Bayesian tip-dated phylogenetics in paleontology: topological effects and stratigraphic fit

K. King

Syst. Biol. 70(2):283–294, 2021

DOI:10.1093/sysbio/syaa057
Advance Access publication July 21, 2020

Bayesian Tip-Dated Phylogenetics in Paleontology: Topological Effects and Stratigraphic Fit

BENEDICT KING[1,2,*]

[1]*Naturalis Biodiversity Center, Postbus 9517, 2300 RA, Leiden, The Netherlands;* [2]*College of Science and Engineering, Flinders University, Adelaide, South Australia 5042, Australia*
[*]*Correspondence to be sent to: Naturalis Biodiversity Center, Leiden, the Netherlands; E-mail: benking315@gmail.com*

Received 04 May 2020; reviews returned 14 July 2020; accepted 15 July 2020
Associate Editor: Robert Thomson

Abstract.—The incorporation of stratigraphic data into phylogenetic analysis has a long history of debate but is not currently standard practice for paleontologists. Bayesian tip-dated (or morphological clock) phylogenetic methods have returned these arguments to the spotlight, but how tip dating affects the recovery of evolutionary relationships has yet to be fully explored. Here I show, through analysis of several data sets with multiple phylogenetic methods, that topologies produced by tip dating are outliers as compared to topologies produced by parsimony and undated Bayesian methods, which retrieve broadly similar trees. Unsurprisingly, trees recovered by tip dating have better fit to stratigraphy than trees recovered by other methods under both the Gap Excess Ratio (GER) and the Stratigraphic Completeness Index (SCI). This is because trees with better stratigraphic fit are assigned a higher likelihood by the fossilized birth-death tree model. However, the degree to which the tree model favors tree topologies with high stratigraphic fit metrics is modulated by the diversification dynamics of the group under investigation. In particular, when net diversification rate is low, the tree model favors trees with a higher GER compared to when net diversification rate is high. Differences in stratigraphic fit and tree topology between tip dating and other methods are concentrated in parts of the tree with weaker character signal, as shown by successive deletion of the most incomplete taxa from two data sets. These results show that tip dating incorporates stratigraphic data in an intuitive way, with good stratigraphic fit an expectation that can be overturned by strong evidence from character data. [fossilized birth-death; fossils; missing data; morphological clock; morphology; parsimony; phylogenetics.]

The question of whether or not the stratigraphic ages of fossils should be taken into account when estimating phylogeny was a major debate within paleontology in the 1990s and early 2000s (Wagner 1995; Lockwood 1998; Siddall 1998; Smith 1998; Fox et al. 1999; Heyning and Thacker 1999; Smith 2000; Geiger et al. 2001; Alroy 2002). Those advocating the incorporation of stratigraphic data into phylogenetic analysis assert that the age of fossils can be informative about the plausibility of competing phylogenetic hypotheses and can therefore be used to improve estimates of phylogenetic relationships (Fisher 2008). Detractors cite, amongst other arguments, the requirement of negative data (fossil absences are taken as true absences (Heyning and Thacker 1999)), that temporal data is not a property of fossil specimens themselves (Geiger et al. 2001) or the incompleteness of the fossil record (Smith 2000). These arguments are reviewed in Fisher (2008).

A number of methods were developed to integrate stratigraphic data with phylogenetic analysis. Wagner (1995) introduced a method to find the most parsimonious cladogram in which sister taxa have overlapping 95% confidence intervals for stratigraphic range. Stratocladistics (Fisher 1994, 2008) is a parsimony method that considers stratigraphic parsimony debt alongside "standard" morphological parsimony debt. Stratigraphic debt is the number of lineage segments that are not preserved in the fossil record, which in stratocladistics are considered *ad hoc* hypotheses requiring differential preservation between lineages (Fisher 2008). The method of Wagner (1998) finds the tree with the maximum likelihood of the observed tree length and stratigraphic debt, calculated using simulations under a defined model of evolution. None of these methods have seen widespread adoption in the paleontological literature, due in part to the arguments against incorporating stratigraphic data, but also partly due to a lack of user-friendly software. StrataPhy is a program for stratocladistics (Marcot and Fox 2008) but was never updated from its original experimental form.

The introduction of Bayesian tip-dated methods (Ronquist et al. 2012a), has seen the return of phylogenetic analysis incorporating stratigraphic data. At the heart of tip-dated methods is the tree prior: the probability distribution of divergence dates and branch lengths. Early tip-dated analyses used the uniform tree prior (Ronquist et al. 2012a), which is relatively uninformative regarding tree shape and can produce unrealistic results (Matzke and Wright 2016). The uniform tree prior has been superseded by serially sampled fossilized birth-death (FBD) tree priors (Stadler 2010), which model diversification, extinction and sampling and allow taxa to be sampled ancestors (Gavryushkina et al. 2014; Heath et al. 2014). These tree priors assume approximately constant rates of diversification, extinction and sampling, although this can be relaxed between different time slices (Gavryushkina et al. 2014). FBD tree priors likely affect tree topology, and indeed assumptions of approximately constant sampling rates between lineages at specified time intervals is a key component of stratocladistics (Fisher 2008).

Tip dating has been used on a number of data sets to infer the phylogeny of clades with fossil members, including Mesozoic birds (Lee et al. 2014b), Mesozoic

mammals (Close et al. 2015), theropod dinosaurs (Lee et al. 2014a; Bapst et al. 2016), pufferfish (Close et al. 2016), and penguins (Gavryushkina et al. 2017). Most of these studies have focused on macroevolutionary patterns and divergence dates. However, Bapst et al. (2016) reported topological differences between tip-dated Bayesian, undated Bayesian and parsimony analyses of a theropod data set. Turner et al. (2017) showed that tip dating can affect tree topology in an analysis of crocodylomorphs by disfavoring long unsampled branches, whilst King et al. (2017) found that tip dating can have major effects on tree topology by attempting to balance inferred rates of evolution.

Stratigraphic fit measures are explicit measures for assessing how well a phylogeny fits with the order of appearance (i.e., geological ages) of its taxa. Historically, they have been put to a number of uses, including assessing the quality of the fossil record (Benton et al. 2000) and comparisons of stratigraphic congruence between taxonomic groups (O'Connor and Wills 2016). These indices include the Stratigraphic Consistency Index (SCI), Manhattan Stratigraphic Measure (MSM), and the Gap Excess Ratio (GER). The SCI (Huelsenbeck 1994) calculates the number of consistent nodes, where a consistent node is one whose oldest descendant is the same age or younger than the oldest descendant of the sister lineage to that node. The MSM (Siddall 1998) and GER (Wills 1999) both rely on measuring the Minimum Implied Gap, a measure of the smallest possible sum of ghost lineages implied by a particular tree topology. For both the SCI and the GER (the two measures used in this study), values can range from 0 (maximally suboptimal) to 1 (optimal).

Here, I investigate the topological differences that result when discrete morphological character matrices are analyzed using parsimony and both undated and tip-dated Bayesian methods, and their relationship with stratigraphic fit, as measured by the GER and SCI. In addition, I explore the hypothesis that taxa with poorly known morphology are responsible for most of the topological differences between tip-dated and other methods.

Materials and Methods

I selected data sets of ichthyosaurs (Ji et al. 2016), eurypterids (Lamsdell and Selden 2017), horseshoe crabs (Bicknell and Pates 2019), baleen whales (Marx and Fordyce 2015), turtles (Perea et al. 2014), sauropods (Poropat et al. 2016), Mesozoic birds (Wang et al. 2015), and early archosauromorphs (Ezcurra 2016). Stratigraphic age data were taken from these publications where available, or otherwise from the Paleobiology Database (http://fossilworks.org/). Stratigraphic ages were converted to age ranges using the International Chronostratigraphy Chart (http://www.stratigraphy.org/index.php/ics-chart-timescale) and the geowhen database (http://www.stratigraphy.org/upload/bak/geowhen/index.html).

