontiers in Genetics*, **5**, 00138.

HO, S. Y. W. and PHILLIPS, M. J. 2009. Accounting for calibration uncertainty in phylogenetic estimation of evolutionary divergence times. *Systematic Biology*, **58** (3), 367–380.

JABŁOŃSKA, J. and TAWFIK, D. S. 2021. The evolution of oxygen-utilizing enzymes suggests early biosphere oxygenation. *Nature Ecology & Evolution*, **5**, 442–448.

KEATING, J. N., SANSOM, R. S., SUTTON, M. D., KNIGHT, C. G. and GARWOOD, R. J. 2020. Morphological phylogenetics evaluated using novel evolutionary simulations. *Systematic Biology*, **69** (5), 897–912.

KING, B. 2021. Bayesian tip-dated phylogenetics in paleontology: topological effects and stratigraphic fit. *Systematic Biology*, **70** (2), 283–294.

KING, B., QIAO, T., LEE, M. S., ZHU, M. and LONG, J. A. 2017. Bayesian morphological clock methods resurrect placoderm monophyly and reveal rapid early evolution in jawed vertebrates. *Systematic Biology*, **66**, 499–516.

KISHINO, H., THORNE, J. L. and BRUNO, W. J. 2001. Performance of a divergence time estimation method under a probabilistic model of rate evolution. *Molecular Biology & Evolution*, **18** (3), 352–361.

KLOPFSTEIN, S., RYER, R., COIRO, M. and SPASOJEVIC, T. 2019. Mismatch of the morphology model is mostly unproblematic in total-evidence dating: insights from an extensive simulation study. bioRxiv, 679084. https://doi.org/10.1101/679084

KSEPKA, D. T., PARHAM, J. F., ALLMAN, J. F., BENTON, M. J., CARRANO, M. T., CRANSTON, K. A., DONOGHUE, P. C. J., HEAD, J. J., HERMSEN, E. J., IRMIS, R. B., JOYCE, W. G., KOHLI, M., LAMM, K. D., LEEHR, D., PATANÉ, J. L., POLLY, P. D., PHILLIPS, M. J., SMITH, N. A., SMITH, N. D., VAN TUINEN, M., WARE, J. L. and WARNOCK, R. C. M. 2015. The fossil calibration database—a new resource for divergence dating. *Systematic Biology*, **64** (5), 853–859.

LANDIS, M. J., QUINTERO, I., MUÑOZ, M. M., ZAPATA, F. and DONOGHUE, M. J. 2022. Phylogenetic inference of where species spread or split across barriers. *Proceedings of the National Academy of Sciences*, **119** (13), e2116948119.

LEE, M. S. Y. 2009. Molecular clock calibrations and metazoan divergence dates. *Journal of Molecular Evolution*, **49**, 385–391.

LEE, M. S. Y. and PALCI, A. 2015. Morphological phylogenetics in the genomic age. *Current Biology*, **25**, R922–R929.

LEE, M. S. Y. and YATES, A. M. 2018. Tip-dating and homoplasy: reconciling the shallow molecular divergences of modern gharials with their long fossil record. *Proceedings of the Royal Society B*, **285**, 20181071.

LEE, M. S. Y., CAU, A., NAISH, D. and DYKE, G. J. 2014. Morphological clocks in paleontology, and a mid-Cretaceous origin of crown aves. *Systematic Biology*, **63** (3), 442–449.

LEPAGE, T., BRYANT, D., PHILIPPE, H. and LARTILLOT, N. 2007. A general comparison of relaxed molecular clock models. *Molecular Biology & Evolution*, **24** (12), 2669–2680.

LEWIS, P. O. 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. *Systematic Biology*, **50** (6), 913–925.

LI, Y., SHEN, X. X., EVANS, B., DUNN, C. W. and ROKAS, A. 2021. Rooting the animal tree of life. *Molecular Biology & Evolution*, **38** (10), 4322–4333.

LLOYD, G. T. 2016. Estimating morphological diversity and tempo with discrete character-taxon matrices: implementation, challenges, progress, and future directions. *Biological Journal of the Linnean Society*, **118**, 131–151.

LLOYD, G. T. and SLATER, G. J. 2021. A total-group phylogenetic metatree for Cetacea and the importance of fossil data in diversification analyses. *Systematic Biology*, **70** (5), 922–939.

