

# Empirical and Methodological Challenges to the Model-Based Inference of Diversification Rates in Extinct Clades

DAVID ČERNÝ<sup>1,\*</sup>, DANIEL MADZIA<sup>2</sup>, AND GRAHAM J. SLATER<sup>1</sup>

<sup>1</sup>Department of the Geophysical Sciences, University of Chicago, 5734 South Ellis Avenue, Chicago, IL 60637, USA; and

<sup>2</sup>Institute of Paleobiology, Polish Academy of Sciences, Twarda 51/55, Warsaw 00-818, Poland

\*Correspondence to be sent to: Department of the Geophysical Sciences, University of Chicago, Chicago, IL 60637, USA;  
E-mail: david.cerny1@gmail.com.

Received 28 October 2020; reviews returned 2 June 2021; accepted 9 June 2021

Associate Editor: Matt Friedman

**Abstract.**—Changes in speciation and extinction rates are key to the dynamics of clade diversification, but attempts to infer them from phylogenies of extant species face challenges. Methods capable of synthesizing information from extant and fossil species have yielded novel insights into diversification rate variation through time, but little is known about their behavior when analyzing entirely extinct clades. Here, we use empirical and simulated data to assess how two popular methods, PyRate and Fossil BMM, perform in this setting. We inferred the first tip-dated trees for ornithischian dinosaurs and combined them with fossil occurrence data to test whether the clade underwent an end-Cretaceous decline. We then simulated phylogenies and fossil records under empirical constraints to determine whether macroevolutionary and preservation rates can be teased apart under paleobiologically realistic conditions. We obtained discordant inferences about ornithischian macroevolution including a long-term speciation rate decline (BMM), mostly flat rates with a steep diversification drop (PyRate) or without one (BMM), and episodes of implausibly accelerated speciation and extinction (PyRate). Simulations revealed little to no conflation between speciation and preservation, but yielded spuriously correlated speciation and extinction estimates while time-smearing tree-wide shifts (BMM) or overestimating their number (PyRate). Our results indicate that the small phylogenetic data sets available to vertebrate paleontologists and the assumptions made by current model-based methods combine to yield potentially unreliable inferences about the diversification of extinct clades. We provide guidelines for interpreting the results of the existing approaches in light of their limitations and suggest how the latter may be mitigated. [BMM; diversification; fossils; macroevolutionary rates; Ornithischia; PyRate.]

Variation in speciation and extinction rates through time and across lineages can establish a key causal link between patterns of biological diversity and the variety of factors that ultimately shape them (Sepkoski 1998; Foote 2000; Rabosky et al. 2012; Schluter and Pennell 2017). Consequently, accurate estimation of macroevolutionary rates is essential for understanding the uneven distribution of species richness across the tree of life and its temporal dynamics, including post-extinction recovery (Stanley 2007; Brayard et al. 2009), the identification of clades in decline (Quental and Marshall 2011; Sakamoto et al. 2016; Burin et al. 2019; Billaud et al. 2020), or time and diversity dependence (Foote et al. 2018; Henao Diaz et al. 2019; Pannetier et al. 2021). Historically, the problem of macroevolutionary rate estimation has been addressed using two broad classes of approaches, relying either on the quantitative analysis of the fossil record (Raup 1985; Foote 2000, 2003; Alroy 2008, 2014; Liow and Finarelli 2014) or time-calibrated molecular phylogenies of extant taxa (Harvey et al. 1994; Nee et al. 1994; Pybus and Harvey 2000; Alfaro et al. 2009; Morlon 2014; Rabosky 2014; Maliet et al. 2019). Both types of methods have seen extensive development and deployment, but often led to drastically different inferences about the underlying macroevolutionary dynamics (Quental and Marshall 2009, 2010; Liow et al. 2010; Morlon et al. 2011; Etienne et al. 2012; Mitchell et al. 2018).