A number of taxa were pruned from the following data sets prior to analysis: for the ichthyosaurs data set, all outgroup taxa except *Hupesuchus*; for the horseshoe crabs data set, all non-xiphosuran taxa except synziphosurines; for the turtles data set, all outgroup taxa, with the tree instead rooted on *Odontochelys*; for the Mesozoic birds data set, all modern taxa deleted. These deletions were used to avoid uncertainties regarding outgroup taxa, or extremely poor taxon sampling within large time divisions. Characters that became invariant following taxon pruning were deleted from the data sets, as these are uninformative in parsimony analyses and Bayesian analyses that assume that only variable characters are included (as used here).

Parsimony analyses in TNT (Goloboff and Catalano 2016) employed new technology search, using sectorial search and tree fusing with default settings for 1000 random addition sequences. This was followed by tree bisection and reconnection (TBR) swapping to fully explore tree islands. Due to the large total number of most-parsimonious trees for some data sets, and the computational burden of downstream analyses, I saved a maximum of 1801 trees (equal to the number of trees in the post-burnin Bayesian samples).

Undated Bayesian analyses were performed in MrBayes 3.2.6 (Huelsenbeck and Ronquist 2001; Ronquist and Huelsenbeck 2003; Ronquist et al. 2012b). The Mkv model (which assumes that only variable characters have been included; Lewis 2001) was used, with gamma-distributed among-character rate variation with four rate categories. The prior on the gamma shape parameter was a uniform distribution between 0 and 10. The prior on branch lengths was the default compound gamma Dirichlet (1, 0.1, 1, 1) (Rannala et al. 2012). Four independent runs of each analysis were run for between 10 and 200 million generations (depending on data set), saving 2001 trees, 10% of which were discarded as burn-in. Convergence of the four runs was confirmed in Tracer (Rambaut et al. 2018) and standard deviation of split frequencies <0.01.

Tip-dated Bayesian analyses were performed in BEAST2.6.2 (Bouckaert et al. 2019). The Mkv model (Lewis 2001) was used, with a gamma distribution with four rate categories to account for rate variation across sites. The prior on the gamma shape parameter was a uniform distribution between 0 and 10. Characters were partitioned according to the number of character states, to ensure consistency with the undated MrBayes analyses. The clock model was an uncorrelated lognormal clock (Drummond et al. 2006) with a lognormal distribution prior (mean -5, standard deviation 1.4) on the mean and an exponential distribution prior (mean 1) on the standard deviation. The tree prior was a sampled-ancestor FBD model (Gavryushkina et al. 2014). Conditioning on rho was used for those data sets containing modern taxa (horseshoe crabs, whales, turtles). Rho was set to 1.0, 0.8, and 0.108, respectively based on the proportion of extant taxa included in the analysis. Tip ages were assigned uniform priors over the range of uncertainty. The prior

distribution for origin time was a uniform distribution (0–2000), for birth rate a lognormal distribution (mean −2, standard deviation 1), for death rate and sampling rate an exponential distribution (mean 0.1). Analyses were run for between 200 million and 400 million generations (depending on data set) with 2001 trees saved, 10% of which were discarded as burn-in. Convergence of four independent runs was confirmed in Tracer (Rambaut et al. 2018) and R We There Yet (RWTY) (Warren et al. 2017).

Each data set was also analyzed without morphological data, to examine the influence of the tree model on topology. The recovered relationships in the resulting tree sample might considered the "effective topology prior" by analogy with node dating (Heled and Drummond 2012). However, since the tip dates form part of the data when using birth-death models (Drummond and Stadler 2016), such an analysis is not technically "sampling from the prior". Although referred in the BEAST2 output as the "prior", the probability of the tree and tip dates given the tree model parameters will be referred to as the "tree model likelihood" throughout this article. This is to be distinguished from the "morphological likelihood" (the probability of the morphological data given the tree and site model parameters, that is, the "likelihood" in the BEAST2 output).

To assess the extent of topological differences between the three phylogenetic methods, I calculated Robinson–Foulds distances (Robinson and Foulds 1981) using the package phangorn (Schliep 2011) in R (R Core Team 2018). A subsample of 91 trees (representing 5% of the posterior sample) was taken from each Bayesian analysis, and 91 most parsimonious trees (MPTs) from parsimony analysis (or all MPTs if less than 91). Each tree from each sample was compared to the trees from both of the other methods. Robinson–Foulds distances were rescaled to a percentage difference following Wright and Hillis (2014). In addition to comparisons between methods, all trees from each set were compared to each other as a measure of topological variation (i.e., resolution) within each method. Tree lengths (as parsimony steps) for the Bayesian posterior samples were calculated in Mesquite (Maddison and Maddison 2015).

Stratigraphic fit, using the GER (Wills 1999) and Stratigraphic Completeness Index (Huelsenbeck 1994) was calculated using the R package strap (Bell and Lloyd 2015), using the entire posterior sample, and up to 1801 MPTs for parsimony. To avoid problems with the different treatment of outgroups between methods, outgroups were removed from all trees prior to calculation of stratigraphic fit measures. The modified GER (Wills et al. 2008) was not used in this study as all trees across all methods returned the maximum value of 1.

I investigated the effect of taxon distribution through time on the tree model likelihood. Random taxon age distributions were generated by drawing 100 samples from an exponential (mean 1) or uniform (0,1) distribution, with the values rescaled so that they ranged from 0 to 100. The resulting age distributions were analyzed in BEAST2 following the template outlined above for empirical data sets. For each analysis (four of each distribution), the posterior distribution of net diversification rate (birth rate–death rate in each sampled generation) and the GER were calculated.

To test the extent to which fossils with a high proportion of missing data were responsible for the topological differences between methods, I examined the effect of the sequential removal of incomplete taxa from data sets. The sauropods (Poropat et al. 2016) and early archosauromorph (Ezcurra 2016) data sets were chosen due to the wide range of data completeness across taxa, as well as the large number of taxa. For each deletion iteration (total of five), I removed the six remaining most incomplete taxa and reanalyzed the data and calculated stratigraphic fit as above. As a control, to test whether or not the act of removing taxa changes stratigraphic fit regardless of the completeness of those taxa, I repeated the process but deleted six random taxa in each iteration.

Results

Tree Topology

Tree topologies of undated Bayesian and parsimony trees were more similar to each other than either was to tip-dated Bayesian trees (Fig. 1), with the exception of the Ezcurra (2016) data set, in which the three methods produced trees that were approximately equally different. Trees output from parsimony were more similar to each other than the output from either of the Bayesian methods (Fig. 1). Topological differences between undated Bayesian and parsimony trees are in general no greater than the differences within the posterior sample of undated Bayesian trees. Trees from tip-dated Bayesian analysis were less parsimonious than trees from undated Bayesian analysis across all data sets (Fig. 2).

Stratigraphic Fit

P-values for all stratigraphic fit calculations were highly significant, showing that all methods produce trees with better fit to stratigraphy than expected by chance. As expected, tip-dated approaches produce trees with a better stratigraphic fit than undated Bayesian or parsimony (Fig. 3). Trees produced when the analysis samples solely from the tree model (sampling from the "prior", without morphological data) have a particularly high stratigraphic fit. Stratigraphic fit is lower for the posterior sample of trees from the tip-dated analysis when analyzed with morphological data but still higher than for trees from the other two methods. There is little difference between the stratigraphic fit measures for the parsimony and undated Bayesian methods relative to the differences with the tip-dated results. However, it should be noted that Sansom et al. (2018) reported