LOUCA, S. and DOEBELI, M. 2017. Efficient comparative phylogenetics on large trees. *Bioinformatics*, **34** (6), 1053–1055.

LUO, A., DUCHÊNE, D. A., ZHANG, C., ZHU, C.-D. and HO, S. Y. W. 2020. A simulation-based evaluation of tip-dating under the fossilized birth–death process. *Systematic Biology*, **69** (2), 325–344.

LUO, A., ZHANG, C., ZHOU, Q.-S., HO, S. Y. W. and ZHU, C.-D. 2023. Impacts of taxon-sampling schemes on Bayesian tip dating under the fossilized birth-death process. *Systematic Biology*, **72**, 781–801.

MARSHALL, C. R. 2019. Using the fossil record to evaluate timetree timescales. *Frontiers in Genetics*, **10**, 1049. https://doi.org/10.3389/fgene.2019.01049

MATSCHINER, M. 2019. Selective sampling of species and fossils influences age estimates under the fossilized birth–death model. *Frontiers in Genetics*, **10**, 1064. https://doi.org/10.3389/fgene.2019.01064

MATZKE, N. J. and IRMIS, R. B. 2018. Including autapomorphies is important for paleontological tip-dating with clocklike data, but not with non-clock data. *PeerJ*, **6**, e4553.

MATZKE, N. J. and WRIGHT, A. 2016. Inferring node dates from tip dates in fossil Canidae: the importance of tree priors. *Biology Letters*, **12**, 20160328.

MAY, M. R. and ROTHFELS, C. J. 2023. Diversification models conflate likelihood and prior, and cannot be compared using conventional model-comparison tools. *Systematic Biology*, **72**, 713–722. https://doi.org/10.1093/sysbio/syad010

MAY, M. R., CONTRERAS, D. L., SUNDUE, M. A., NAGALINGUM, N. S., LOOY, C. V. and ROTHFELS, C. J. 2021. Inferring the total-evidence timescale of marattialean fern evolution in the face of model sensitivity. *Systematic Biology*, **70** (6), 1232–1255.

MILLS, D. B., BOYLE, R. A., DAINES, S. J., SPERLING, E. A., PISANI, D., DONOGHUE, P. C. J. and LENTON, T. M. 2022. Eukaryogenesis and oxygen in Earth history. *Nature Ecology & Evolution*, **6**, 520–532.

MITCHELL, J. S. 2015. Extant only comparative methods fail to recover the disparity preserved in the bird fossil record. *Evolution*, **69**, 2414–2424.

MONGIARDINO KOCH, N. 2021. Exploring adaptive landscapes across deep time: a case study using echinoid body size. *Evolution*, **75** (6), 1567–1581.

MONGIARDINO KOCH, N. and PARRY, L. A. 2020. Death is on our side: paleontological data drastically modify phylogenetic hypotheses. *Systematic Biology*, **69** (6), 1052–1067.

MONGIARDINO KOCH, N. and THOMPSON, J. T. 2021. A total-evidence dated phylogeny of Echinoidea combining phylogenomic and paleontological data. *Systematic Biology*, **70** (3), 421–439.

MONGIARDINO KOCH, N., GARWOOD, R. J. and PARRY, L. A. 2021. Fossils improve phylogenetic analyses of morphological characters. *Proceedings of the Royal Society B*, **288**, 20210044.

MONGIARDINO KOCH, N., THOMPSON, J. R., HILEY, A. S., McCOWIN, M. F., ARMSTRONG, A. F., COPPARD, S. E., AGUILERA, F., BRONSTEIN, O., KROH, A., MOOI, R. and ROUSE, G. W. 2022a. Phylogenomic analyses of echinoid diversification prompt a re-evaluation of their fossil record. *eLife*, **11**, e72460.

MONGIARDINO KOCH, N., GARWOOD, R. J. and PARRY, L. A. 2022b. Inaccurate fossil placement does not compromise tip-dated divergence times. v1. Zenodo. https://doi.org/10.5281/zenodo.7125200

MONGIARDINO KOCH, N., TILIC, E., MILLER, A. K., STILLER, J. and ROUSE, G. W. 2023. Confusion will be my epitaph: genome-scale discordance stifles phylogenetic resolution of Holothuroidea. *Proceedings of the Royal Society B*, **290** (2002), 20230988.