In recent years, concerns about estimating extinction rates in the absence of fossil data (Rabosky 2010, 2016) and the widely noted discrepancy between

diversification rates estimated from paleobiological and neontological data (Hunt and Slater 2016; Marshall 2017; Silvestro et al. 2018) have fueled the development of approaches that can incorporate both extant and extinct taxa. Examples of such frameworks include PyRate (Silvestro et al. 2014a,b, 2019), developed as an extension of quantitative paleobiological methods, and Fossil BMM (Bayesian Analysis of Macroevolutionary Mixtures) (Mitchell et al. 2018), which extends extant-only phylogenetic approaches to extinct lineages. Despite their origins in different research traditions, the two methods share fundamental similarities in being Bayesian, utilizing birth–death models, and averaging the parameters of interest across models of different dimensionality. These properties make them well suited for the study of diversification dynamics through time, and both frameworks are consequently beginning to see widespread adoption (Silvestro et al. 2015; Condamine et al. 2016; Raia et al. 2016; Pires et al. 2018; Crouch et al. 2019; Crouch 2020; Lloyd and Slater 2021).

Recently, several lines of evidence have cast doubt on the ability of birth–death approaches to tease apart variation in rates of speciation, extinction, and fossil sampling. Diversification scenarios involving time-varying speciation and extinction rates are not identifiable from time trees of extant taxa alone (Louca and Pennell 2020), and while the ability of PyRate and Fossil BMM to incorporate fossil taxa could potentially alleviate this problem, the fossil preservation rate may itself be subject to similar identifiability issues. Covariation between extinction and fossil preservation has long been

recognized in paleobiology (Sepkoski 1975; Foote 2003, 2007; Foote et al. 2019) and motivated the development of approaches that jointly estimate the rates of speciation, extinction, and preservation (Connolly and Miller 2001; Foote 2003; Liow and Finarelli 2014), but even these approaches can be misled by heterogeneity in the rate of sampling and introduce spurious shifts into macroevolutionary rates as a result (Smiley 2018). Similarly, episodes of exceptional sampling were suggested to inflate PyRate estimates of net diversification rates (Condamine et al. 2016). Moreover, while both frameworks were tested and validated on empirical and simulated data upon their introduction (Silvestro et al. 2014b, 2019; Mitchell et al. 2018), these tests did not cover the full range of potential use cases. In particular, the availability of time trees for extinct clades (Bapst et al. 2016; Paterson et al. 2019) makes Fossil BMM applicable not only to groups with living representatives, as in the original simulations and current empirical applications (Mitchell et al. 2018; Crouch et al. 2019; Lloyd and Slater 2021), but also to clades that are wholly extinct, where its performance remains unknown.

In this study, we apply Fossil BMM and PyRate to the clade Ornithischia, one of the three major radiations of dinosaurs (Gauthier 1986; Baron et al. 2017; Langer et al. 2017), whose fossil record is ideally suited to testing the performance of the two frameworks. First, the ornithischians represent a wholly extinct, species-rich clade with an evolutionary history spanning at least a ~134 myr-long period from the Early Jurassic at the latest (Agnolín and Rozadilla 2018; Baron 2019) to the Cretaceous–Paleogene (K–Pg) extinction event (Archibald and Fastovsky 2004; Brusatte et al. 2015). Second, the ornithischians have been subject to extensive phylogenetic research (Boyd 2015; Han et al. 2018; Dieudonné et al. 2020), allowing phylogenetic information to be used in diversification rate estimation. Third, the ornithischians have a fossil record of adequate but not exceptional quality, marked by global episodes of poor sampling, patchy local records, and a substantial proportion of singletons (taxa known from a single stratigraphically unique occurrence) (Pol et al. 2011; Tennant et al. 2018). In terms of size and completeness, ornithischian phylogenetic and occurrence data are thus typical of the data sets widely in use by vertebrate paleontologists, whose broad availability may lead them to be repurposed for diversification rate estimation as methods like Fossil BMM and PyRate grow in popularity. Finally, there is an ongoing controversy concerning ornithischian diversity trends and diversification rates, especially those prior to the K–Pg extinction (Wang and Dodson 2006; Barrett et al. 2009; Lloyd 2012; Brusatte et al. 2015; Sakamoto et al. 2016; Chiarenza et al. 2019; Bonsor et al. 2020), further motivating the application of state-of-the-art diversification rate analysis to their fossil record.

Here, we infer the first Bayesian tip-dated phylogenies for Ornithischia based on multiple character matrices. We then use the resulting time trees and an expert-curated set of fossil occurrences to assess the performance of the diversification rate estimation

implemented in Fossil BMM and PyRate in an entirely extinct clade and under a wide range of analytical settings. We evaluate the congruence and robustness of the resulting diversification scenarios, and examine how model, prior, and algorithm choice affect the magnitude of speciation and extinction rate estimates as well as the number and position of inferred rate shifts. To assess the overall plausibility of our results, we additionally simulate birth–death trees and fossil records under empirically informed constraints to determine the extent to which the methods in question can disentangle the rates of speciation, extinction, and fossil sampling. We conclude by discussing the diversification scenarios inferred by our analyses in light of previous hypotheses about the clade’s macroevolutionary dynamics, and urge caution when applying current model-based methods to the low-power data sets common in vertebrate paleontology.