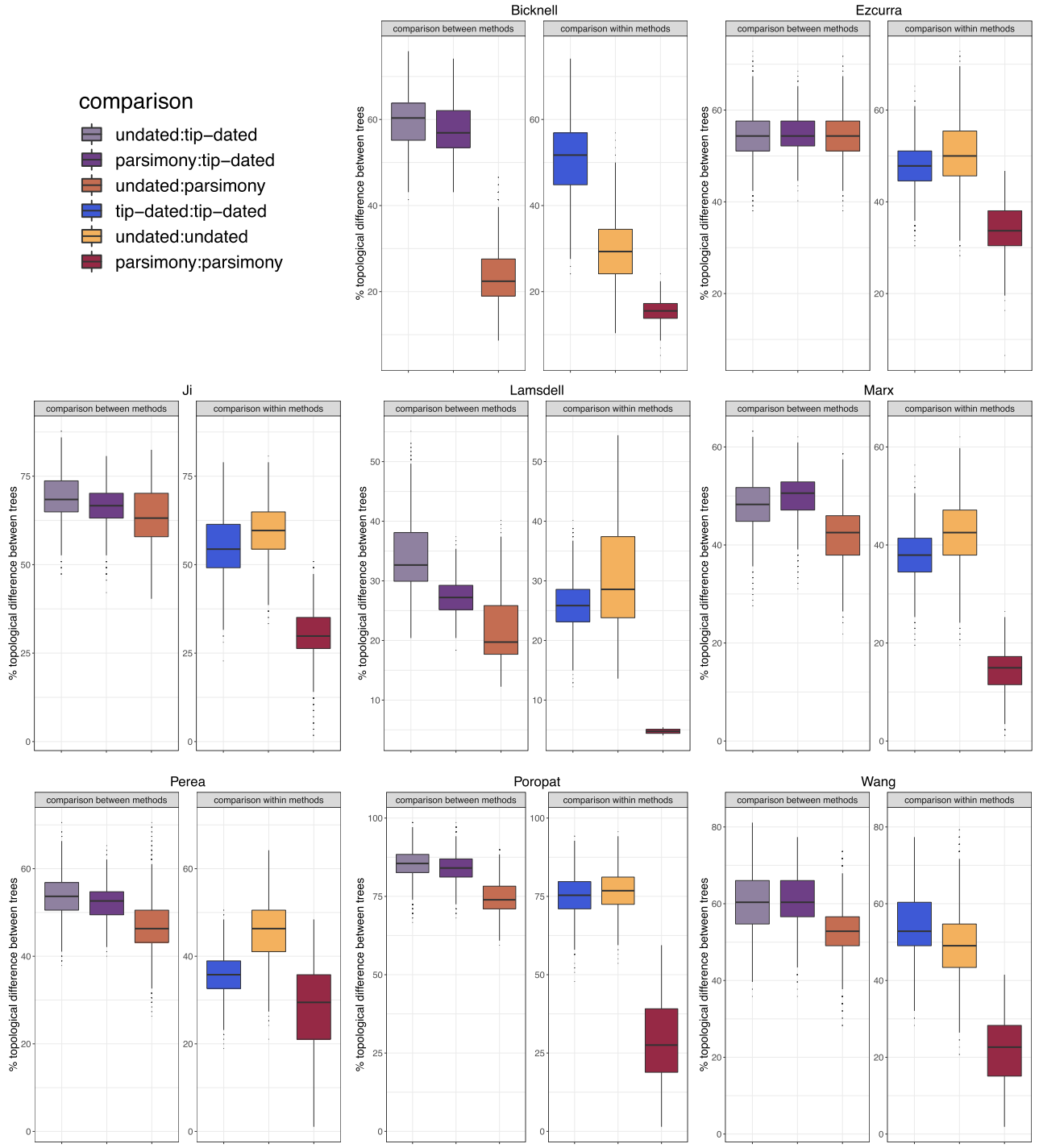

FIGURE 1. Tree topologies resulting from Bayesian tip-dated methods are outliers as compared to trees resulting from undated Bayesian and parsimony methods. % topological difference (rescaled Robinson–Foulds distance) is plotted for each comparison, across seven data sets. Every tree from a subsample of the posterior or set of shortest trees is compared to the sample from an alternative method, and the resulting range of values shown as a boxplot. Tree samples from each method are also compared to themselves as a measure of resolution.

a small but statistically significant difference favoring parsimony based on many data sets (although results vary within particular data sets).

Trees with a higher stratigraphic fit are assigned higher tree model likelihood in tip dating. Figure 4 shows plots of the stratigraphic fit against tree model likelihood for the distribution of trees produced from a tip-dated analysis (without morphological data) for each data set. The GER shows significant positive correlations with tree model likelihood across all data sets, whereas the

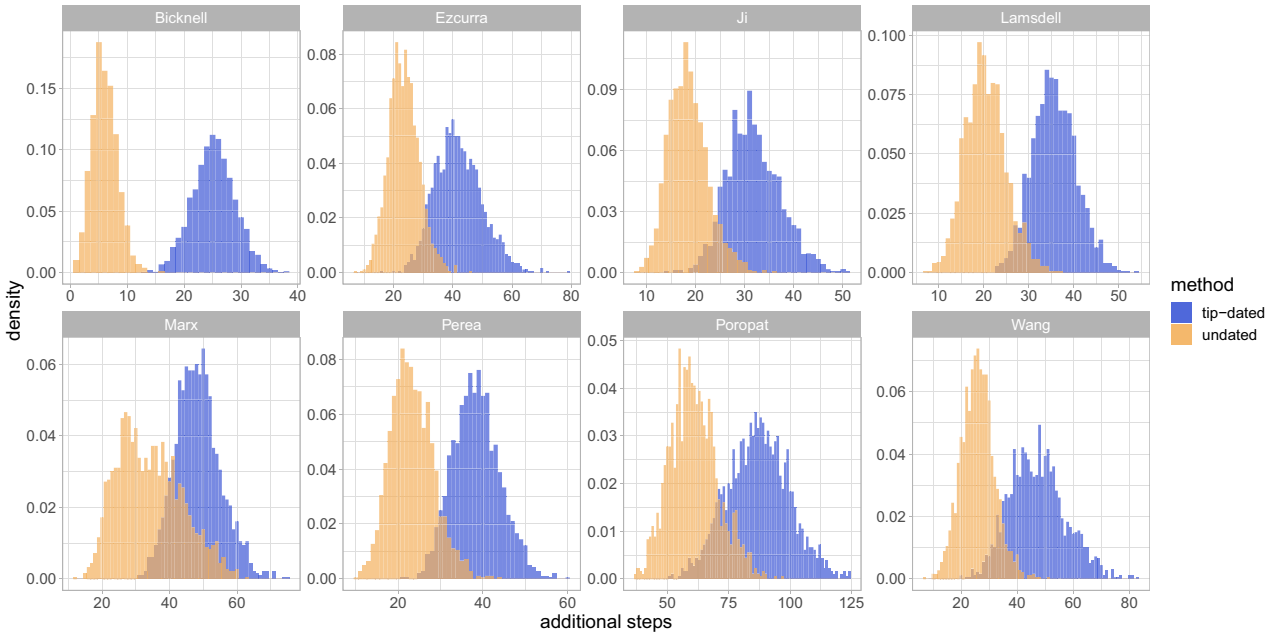

FIGURE 2. Bayesian tip dating produces trees that are less parsimonious than trees from undated Bayesian analysis. Histograms of the number of additional steps required by the posterior sample of trees produced by tip dating and undated Bayesian analysis, compared to the tree length of the most parsimonious trees.

SCI shows significant positive correlations for six of the eight data sets.

Sampling Distribution

Diversification dynamics affect the extent to which the tree model favors trees with good stratigraphic fit. The whales data set of Marx and Fordyce (2015) represents an outlier when the difference between the stratigraphic fit of tip-dated trees analyzed with and without morphological data is considered (Fig. 3). For this data set, the estimated GER values for tip dating without morphological data are on average lower than for tip dating with morphological data. In contrast to the other data sets, the frequency of sampled taxa in this data set increases through time (Supplementary Fig. S1A available on Dryad at http://dx.doi.org/10.5061/dryad.xksn02vc9), a pattern that is still in evidence when extant taxa are excluded (Supplementary Fig. S1B available on Dryad). The estimated net diversification rate is therefore higher for this data set than any other (Supplementary Fig. S1C available on Dryad). The effect of net diversification rate on the stratigraphic fit of trees drawn from the tree model distribution is clear in analyses using randomized taxon ages. Tip ages sampled from an exponential distribution, representing positive net diversification through time (assuming a constant fossil recovery rate), result in trees with lower GER values than trees resulting from tip ages sampled from a uniform distribution (Fig. 5).