MOODY, E. R. R., MAHENDRARAJAH, T. A., DOMBROWSKI, N., CLARK, J. W., PETITJEAN, C., OFFRE, P., SZÖLLŐSI, G. J., SPANG, A. and WILLIAMS, T. A. 2022. An estimate of the deepest branches of the tree of life from ancient vertically evolving genes. *eLife*, **11**, e66695.

MULHAIR, P. O., McCARTHY, C. G., SIU-TING, K., CREEVEY, C. J. and O'CONNELL, M. J. 2022. Filtering artifactual signal increases support for Xenacoelomorpha and Ambulacraria sister relationship in the animal tree of life. *Current Biology*, **32** (23), 5180–5188.

NEAR, T. J., MEYLAN, P. A. and SHAFFER, H. B. 2005. Assessing the concordance of fossil calibration points in molecular clock studies: an example using turtles. *The American Naturalist*, **165** (2), 137–146.

NORELL, M. A., WIEMANN, J., FABBRI, M., YU, C., MARSICANO, C. A., MOORE-NALL, A., VARRICCHIO, D. J., POL, D. and ZELENITSKY, D. K. 2020. The first dinosaur egg was soft. *Nature*, **583** (7816), 406–410.

O'HARA, T. D., HUGALL, A. F., WOOLLEY, S. N. C., BRIBIESCA-CONTRERAS, G. and BAX, N. J. 2019. Contrasting processes drive ophiuroid phylodiversity across shallow and deep seafloors. *Nature*, **565**, 636–639.

ONSTEIN, R. E., KISSLING, W. D. and LINDER, H. P. 2022. The megaherbivore gap after the non-avian dinosaur extinctions modified trait evolution and diversification of tropical palms. *Proceedings of the Royal Society B*, **289**, 20212633.

O'REILLY, J. E. and DONOGHUE, P. C. J. 2016. Tips and nodes are complementary not competing approaches to the calibration of molecular clocks. *Biology Letters*, **12**, 20150975.

O'REILLY, J. E. and DONOGHUE, P. C. J. 2020. The effect of fossil sampling on the estimation of divergence times with the fossilized birth–death process. *Systematic Biology*, **69** (1), 124–138.

O'REILLY, J. E., dos REIS, M. and DONOGHUE, P. C. J. 2015. Dating tips for divergence-time estimation. *Trends in Genetics*, **31** (11), 637–650.

O'REILLY, J. E., PUTTICK, M. N., PARRY, L. A., TANNER, A. R., TARVER, J. E., FLEMING, J., PISANI, D. and DONOGHUE, P. C. J. 2016. Bayesian methods outperform parsimony but at the expense of precision in the estimation of phylogeny from discrete morphological data. *Biology Letters*, **12** (4), 20160081.

PALAZZESI, L., HIDALGO, O., BARREDA, V. D., FOREST, F. and HÖHNA, S. 2022. The rise of grasslands is linked to atmospheric CO_2 decline in the late Palaeogene. *Nature Communications*, **13**, 293.

PARADIS, E. and SCHLIEP, K. 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, **35**, 526–528.

PARHAM, J. F., DONOGHUE, P. C. J., 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. and BENTON, M. J. 2012. Best practices for justifying fossil calibrations. *Systematic Biology*, **61** (2), 346–359.

PARINS-FUKUCHI, C. and BROWN, J. W. 2017. What drives results in Bayesian morphological clock analyses? bioRxiv, 219048. https://doi.org/10.1101/219048

POLLY, P. D. 2001. On morphological clocks and paleophylogeography: towards a timescale for *Sorex* hybrid zones. *Genetica*, **112–113**, 339–357.

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

PYRON, R. A. 2015. Post-molecular systematics and the future of phylogenetics. *Trends in Ecology & Evolution*, **30** (7), 384–389.

QUENTAL, T. B. and MARSHALL, C. R. 2010. Diversity dynamics: molecular phylogenies need the fossil record. *Trends in Ecology & Evolution*, **25** (8), 434–441.

R CORE TEAM. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

REVELL, L. J. 2012. phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology & Evolution*, **3**, 217–223.

RONQUIST, F., KLOPFSTEIN, S., VILHELMSEN, L., SCHULMEISTER, S., MURRAY, D. L. and RASNITSYN, A. P. 2012a. A total-evidence approach to dating with fossils, applied to the early radiation of the Hymenoptera. *Systematic Biology*, **61** (6), 973–999.