## MATERIALS AND METHODS

### *Occurrence Data*

We downloaded all ornithischian species-level body fossil occurrences that were present in the Paleobiology Database (PBDB; <http://www.paleobiodb.org>) on 1 August 2019; the full search settings and URL are provided in the Supplementary Information available on Dryad at <http://dx.doi.org/10.5061/dryad.sbccc2fr4x>. The manually curated data set comprised 1240 occurrences from 398 species (occurrences per species: median = 1, standard deviation = 5.22); 60.8% of the species were singletons.

### *Character Data and Tip Ages*

We updated and expanded three of the largest character matrices previously used for phylogenetic analyses of ornithischian dinosaurs: Han et al. (2018), Herne et al. (2019), and Madzia et al. (2018), hereafter referred to as “HaEA”, “HeEA”, and “MEA”, respectively (Supplementary Information, Phylogenetic Analyses available on Dryad). Following Barido-Sottani et al. (2019a), our time tree analyses sampled tip dates from their stratigraphic age ranges. Where possible, we extracted these from the PBDB data. The stratigraphic position of most occurrences could not be constrained beyond the stage level, leading to relatively wide ranges. We used the maximum first appearance date and the minimum last appearance date for taxa represented by multiple occurrences, and the maximum and minimum ages of the only occurrence for singletons. Several tips corresponded to unnamed specimens or recently described taxa absent from the PBDB; for these, age ranges were taken from the literature.

### *Phylogenetic Analyses and Tip-Dating*

Following preliminary maximum likelihood analyses using RAXML (Stamatakis 2014), we performed 16 tip-dating analyses using BEAST 2.6 (Bouckaert et al. 2019),

varying the treatment of topology (unconstrained vs. fixed to the RAxML estimate), the extant sampling parameter  $\rho$  (zero or nonzero), and (for the HeEA data set) the partitioning scheme. All analyses used the sampled-ancestor fossilized birth–death model (Gavryushkina et al. 2014) as the tree prior and a lognormal relaxed clock model with empirically derived hyperpriors (Fig. 1). Full details of phylogenetic analyses are provided in the Supplementary Information available on Dryad.

### Diversification Rate Analyses

**Fossil BAMB.**—We inferred the diversification dynamics of the ornithischians using BAMB v2.6 (Fossil BAMB; Mitchell et al. 2018). Like prior versions of BAMB (Rabosky 2014), Fossil BAMB uses a time-scaled phylogenetic tree to identify shifts separating distinct diversification regimes, without requiring the user to *a priori* specify their number or locations. Unlike the original version of the method, Fossil BAMB does not require the tree to be ultrametric, enabling the inclusion of non-contemporaneous tips, and its likelihood function incorporates a fossil preservation rate parameter estimated from the total number of stratigraphically unique species-level occurrences associated with the lineages represented in the tree

(Mitchell et al. 2018). The method automatically detects clade-specific rate shifts by using reversible-jump Markov chain Monte Carlo (rjMCMC) to move between models with different numbers of parameters, which in this case represent different diversification rates. Speciation rates may be constant or vary over time following an exponential change function (Rabosky 2014), while the extinction rate is assumed to be time-constant within each regime, and the rate of fossil preservation is held constant throughout the tree (Mitchell et al. 2018).

Before each BAMB analysis, we set priors on the initial rates of speciation and extinction and on the exponential rate change parameter using the `setBAM-Mpriors()` function from the R (R Core Team 2019) package `BAMMtools` v2.1.6 (Rabosky et al. 2014). Each analysis was run for 20 million generations, sampling every 10,000 generations and excluding the first 10% of samples as burnin following a visual inspection of the posterior traces. Markov chain Monte Carlo (MCMC) convergence was diagnosed using the package `coda` v0.19-2 (Plummer et al. 2006) by ensuring that the effective sample size (ESS) of all parameters exceeded 200, an arbitrary but widely used threshold (Lanfear et al. 2016; Ali et al. 2017). Using `BAMMtools`, we compared the prior and posterior support of models involving