Effect of Incomplete Taxa

Successive deletion of incomplete taxa from the Poropat et al. (2016) and Ezcurra (2016) data sets results in an increase in stratigraphic fit of trees from undated Bayesian analyses, whereas the stratigraphic fit of trees from tip dating are essentially unchanged (Fig. 6). Topological differences between the methods generally decline with each deletion. Random deletion of taxa does not lead to changes in stratigraphic fit for either method and topological differences do not change, showing that the observed patterns are due to the deletion of incomplete taxa, not merely a result of deletion of taxa in general. Parsimony analyses of the same data sets, led to congruent, although somewhat less consistent results (Supplementary Fig. S2 available on Dryad).

DISCUSSION

Differences in Tree Topology Between Different Methods

Tip dating has received relatively little attention during debates over which phylogenetic methods are most suitable for analyzing morphology-only data sets (Wright and Hillis 2014; O'Reilly et al. 2016; Goloboff et al. 2017, 2018a, 2018b; O'Reilly et al. 2018; Sansom et al. 2018; Smith 2019; Keating et al. 2020); these debates have focused on the merits of parsimony or undated Bayesian methods. However, the results presented here show that tree topologies produced by tip dating are outliers as compared to trees produced by parsimony and undated Bayesian methods, a result also found for one data set in Turner et al. (2017). Studies on empirical problems have already shown that use of tip dating can produce radically different tree topologies from other methods (King et al. 2017; Lee and Yates 2018).

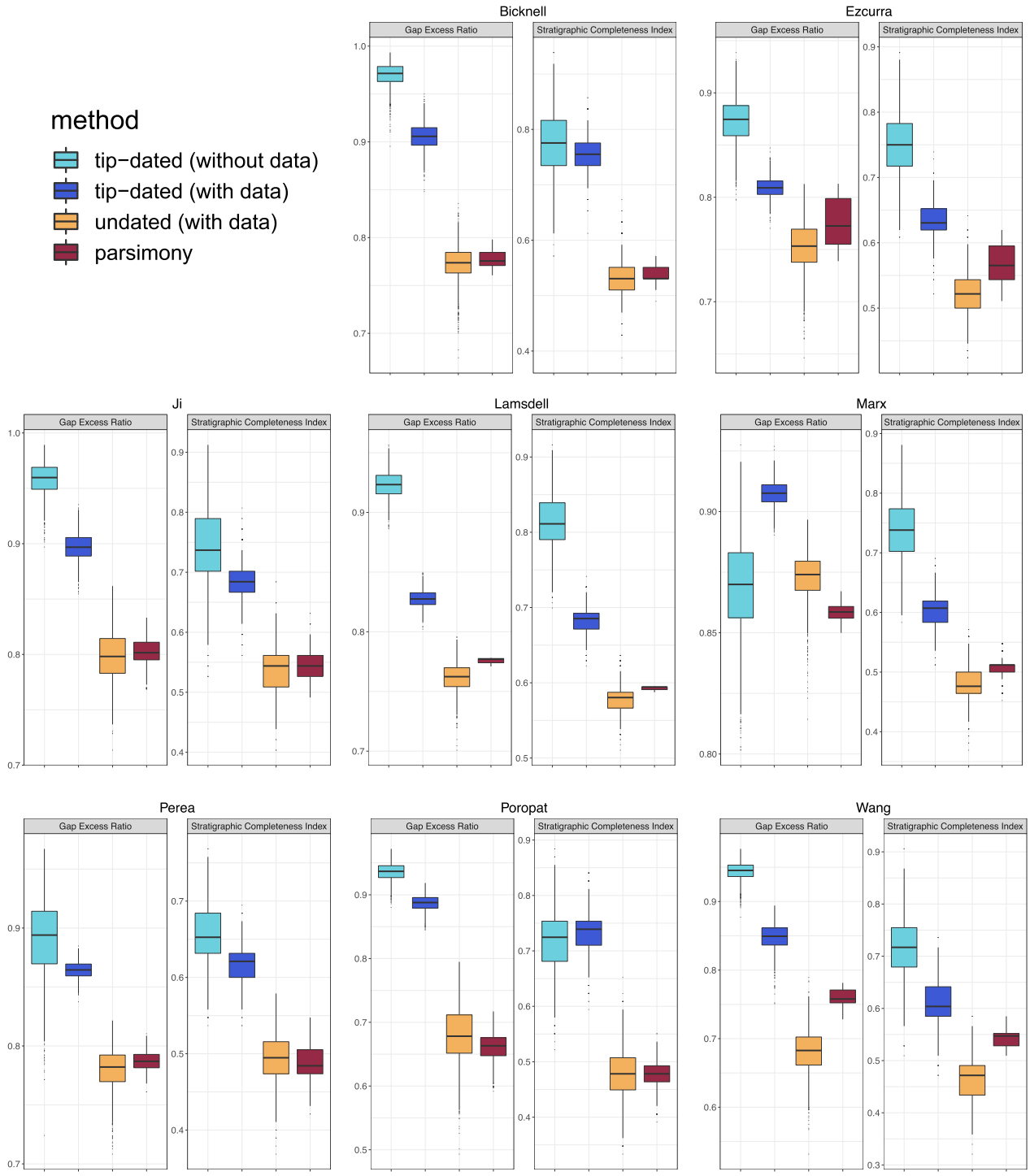

FIGURE 3. Bayesian tip-dated methods recover trees with better fit to stratigraphy than other methods. Gap Excess Ratio and Stratigraphic Completeness Index for every tree in each posterior sample (Bayesian analysis) or MPTs (parsimony) is shown as a box plot.

Much of the debate between parsimony and undated Bayesian methods has centered on consensus trees and their degree of resolution (O'Reilly et al. 2016; Brown et al. 2017; Goloboff et al. 2017). My results comparing topological differences within methods show that parsimony MPTs are more topologically similar than Bayesian posterior samples (Fig. 1), aligning with previously research showing consensus trees from parsimony to be more resolved than those from Bayesian analysis (O'Reilly et al. 2016). Differences between trees within the posterior sample from Bayesian tip-dated and undated analyses are in general approximately