RONQUIST, F., TESLENKO, M., VAN DER MARK, P., AYRES, D. L., DARLING, A., HÖHNA, S., LARGET, B., LIU, L., SUCHARD, M. A. and HUELSENBECK, J. P. 2012b. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, **61** (3), 539–542.

RONQUIST, F., LARTILLOT, N. and PHILLIPS, M. J. 2016. Closing the gap between rocks and clocks using total-evidence dating. *Philosophical Transactions of the Royal Society B*, **371**, 20150136.

ROTHWELL, G. W., ESCAPA, I. H. and TOMESCU, A. M. 2018. Tree of death: the role of fossils in resolving the overall pattern of plant phylogeny. *American Journal of Botany*, **105**, 1239–1242.

SANDERSON, M. J. 1998. Estimating rate and time in molecular phylogenies: beyond the molecular clock? 242–264. *In* SOLTIS, D. E., SOLTIS, P. S. and DOYLE, J. J. (eds) *Molecular systematics of plants II*. Springer.

SANSOM, R. S. and WILLS, M. A. 2013. Fossilization causes organisms to appear erroneously primitive by distorting evolutionary trees. *Scientific Reports*, **3**, 2545.

SANSOM, R. S., GABBOTT, S. E. and PURNELL, M. A. 2010. Non-random decay of chordate characters causes bias in fossil interpretation. *Nature*, **463**, 797–800.

SAUQUET, H., RAMÍREZ-BARAHONA, S. and MAGALLÓN, S. 2022. What is the age of flowering plants? *Journal of Experimental Botany*, **73**, 3840–3853.

SCHLIEP, K. P. 2011. phangorn: phylogenetic analysis in R. *Bioinformatics*, **27** (4), 592–593.

SCOTLAND, R. W., OLMSTEAD, R. G. and BENNETT, J. R. 2003. Phylogeny reconstruction: the role of morphology. *Systematic Biology*, **52** (4), 539–548.

SIMÕES, T. R. and PIERCE, S. E. 2021. Sustained high rates of morphological evolution during the rise of tetrapods. *Nature Ecology & Evolution*, **10**, 1403–1414.

SIMÕES, T. R., CALDWELL, M. W., PALCI, A. and NYDAM, R. L. 2017. Giant taxon-character matrices: quality of character constructions remains critical regardless of size. *Cladistics*, **33** (2), 198–219.

SIMÕES, T. R., CALDWELL, M. W. and PIERCE, S. E. 2020. Sphenodontian phylogeny and the impact of model choice in Bayesian morphological clock estimates of divergence times and evolutionary rates. *BMC Biology*, **18**, 191.

SIMÕES, T. R., VERNYGORA, O. V., de MEDEIROS, B. A. S. and WRIGHT, A. M. 2023. Handling logical character dependency in phylogenetic inference: extensive performance testing of assumptions and solutions using simulated and empirical data. *Systematic Biology*, **72**, syad006.

SLATER, G. J. 2013. Phylogenetic evidence for a shift in the mode of mammalian body size evolution at the Cretaceous-Palaeogene boundary. *Methods in Ecology & Evolution*, **4**, 734–744.

SLATER, G. J. and HARMON, L. J. 2013. Unifying fossils and phylogenies for comparative analyses of diversification and trait evolution. *Methods in Ecology & Evolution*, **4**, 699–702.

SPASOJEVIC, T., BROAD, G. R., SÄÄKSJÄRVI, I. E., SCHWARZ, M., ITO, M., KORENKO, S. and KLOPFSTEIN, S. 2021. Mind the outgroup and bare branches in total-evidence dating: a case study of pimpliform Darwin wasps (Hymenoptera, Ichneumonidae). *Systematic Biology*, **70** (2), 322–339.

STRASSERT, J. F. H., IRISARRI, I., WILLIAMS, T. A. and BURKI, F. 2021. A molecular timescale for eukaryote evolution with implications for the origin of red algal-derived plastids. *Nature Communications*, **12**, 1879.

SUISSA, J. S. and FRIEDMAN, W. E. 2022. Rapid diversification of vascular architecture underlies the Carboniferous fern radiation. *Proceedings of the Royal Society B*, **289** (1973), 20212209.

TURNER, A. H., PRITCHARD, A. C. and MATZKE, N. J. 2017. Empirical and Bayesian approaches to fossil-only divergence times: a study across three reptile clades. *PLoS One*, **12**, e0169885.