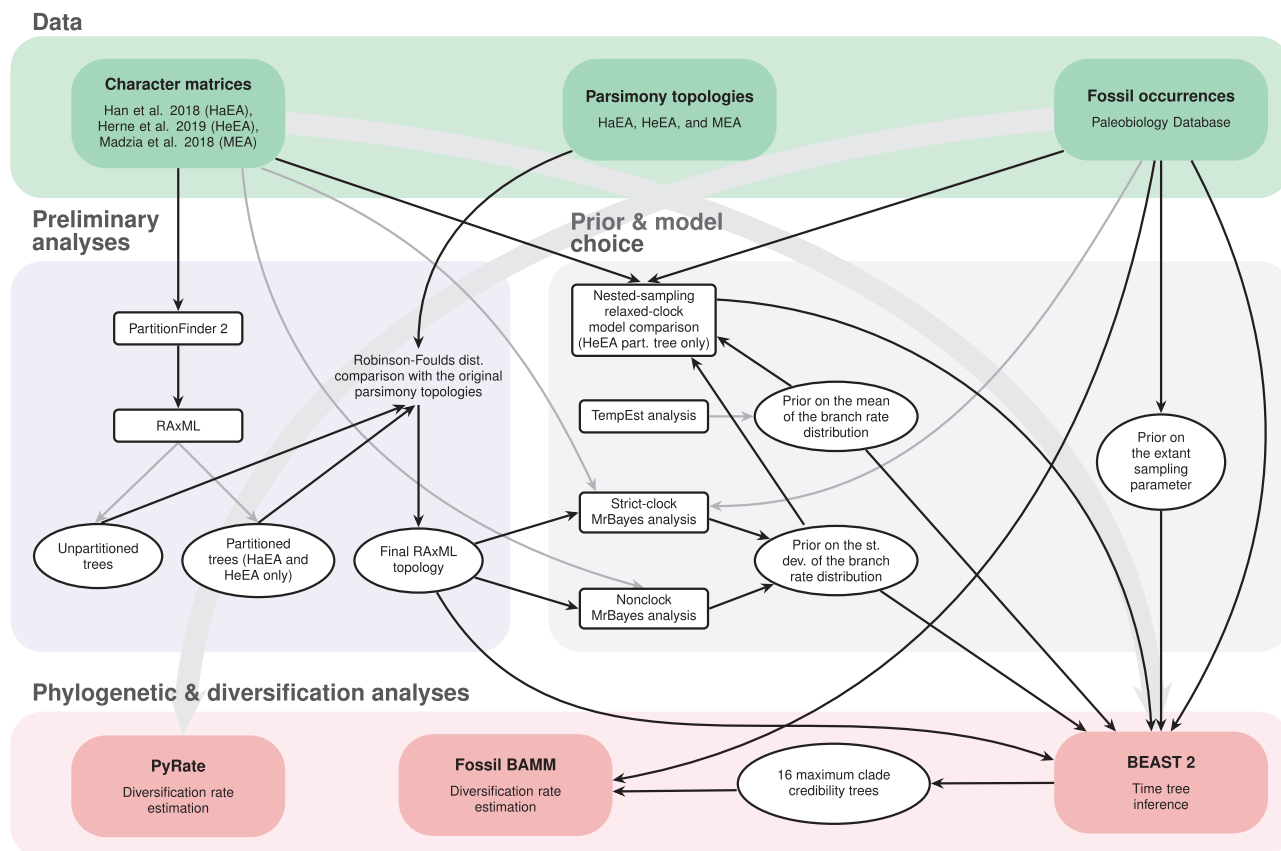


FIGURE 1. Flowchart showing the workflow of the phylogenetic and diversification rate analyses performed in this study. The protocol used for diversification rate estimation is further described below; for a detailed explanation and justification of the phylogenetic pipeline, see the Supplementary Information available on Dryad.



different numbers of rate shifts, and calculated their Bayes factor (BF) relative to the no-shift model. We also extracted the 95% credible set of configurations differing in the number and location of “core” shifts, or shifts that were sampled more often than expected under the prior alone. We used a marginal odds ratio of 5 (default BAMMtools threshold) to distinguish between core and non-core shifts. To determine overall trends in marginal rates, we used a modified BAMMtools function to plot the medians of the rate distributions through time across multiple analyses.