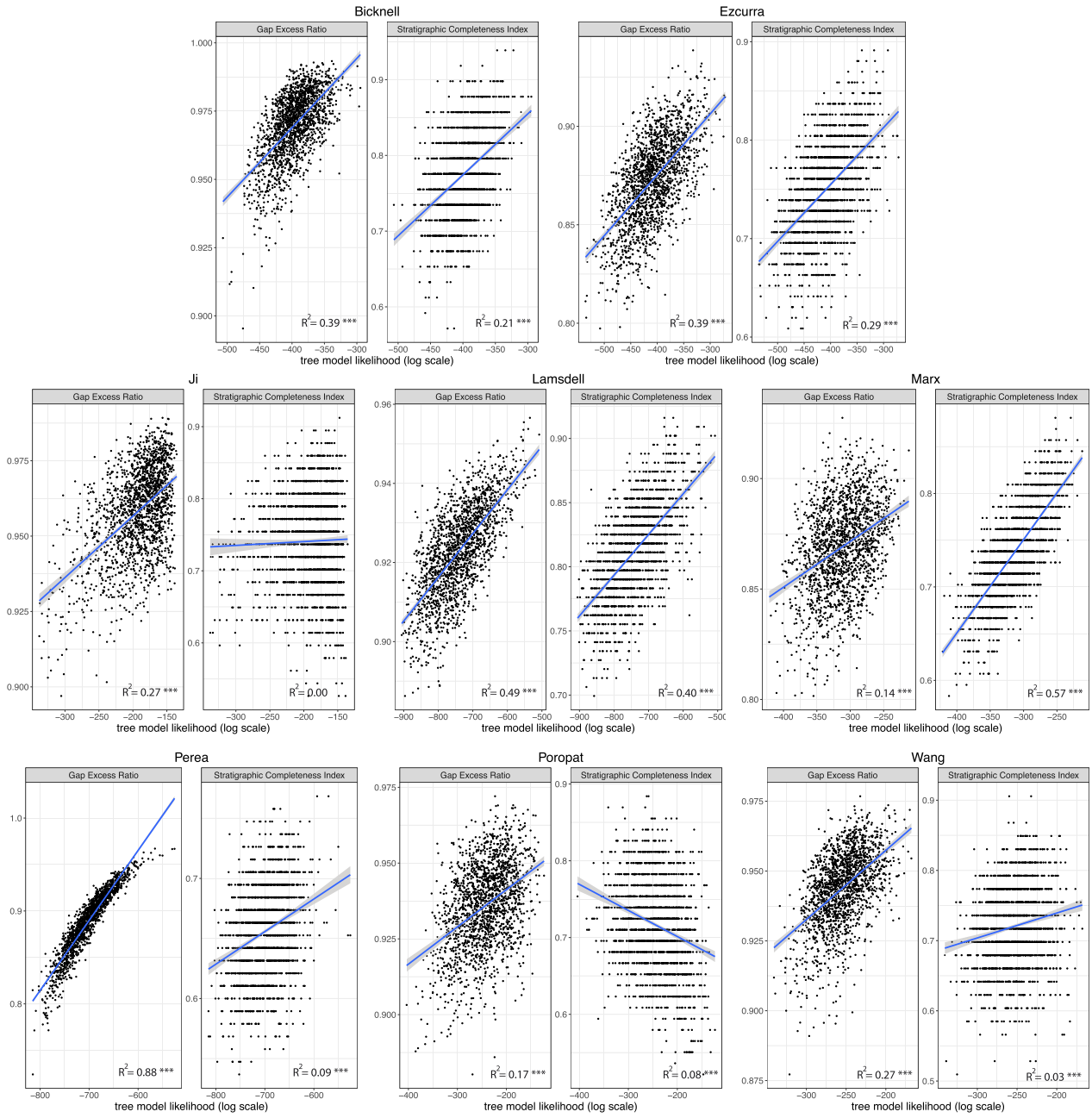

FIGURE 4. Trees with better stratigraphic fit are assigned higher tree model likelihood. Each data point represents a tree from a tip-dated analysis without morphological data. A significant positive correlation exists between Gap Excess Ratio and tree model likelihood for all eight data sets, and between Stratigraphic Completeness Index and tree model likelihood for six of eight data sets. Stars denote significance level (***$P < 0.001$).

equal, with the results varying between data sets (Fig. 1).

The results presented here suggest that the incorporation of stratigraphic data into a tip-dated analysis has more effect on tree topology than any other feature of a Bayesian analysis. Topological differences between undated Bayesian and parsimony methods are mainly driven by a bias against characters with incomplete coding under Bayesian analysis, whereby incompletely coded characters have a negligible effect on tree topology (Simmons 2012, 2014; King 2019). Tip-dated analyses are likely also affected by this bias, and also share with undated Bayesian analyses the use of gamma-distributed among-character rate variation (Yang 1996), which may affect tree topology (Lee and Worthy 2011), and the effective up-weighting of characters with a higher number of states (King et al. 2017). Although the bias against incompletely coded

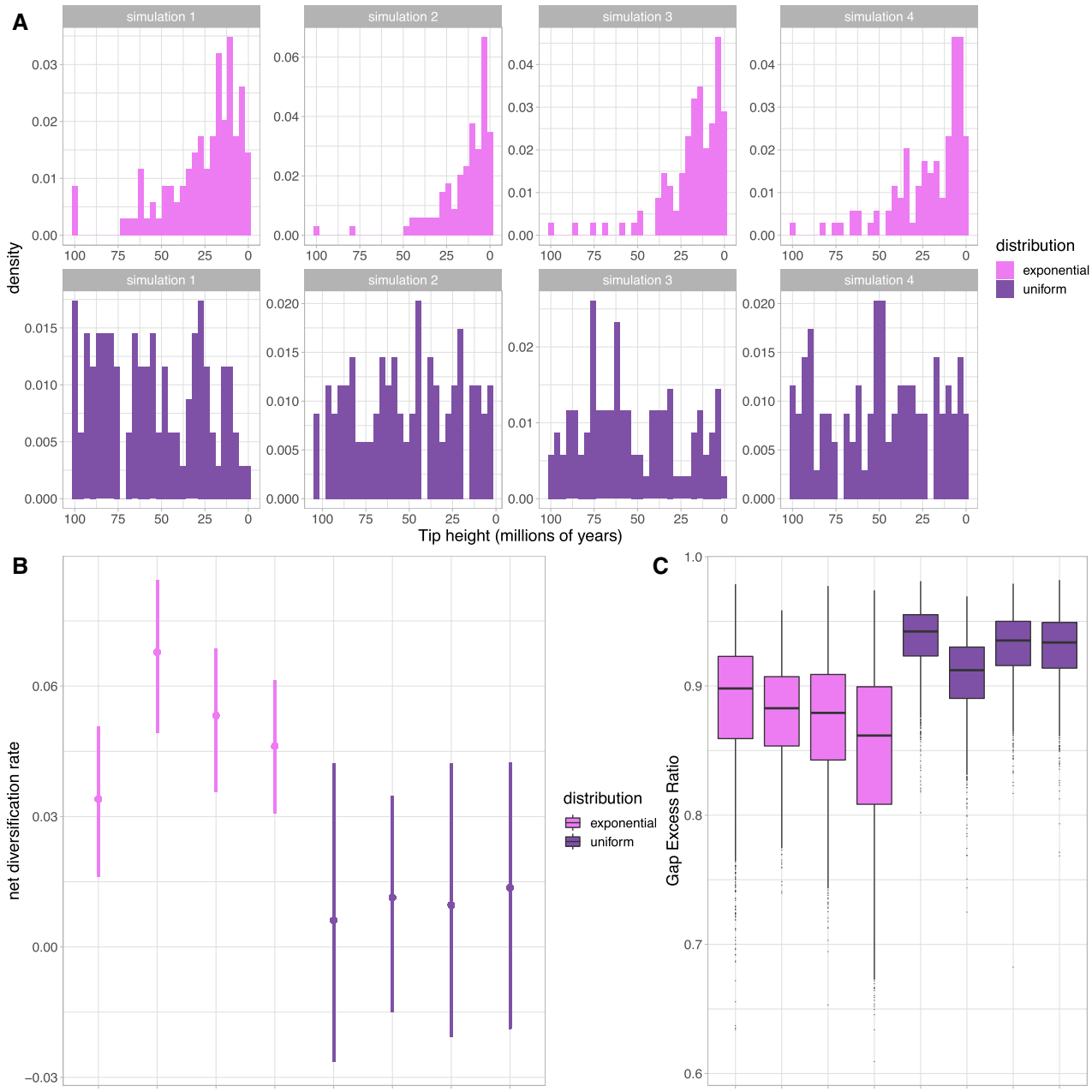

FIGURE 5. Diversification dynamics affect the extent to which trees with a high Gap Excess Ratio are favored by the tree model. a) Histograms of age distributions for taxa sampled from an exponential distribution (top) and uniform distribution (bottom). b) Net diversification rate estimates (calculated as birth rate–death rate in each sample from a tip-dated analysis) when tip ages are drawn from an exponential distribution are higher than when tip ages are drawn from a uniform distribution. c) Gap Excess Ratios for trees resulting from a tip-dated analysis of taxa with ages sampled from a uniform distribution are higher than for trees resulting from a tip-dated analysis of taxa with ages sampled from an exponential distribution.

characters in Bayesian analysis can lead to strongly supported topology differences from parsimony (King 2019), this is likely to be uncommon, only occurring when characters with different levels of coding completeness support conflicting phylogenetic relationships.

When will Stratigraphic Data Overrule Morphological Data?

Both the tree model and the morphological data influence tree topologies in the posterior sample resulting from tip dating. Topologies with a better stratigraphic fit (e.g., requiring shorter ghost lineages) have a higher tree model likelihood (Fig. 4). The stratigraphic fit of the posterior distribution of tree topologies is intermediate between the prior and the values from undated Bayesian and parsimony approaches (Fig. 3), reflecting the interplay of the tree model and evidence from the morphological character data.