VENABLES, W. N. and RIPLEY, B. D. 2002. *Modern applied statistics with S*, Fourth edition. Springer.

VINTHER, J., SPERLING, E. A., BRIGGS, D. E. and PETERSON, K. J. 2012. A molecular palaeobiological hypothesis for the origin of aplacophoran molluscs and their derivation from chiton-like ancestors. *Proceedings of the Royal Society B*, **279** (1732), 1259–1268.

VINTHER, J., PARRY, L. A., BRIGGS, D. E. G. and VAN ROY, P. 2017. Ancestral morphology of crown-group molluscs revealed by a new Ordovician stem aculiferan. *Nature*, **542** (7642), 471–474.

WAGNER, P. 1998. A likelihood approach for evaluating estimates of phylogenetic relationship among fossil taxa. *Paleobiology*, **24**, 430–449.

WARNOCK, R. C. M., PARHAM, J. F., JOYCE, W. G., LYSON, T. R. and DONOGHUE, P. C. J. 2015. Calibration uncertainty in molecular dating analyses: there is no substitute for the prior evaluation of time priors. *Proceedings of the Royal Society B*, **282**, 20141013.

WARNOCK, R. C. M. and WRIGHT, A. 2021. *Understanding the tripartite approach to Bayesian divergence time estimation*. Cambridge University Press Elements of Paleontology.

WARNOCK, R. C. M., YANG, Z. and DONOGHUE, P. C. J. 2011. Exploring uncertainty in the calibration of the molecular clock. *Biology Letters*, **8** (1), 156–159.

WHITFIELD, J. B. and LOCKHART, P. J. 2007. Deciphering ancient rapid radiations. *Trends in Ecology & Evolution*, **22** (5), 258–265.

WICKHAM, H. 2016. *ggplot2: Elegant graphics for data analysis*. Springer.

WICKHAM, H. 2019. stringr: simple, consistent wrappers for common string operations. R package version 1.4.0. https://CRAN.R-project.org/package=stringr

WICKHAM, H. 2021. rvest: easily harvest (scrape) web pages. R package version 1.0.0. https://CRAN.R-project.org/package=rvest

WICKHAM, H. and HESTER, J. 2020. readr: read rectangular text data. R package version 1.4.0. https://CRAN.R-project.org/package=readr

WICKHAM, H., FRANÇOIS, R., HENRY, L. and MÜLLER, K. 2021. dplyr: a grammar of data manipulation. R package version 1.0.5. https://CRAN.R-project.org/package=dplyr

WIENS, J. and MOEN, D. 2008. Missing data and the accuracy of Bayesian phylogenetics. *Journal of Systematics & Evolution*, **46**, 307–314.

WISNIEWSKI, A. L., LLOYD, G. T. and SLATER, G. J. 2022. Extant species fail to estimate ancestral geographical ranges at older nodes in primate phylogeny. *Proceedings of the Royal Society B*, **289** (1975), 20212535.

WRIGHT, A. M. and HILLIS, D. M. 2014. Bayesian analysis using a simple likelihood model outperforms parsimony for estimation of phylogeny from discrete morphological data. *PLoS One*, **9** (10), e109210.

WRIGHT, A. M., WAGNER, P. and WRIGHT, D. 2021. *Testing character evolution models in phylogenetic paleobiology: A case study with Cambrian echinoderms*. Cambridge University Press, Elements of Paleontology.

WRIGHT, A. M., BAPST, D. W., BARIDO-SOTTANI, J. and WARNOCK, R. C. M. 2022. Integrating fossil observations into phylogenetics using the fossilized birth-death model. *Annual Review of Ecology, Evolution, & Systematics*, **53**, 12.1–12.23.

WOLFE, J. M. and FOURNIER, G. P. 2018. Horizontal gene transfer constrains the timing of methanogen evolution. *Nature Ecology & Evolution*, **5**, 897–903.

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** (1), 212–226.

ZHANG, C., STADLER, T., KLOPFSTEIN, S., HEATH, T. A. and RONQUIST, F. 2016. Total-evidence dating under the fossilized birth-death process. *Systematic Biology*, **65**, 228–249.

ZOU, Z. and ZHANG, J. 2016. Morphological and molecular convergences in mammalian phylogenetics. *Nature Communications*, **7**, 12758.