To accommodate topological and divergence-time uncertainty, we ran Fossil BAMM on each of the 16 BEAST maximum clade credibility trees, using the PBDB data to calculate the number of occurrences associated with their tips. Taxa present in the trees but not in the PBDB data were assumed to have a single occurrence each. The global sampling fraction (Chang et al. 2020) was calculated as the number of ingroup species in each tree divided by the total number of valid species in our occurrence data (HaEA, MEA: 0.186; HeEA: 0.136). For each tree, we conducted both a time-constant and time-varying analysis. To determine the strength of the signal for rate variation through time, we modified the R scripts of Friedman et al. (2019) to calculate 95% credible intervals (CIs) about the means of the speciation and net diversification rates at the root and at the tips and assessed their overlap. To evaluate prior sensitivity (Mitchell and Rabosky 2017), we additionally ran each analysis under three different values of the expected number of rate shifts (0.1, 1, 10), for a total of 96 analyses.

**PyRate.**—We used PyRate (Silvestro et al. 2014a,b, 2019) to estimate ornithischian macroevolutionary rates from fossil occurrence data. PyRate implements a hierarchical Bayesian approach that estimates the speciation and extinction times of every lineage based on the temporal distribution of fossil occurrences, a Poisson fossil sampling process, and a birth–death prior (Silvestro et al. 2014b). The speciation and extinction hyperparameters of the birth–death prior are assumed to be only piecewise-constant, allowing for an arbitrary number of rates separated by tree-wide shifts, and are themselves estimated from the data along with the speciation and extinction times and the fossil preservation rate using a transdimensional MCMC sampler. All lineages are assumed to be connected by an unknown underlying phylogeny in which every species has been sampled at least once (Silvestro et al. 2014b). The method makes use of all fossil occurrences, including those that fall between the first and last appearance, and those that constitute the only known record of a given lineage (singletons) (Silvestro et al. 2019).

We created two types of data sets: one in which the age of every occurrence was independently drawn from the corresponding range, and one “site-linked” data set. Site-linking accommodates the uncertainty regarding the age of a given fossil site while simultaneously accounting for the fact that the ages of all fossils from that site are approximately equal (King and Rücklin 2020). In a site-linked analysis, a single age is drawn from the

temporal range of every site and assigned to all the occurrences from the same site. We treated collection number metadata as a proxy for unique sites; manually added entries without metadata were assumed to each come from a unique site, resulting in 1087 sites per 1240 occurrences. We accounted for occurrence age uncertainty by replicating this procedure 10 times. Following Silvestro et al. (2019), we used the PyRate model selection tool based on the corrected Akaike information criterion (AICc) to compare the fit of the homogeneous Poisson preservation (HPP) model, corresponding to the frequentist null hypothesis; the nonhomogeneous HPP (NHPP) model, which allows the rate of preservation to change over the lifespan of each lineage; and two time-varying (TPP) models that were piecewise-homogeneous within each geochronological period (5 time bins: Early Jurassic through Late Cretaceous) or age (23 time bins: Hettangian through Maastrichtian). Based on the results, we limited all subsequent analyses to the HPP, NHPP, and TPP-age models, as the TPP-period model was intermediate between the above in terms of both fit and complexity (Supplementary Information, Table S5 available on Dryad). All three models were coupled with a four-category discrete  $\Gamma$  model of among-lineage preservation rate heterogeneity.

A total of 120 PyRate analyses were performed, varying the data set (site-linked vs. site-unlinked), replicate (10 per data set), preservation model (HPP, NHPP, TPP-age), and model averaging algorithm: in addition to the rjMCMC algorithm described by Silvestro et al. (2019), we also used birth–death Markov chain Monte Carlo (bdMCMC; Silvestro et al. 2014b), an alternative transdimensional MCMC sampler in which new rate parameters are introduced and removed following a birth–death process. Our preliminary runs showed that rjMCMC took longer to reach stationarity than bdMCMC; therefore, rjMCMC analyses were run for 125 million generations (sampling every 50,000), as opposed to 75 million generations at the same sampling frequency for bdMCMC. We considered the analyses to have converged if, after removing the first 10% of the chain as burnin, the ESS values were  $<200$  for no more than 10 parameters. We combined the post-burnin posteriors of all replicates that reached convergence, and summarized the posterior probabilities of models involving different numbers of speciation and extinction rates. We further extracted the marginal posterior distributions of speciation, extinction, and net diversification rates within 1 myr time bins, and calculated their means and 95% CIs to build rates-through-time (RTT) plots.