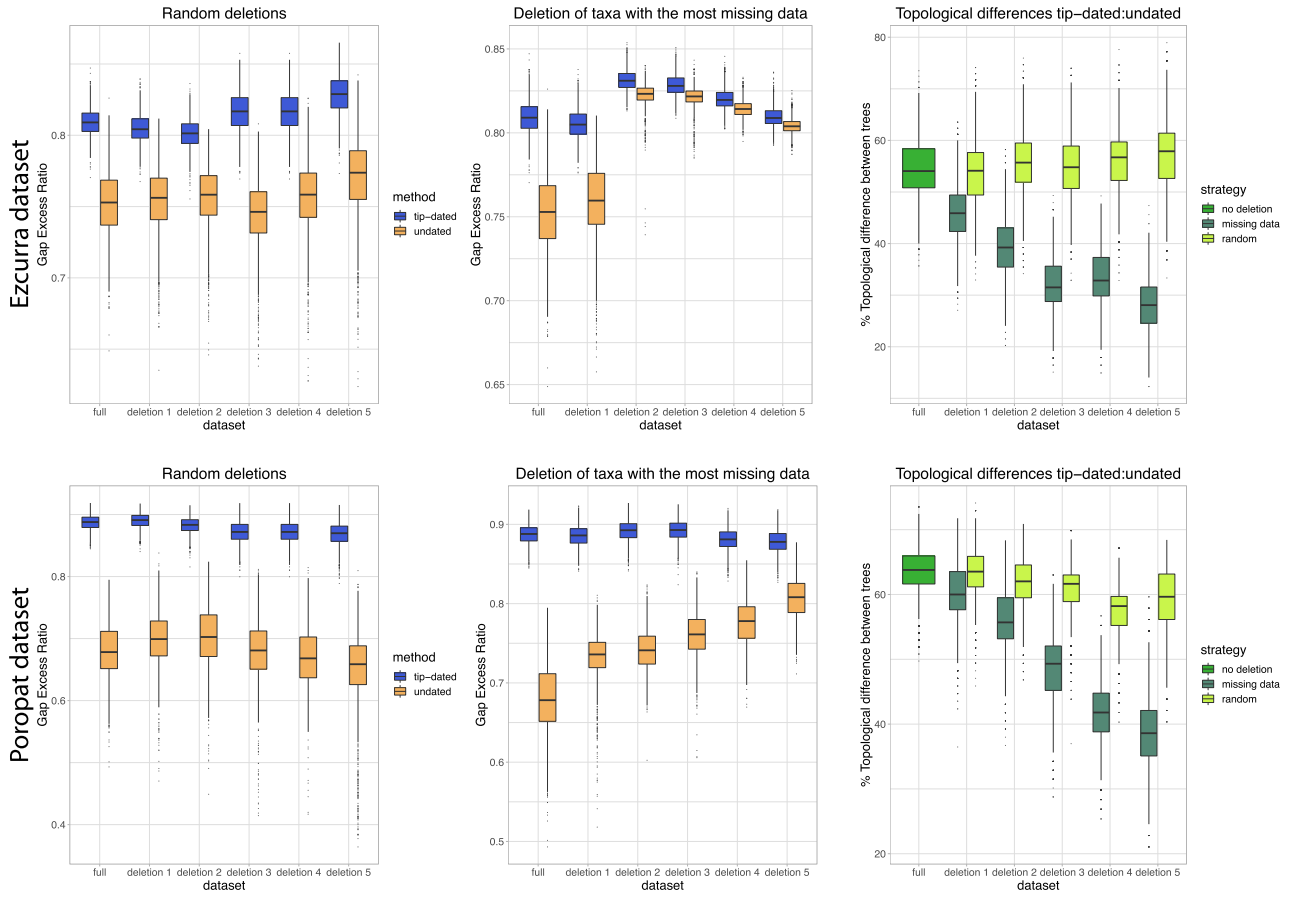

FIGURE 6. Differences between tip-dated and undated Bayesian methods in terms of topology and stratigraphic fit are concentrated in uncertain parts of the phylogeny. This figure shows the effect of taxon deletion on the stratigraphic fit of Bayesian tip-dated and undated trees, and the topological difference between these methods. Successive deletion if random taxa (left) does not result in changes in the stratigraphic fit of trees under either method. Successive deletion of the most incomplete taxa (center) leads to an increase in stratigraphic fit for undated analysis, but relatively little difference for tip-dated trees. Topological differences between these methods also successively decrease when taxa with the most missing data are deleted (right), but not when random taxa are deleted.

The topological differences between tip dating and other phylogenetic methods is driven by the placement of incomplete taxa (Fig. 6). In undated analyses (under either parsimony or model-based optimality criteria), these incomplete taxa may fit into several positions on the tree even if these are inconsistent with their age, but temporally inconsistent positions require stronger morphological evidence in the tip-dated approach. These results suggest that, as morphological data for incomplete taxa improve, topologies recovered from undated Bayesian or parsimony approaches might become more similar to those produced by tip dating. The results of Benton and Storrs (1994), showing a small improvement in stratigraphic fit of trees for various groups following the accumulation of data over 25 years, provides some evidence in support of this idea.

The results support the view that inclusion of stratigraphic age data in tip-dated Bayesian analysis directly affects tree topology. Highly incomplete taxa are more constrained to positions congruent with their age in tip-dated analysis, even if the (limited) character data are also consistent with other, less temporally congruent positions. Conversely, taxa for which abundant data are available can be placed in temporally incongruent positions if there is sufficient character support. This is intuitive, and in some respect resembles the approach already informally taken by most paleontologists: extraordinary claims require extraordinary evidence.

It should be noted that the tree model in tip-dated analyses does not necessarily "maximize" stratigraphic fit measures. High values approaching 1 for the GER are only possible for highly unbalanced trees (Wills et al. 2008), but such tree shapes are only favored by the FBD model in certain circumstances. Whereas the GER measures the length of required ghost lineages, the FBD model finds trees with approximately constant rates of speciation and extinction. These two factors align best when net diversification is low, leading to approximately constant diversity through time (Fig. 5). However, when diversity increases through time (positive net diversification), minimizing ghost lineages would produce a tree with many nodes near the

present, implying rates of speciation increase through time. Such a tree would violate the assumptions of the FBD model; instead, the tree model would tend to favor trees with longer ghost lineages but more constant rates of diversification. The degree to which stratigraphic fit metrics are "maximized" by the FBD model therefore varies across data sets, but regardless of this variation, stratigraphic fit is higher for tip-dated trees compared with trees from other methods (Fig. 2).

Is the Inclusion of Stratigraphic Data in Phylogenetic Analysis Justified?

"It is merely eccentric to claim that time is not a desirable parameter in working out phylogenies, even though what is desirable is not invariably available or clear in its significance" (Simpson 1975).

The fossil record need only be adequate, not perfect, to justify its utilization (Paul 1985). The high level of congruence between phylogenetic and stratigraphic data (Gauthier et al. 1988; Norell and Novacek 1992; Benton and Hitchin 1997) shows that stratigraphic data is generally informative regarding clade age and branching order. Including both stratigraphic and morphological data when estimating phylogeny allows these two generally congruent data types to reciprocally illuminate in the minority of occasions in which they disagree.

Stratigraphic data will always be incomplete (in the sense that not every species will be present in the fossil record), but the same can be said of morphological data. No fossil preserves 100% of morphological data (even a fossil being scored for 100% of characters within a particular morphological data matrix is rare). In addition, convergent evolution means that morphological data can sometimes be misleading, a fact becoming increasingly apparent with the availability of increasing amounts of molecular data (Lee and Palci 2015). Stratigraphic data provides an opportunity to identify probable convergent evolution in clades widely separated by time, a point that has already been demonstrated in tip-dated phylogenetic analyses of paleognath birds (Worthy et al. 2017) and gharials (Lee and Yates 2018).

Stratigraphic and morphological data represent a way of assessing the quality of each other. Stratigraphic incongruence cannot be recognized independently from morphological data, but then how do we know that the morphological data are correct? In the absence of molecular data, stratigraphic data are an important independent assessment of the quality of morphological data. Simultaneous analysis of both forms of data is therefore a good approach and should avoid the recovery of highly temporally incongruent trees based on scant evidence. However, morphological data have generally been afforded precedence. For example, Turner et al. (2017) suggest, based on analysis of a crocodylomorphs data set, that tip dating should not be used in cases where sampling is uneven, leading to long un-sampled branches which are then disfavored by tip dating. However, their recognition of uneven sampling was not independent from phylogeny and was based on a time-scaled parsimony tree of the same data set. On the contrary, an expectation that long un-sampled branches should be rare is a reasonable expectation, one that requires solid data to overturn. This expectation is particularly relevant for groups with a good stratigraphic record. However, even in groups with a poor stratigraphic record, *relatively* long un-sampled branches should still be unexpected.

The results presented here show the effect of FBD models on tree topology, which is particularly important in parts of the tree with weak character support. This study only investigated a time-homogenous FBD model: further research should test the effect of time heterogeneous models, which can accommodate heterogeneity through time slices (Stadler et al. 2013; Gavryushkina et al. 2014) or amongst lineages (Barido-Sottani et al. 2020).

Supplementary Material

Data available from the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.xksn02vc9.

Acknowledgments

I thank Robin Beck, Mike Lee, John Long, and three anonymous referees for comments on the manuscript, and Graeme Lloyd for discussions. Martin Rücklin supported the completion of this paper. Tnt is made freely available through the Willi Hennig society.

References

Alroy J. 2002. Stratigraphy in phylogeny reconstruction—reply to Smith (2000). J. Paleontol. 76:587–589.

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

Barido-Sottani J., Vaughan T.G., Stadler T. 2020. A multitype birth-death model for Bayesian inference of lineage-specific birth and death rates. Syst. Biol. syaa016.

Bell M.A., Lloyd G.T. 2015. strap: an R package for plotting phylogenies against stratigraphy and assessing their stratigraphic congruence. Palaeontology. 58:379–389.

Benton M.J., Hitchin R. 1997. Congruence between phylogenetic and stratigraphic data on the history of life. Proc. R. Soc. Lond. B Biol. Sci. 264:885–890.

Benton M.J., Storrs G.W. 1994. Testing the quality of the fossil record: paleontological knowledge is improving. Geology. 22:111–114.

Benton M.J., Wills M., Hitchin R. 2000. Quality of the fossil record through time. Nature. 403:534–537.

Bicknell R.D., Pates S. 2019. Xiphosurid from the Tournaisian (Carboniferous) of Scotland confirms deep origin of Limuloidea. Sci. Rep. 9:1–13.

Bouckaert R., Vaughan T.G., Barido-Sottani J., Duchêne S., Fourment M., Gavryushkina A., Heled J., Jones G., Kühnert D., De Maio N., Matschiner M., Mendes F.K., Müller N.F., Ogilvie H.A., du Plessis L., Popinga A., Rambaut A., Rasmussen D., Siveroni I.,

Suchard M.A., Wu C.-H., Xie D., Zhang C., Stadler T., Drummond A.J. 2019. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. PLoS Comput. Biol. 15:e1006650.
Brown J.W., Parins-Fukuchi C., Stull G.W., Vargas O.M., Smith S.A. 2017. Bayesian and likelihood phylogenetic reconstructions of morphological traits are not discordant when taking uncertainty into consideration: a comment on Puttick *et al*. Proc. R. Soc. B 284:20170986.
Close R.A., Friedman M., Lloyd G.T., Benson R.B. 2015. Evidence for a mid-Jurassic adaptive radiation in mammals. Curr. Biol. 25:2137–2142.
Close R.A., Johanson Z., Tyler J.C., Harrington R.C., Friedman M. 2016. Mosaicism in a new Eocene pufferfish highlights rapid morphological innovation near the origin of crown tetraodontiforms. Palaeontology. 59:499–514.
Drummond A.J., Ho S.Y., Phillips M.J., Rambaut A. 2006. Relaxed phylogenetics and dating with confidence. PLoS Biol. 4:e88.
Drummond A.J., Stadler T. 2016. Bayesian phylogenetic estimation of fossil ages. Philos. Trans. R. Soc. B. 371:20150129.
Ezcurra M.D. 2016. The phylogenetic relationships of basal archosauromorphs, with an emphasis on the systematics of proterosuchian archosauriforms. PeerJ. 4:e1778.
Fisher D.C. 1994. Stratocladistics: morphological and temporal patterns and their relation to phylogenetic process. In: Grand L., Rieppel O., editors. Interpreting the hierarchy of nature. San Diego: Academic Press, p. 133–171.
Fisher D.C. 2008. Stratocladistics: integrating temporal data and character data in phylogenetic inference. Annu. Rev. Ecol. Evol. Syst. 39:365–385.
Fox D.L., Fisher D.C., Leighton L.R. 1999. Reconstructing phylogeny with and without temporal data. Science. 284:1816–1819.
Gauthier J., Kluge A.G., Rowe T. 1988. Amniote phylogeny and the importance of fossils. Cladistics. 4:105–209.
Gavryushkina A., Heath T.A., Ksepka D.T., Stadler T., Welch D., Drummond A.J. 2017. Bayesian total evidence dating reveals the recent crown radiation of penguins. Syst. Biol. 66:57–73.
Gavryushkina A., Welch D., Stadler T., Drummond A.J. 2014. Bayesian inference of sampled ancestor trees for epidemiology and fossil calibration. PLoS Comput. Biol. 10:e1003919.
Geiger D.L., Fitzhugh K., Thacker C.E. 2001. Timeless characters: a response to Vermeij (1999). Paleobiology. 27:177–178.
Goloboff P.A., Catalano S.A. 2016. TNT version 1.5, including a full implementation of phylogenetic morphometrics. Cladistics. 32:221–238.
Goloboff P.A., Pittman M., Pol D., Xu X. 2018a. Morphological data sets fit a common mechanism much more poorly than DNA sequences and call into question the Mkv model. Syst. Biol. 68:494–504.
Goloboff P.A., Torres A., Arias J.S. 2017. Weighted parsimony outperforms other methods of phylogenetic inference under models appropriate for morphology. Cladistics. 34:407–437.
Goloboff P.A., Torres Galvis A., Arias J.S. 2018b. Parsimony and model-based phylogenetic methods for morphological data: comments on O'Reilly *et al*. Palaeontology. 61:625–630.
Heath T.A., Huelsenbeck J.P., Stadler T. 2014. The fossilized birth-death process for coherent calibration of divergence-time estimates. Proc. Natl. Acad. Sci. USA. 111:E2957–E2966.
Heled J., Drummond A.J. 2012. Calibrated tree priors for relaxed phylogenetics and divergence time estimation. Syst. Biol. 61:138–149.
Heyning J.E., Thacker C. 1999. Phylogenies, temporal data, and negative evidence. Science. 285:1179–1179.
Huelsenbeck J.P. 1994. Comparing the stratigraphic record to estimates of phylogeny. Paleobiology. 20:470–483.
Huelsenbeck J.P., Ronquist F. 2001. MRBAYES: Bayesian inference of phylogeny. Bioinformatics. 17:754–755.
Ji C., Jiang D.-Y., Motani R., Rieppel O., Hao W.-C., Sun Z.-Y. 2016. Phylogeny of the Ichthyopterygia incorporating recent discoveries from South China. J. Vertebr. Paleontol. 36:e1025956.
Keating J.N., Sansom R.S., Sutton M.D., Knight C.G., Garwood R.J. 2020. Morphological phylogenetics evaluated using novel evolutionary simulations. Syst. Biol. syaa012.
King B. 2019. Which morphological characters are influential in a Bayesian phylogenetic analysis? Examples from the earliest osteichthyans. Biol. Lett. 15:20190288.
King B., Qiao T., Lee M.S., Zhu M., Long J.A. 2017. Bayesian morphological clock methods resurrect placoderm monophyly and reveal rapid early evolution in jawed vertebrates. Syst. Biol. 66:499–516.
Lamsdell J.C., Selden P.A. 2017. From success to persistence: Identifying an evolutionary regime shift in the diverse Paleozoic aquatic arthropod group Eurypterida, driven by the Devonian biotic crisis. Evolution. 71:95–110.
Lee M.S., Cau A., Naish D., Dyke G.J. 2014a. Sustained miniaturization and anatomical innovation in the dinosaurian ancestors of birds. Science. 345:562–566.
Lee M.S.Y., Cau A., Naish D., Dyke G.J. 2014b. Morphological clocks in paleontology, and a mid-Cretaceous origin of crown Aves. Syst. Biol. 63:442–449.
Lee M.S.Y., Palci A. 2015. Morphological phylogenetics in the genomic age. Curr. Biol. 25:R922–R929.
Lee M.S.Y., Worthy T.H. 2011. Likelihood reinstates *Archaeopteryx* as a primitive bird. Biol. Lett. 8:299–303.
Lee M.S.Y., Yates A.M. 2018. Tip-dating and homoplasy: reconciling the shallow molecular divergences of modern gharials with their long fossil record. Proc. R. Soc. B. 285:20181071.
Lewis P.O. 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. Syst. Biol. 50:913–925.
Lockwood C. 1998. Stratigraphy and phylogeny. Evol. Anthropol. 6:153–154.
Maddison W.P., Maddison D.R. 2015. Mesquite: a modular system for evolutionary analysis. Version 3.04. Available from: http://mesquiteproject.org
Marcot J.D., Fox D.L. 2008. StrataPhy: a new computer program for stratocladistic analysis. Palaeontol. Electron. 11:5A.
Marx F.G., Fordyce R.E. 2015. Baleen boom and bust: a synthesis of mysticete phylogeny, diversity and disparity. Roy. Soc. Open Sci. 2:140434.
Matzke N.J., Wright A. 2016. Inferring node dates from tip dates in fossil *Canidae*: the importance of tree priors. Biol. Lett. 12:20160328.
Norell M.A., Novacek M.J. 1992. The fossil record and evolution: comparing cladistic and paleontologic evidence for vertebrate history. Science. 255:1690–1693.
O'Connor A., Wills M.A. 2016. Measuring stratigraphic congruence across trees, higher taxa, and time. Syst. Biol. 65:792–811.
O'Reilly J.E., Puttick M.N., Parry L., Tanner A.R., Tarver J.E., Fleming J., Pisani D., Donoghue P.C. 2016. Bayesian methods outperform parsimony but at the expense of precision in the estimation of phylogeny from discrete morphological data. Biol. Lett. 12:20160081.
O'Reilly J.E., Puttick M.N., Pisani D., Donoghue P.C. 2018. Empirical realism of simulated data is more important than the model used to generate it: a reply to Goloboff et al. Palaeontology. 61:631–635.
Paul C.R.C. 1985. The adequacy of the fossil record revisited. Spec. Papers Palaeontol. 33:7–15.
Perea D., Soto M., Sterli J., Mesa V., Toriño P., Roland G., Da Silva J. 2014. *Tacuarembemys kusterae*, gen. et sp. nov., a new late Jurassic–?Earliest Cretaceous continental turtle from western Gondwana. J. Vertebr. Paleontol. 34:1329–1341.
Poropat S.F., Mannion P.D., Upchurch P., Hocknull S.A., Kear B.P., Kundrát M., Tischler T.R., Sloan T., Sinapius G.H.K., Elliott J.A. 2016. New Australian sauropods shed light on Cretaceous dinosaur palaeobiogeography. Sci. Rep. 6:34467.
R Core Team. 2018. R: a language and environment for statistical computing. Vienna (Austria): R foundation for Statistical Computing. Available from: https://www.R-project.org/
Rambaut A., Drummond A.J., Xie D., Baele G., Suchard M.A. 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Syst. Biol. 67:901–904.
Rannala B., Zhu T., Yang Z. 2012. Tail paradox, partial identifiability, and influential priors in Bayesian branch length inference. Mol. Biol. Evol. 29:325–335.
Robinson D.F., Foulds L.R. 1981. Comparison of phylogenetic trees. Math. Biosci. 53:131–147.
Ronquist F., Huelsenbeck J.P. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics. 19:1572–1574.
Ronquist F., Klopfstein S., Vilhelmsen L., Schulmeister S., Murray D.L., Rasnitsyn A.P. 2012a. A total-evidence approach to dating with