### Simulations

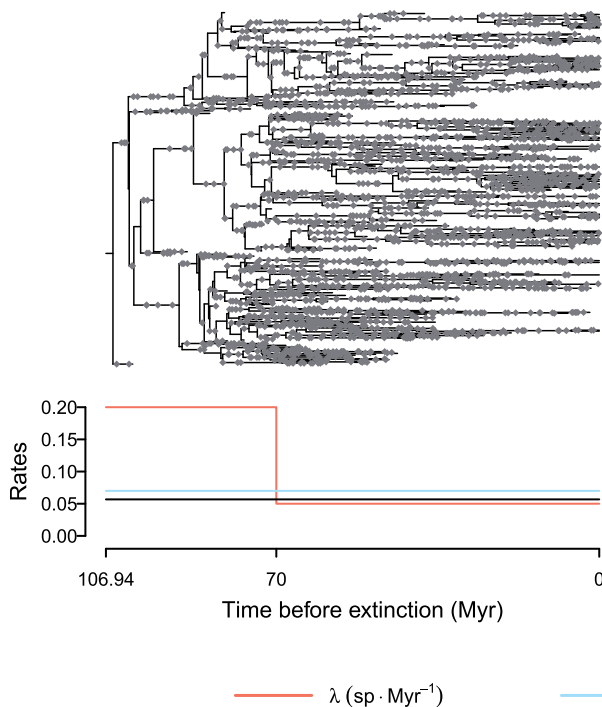
To assess how well PyRate and Fossil BAMM estimate the macroevolutionary and preservation rates, we simulated phylogenies and fossil records under two different scenarios. In the Decreasing Speciation scenario, the extinction and sampling rates were held constant at  $\mu = 0.07$  sp·myr<sup>−1</sup>,  $\psi = 0.567$  occ·sp<sup>−1</sup>·myr<sup>−1</sup>, while the rate

of speciation underwent a 4-fold decrease from  $\lambda_1 = 0.2 \text{ sp} \cdot \text{myr}^{-1}$  to  $\lambda_2 = 0.05 \text{ sp} \cdot \text{myr}^{-1}$  70 myr before the end of the process (Fig. 2). In the Decreasing Preservation scenario, the speciation and extinction rates were held constant at  $\lambda = 0.1 \text{ sp} \cdot \text{myr}^{-1}$ ,  $\mu = 0.07 \text{ sp} \cdot \text{myr}^{-1}$ , while the rate of fossil preservation decreased 4-fold from  $\psi_1 = 1.135 \text{ occ} \cdot \text{sp}^{-1} \cdot \text{myr}^{-1}$  to  $\psi_2 = 0.284 \text{ occ} \cdot \text{sp}^{-1} \cdot \text{myr}^{-1}$  at the same time horizon (Fig. 2; see the Supplementary Information available on Dryad for a justification of the values used). This choice of simulation settings was motivated by several considerations. First, previous analyses employing PyRate (Condamine et al. 2016) as well as quantitative paleobiological methods (Smiley 2018) showed speciation to track fossil preservation. This expected positive correlation facilitates interpretation: if the two rates are confounded, both scenarios should yield similar estimated speciation rates despite the different generating rates. Second, both scenarios tie in well with the empirical case, since a decline in the rate of speciation and its drop below the rate of extinction prior to the K-Pg boundary has been proposed for the ornithischians (Sakamoto et al. 2016), and the clade displays sharp variations in the preservation rate over its lifespan (Starrfelt and Liow 2016).

Using the R package TreeSim (Stadler 2011a), we simulated 100 birth–death trees under each scenario and employed rejection sampling to ensure maximum congruence with the empirical data, selecting 10 trees per

scenario (Supplementary Information, Table S7 available on Dryad). We used the package FossilSim (Barido-Sottani et al. 2019b) to simulate fossil records on the synthetic phylogenies according to the preservation rates above. Exact sampling times were recorded for all occurrences to eliminate stratigraphic age uncertainty. To ensure that the rate estimates are not confounded by topological or divergence time error, Fossil BMM was applied to the true (generating) phylogenies after pruning unsampled tips and truncating terminal edges to the age of their last occurrence. Additionally, we analyzed proportionally subsampled and 70-tip trees designed to mimic the phylogenetic data sets available for Ornithischia, which only include a fraction of the known diversity of the clade (Supplementary Information, Fossil BMM available on Dryad). Next, we used the AICc-based model selection implemented in PyRate to choose between the homogeneous Poisson preservation model (“HPP”), a time-varying model with a single shift at 136 Ma (“TPP-136”), and an arbitrarily overparameterized model allowing the preservation rate to change every 10 myr (“TPP-by10”), similar to the TPP-age model favored by the empirical analyses. We then analyzed each data set under both the best-performing model and the true model (HPP for Decreasing Speciation, TPP-136 for Decreasing Preservation) whenever the two differed (Supplementary Information, Table S8 available on Dryad). All PyRate analyses were further