fossils, applied to the early radiation of the Hymenoptera. Syst. Biol. 61:973–999.
Ronquist F., Teslenko M., van der Mark P., Ayres D.L., Darling A., Höhna S., Larget B., Liu L., Suchard M.A., Huelsenbeck J.P. 2012b. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. 61:539–542.
Sansom R.S., Choate P.G., Keating J.N., Randle E. 2018. Parsimony, not Bayesian analysis, recovers more stratigraphically congruent phylogenetic trees. Biol. Lett. 14:20180263.
Schliep K.P. 2011. phangorn: phylogenetic analysis in R. Bioinformatics. 27:592–593.
Siddall M.E. 1998. Forum. stratigraphic fit to phylogenies: a proposed solution. Cladistics. 14:201–208.
Simmons M.P. 2012. Radical instability and spurious branch support by likelihood when applied to matrices with non-random distributions of missing data. Mol. Phylogenet. Evol. 62:472–484.
Simmons M.P. 2014. A confounding effect of missing data on character conflict in maximum likelihood and Bayesian MCMC phylogenetic analyses. Mol. Phylogenet. Evol. 80:267–280.
Simpson G.G. 1975. Recent advances in methods of phylogenetic inference. In: Luckett W.P., Szalay F.S., editors. Phylogeny of the primates. New York: Springer, p. 3–19.
Smith A.B. 1998. Is the fossil record adequate. Nature. doi:10.1038/nature28134
Smith A.B. 2000. Stratigraphy in phylogeny reconstruction. J. Paleontol. 74:763–766.
Smith M.R. 2019. Bayesian and parsimony approaches reconstruct informative trees from simulated morphological datasets. Biol. Lett. 15:20180632.
Stadler T. 2010. Sampling-through-time in birth-death trees. J. Theor. Biol. 267:396–404.
Stadler T., Kuhnert D., Bonhoeffer S., Drummond A.J. 2013. Birth-death skyline plot reveals temporal changes of epidemic spread in HIV and hepatitis C virus (HCV). Proc. Natl. Acad. Sci. USA. 110:228–233.
Turner A.H., Pritchard A.C., Matzke N.J. 2017. Empirical and Bayesian approaches to fossil-only divergence times: a study across three reptile clades. PLoS One. 12:e0169885.
Wagner P.J. 1995. Stratigraphic tests of cladistic hypotheses. Paleobiology. 21:153–178.
Wagner P.J. 1998. A likelihood approach for evaluating estimates of phylogenetic relationships among fossil taxa. Paleobiology. 24:430–449.
Wang M., Zheng X., O'Connor J.K., Lloyd G.T., Wang X., Wang Y., Zhang X., Zhou Z. 2015. The oldest record of Ornithuromorpha from the early cretaceous of China. Nat. Commun. 6:6987.
Warren D.L., Geneva A.J., Lanfear R. 2017. RWTY (R We There Yet): An R package for examining convergence of Bayesian phylogenetic analyses. Mol. Biol. Evol. 34:1016–1020.
Wills M.A. 1999. Congruence between phylogeny and stratigraphy: randomization tests and the gap excess ratio. Syst. Biol. 48:559–580.
Wills M.A., Barrett P.M., Heathcote J.F. 2008. The modified gap excess ratio (GER*) and the stratigraphic congruence of dinosaur phylogenies. Syst. Biol. 57:891–904.
Worthy T.H., Degrange F.J., Handley W.D., Lee M.S. 2017. The evolution of giant flightless birds and novel phylogenetic relationships for extinct fowl (Aves, Galloanseres). Roy. Soc. Open Sci. 4:170975.
Wright A.M., Hillis D.M. 2014. Bayesian analysis using a simple likelihood model outperforms parsimony for estimation of phylogeny from discrete morphological data. PLoS One. 9:e109210.
Yang Z. 1996. Among-site rate variation and its impact on phylogenetic analyses. Trends Ecol. Evol. 11:367–372.