### a) Decreasing Speciation



### b) Decreasing Preservation

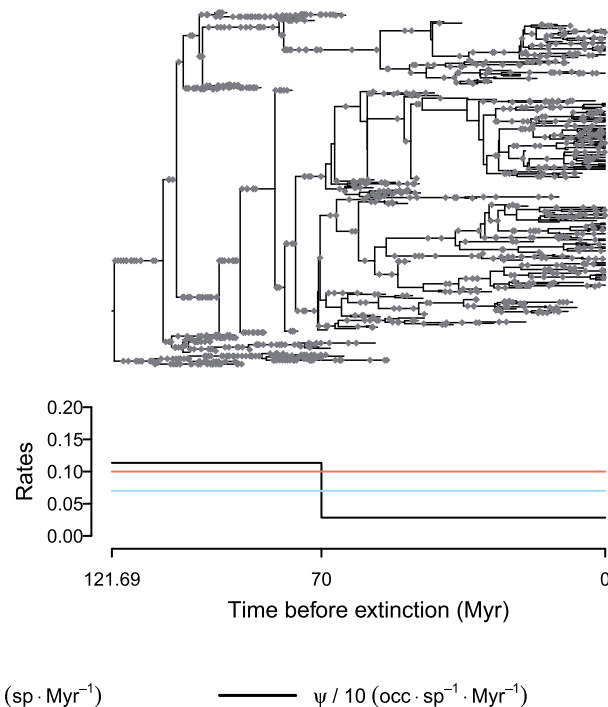


FIGURE 2. Examples of birth–death trees and the corresponding fossil records generated under the two simulation scenarios employed in this study. Each gray rectangle represents one fossil occurrence. a) Under Decreasing Speciation, the rates of extinction and preservation were held constant, while the rate of speciation underwent a 4-fold decrease. b) Under Decreasing Preservation, the rates of speciation and extinction were held constant, and the decrease affected the rate of preservation instead.

repeated on subsampled data sets excluding sampled ancestors, which were abundant in the simulations but absent from the empirical phylogenies (Supplementary Information, Phylogenetic Analyses available on Dryad). For both Fossil BMM and PyRate, we calculated the summary statistics described for the empirical analyses as well as the relative error and relative precision (width of the 95% CI divided by the mean rate) following [Silvestro et al. \(2019\)](#), averaging them first across time bins within a replicate and then across replicates.

## RESULTS

### *Diversification Rate Analyses*

**Fossil BMM.**—We found no evidence of shifts between distinct diversification regimes throughout ornithischian evolutionary history. The 95% credible set of rate shift configurations contained a single configuration for all but two analyses (both based on unconstrained HaEA topologies, with constant within-regime speciation and one expected shift). The maximum *a posteriori* (MAP) configuration contained no shifts in any of the 96 analyses performed; in the two analyses where the 95% credible set included an additional configuration, this second-best configuration contained a single shift associated with the clade Ornithopoda. The posterior probability of the zero-shift configuration always exceeded its prior probability, even in cases where the zero-shift configuration was already strongly favored by the prior (expected number of shifts = 0.1). Model comparisons using BFs found no evidence in favor of models containing rate shifts relative to the zero-shift model, with BFs for alternative models ranging from 0.002 to 0.781.

Analyses assuming constant within-regime speciation yielded model-averaged rate estimates that were almost perfectly flat through time (Fig. 3); we thus report the average of mean rates at the root and at the youngest tips. Across the 48 time-constant analyses, the mean of root-tip mean net diversification rates was 0.00501 sp·myr<sup>-1</sup> (range of root-tip mean rates = [0.0038, 0.0083] sp·myr<sup>-1</sup>), a low value reflecting low and nearly equal rates of speciation (0.0367 [0.0273, 0.0673] sp·myr<sup>-1</sup>), and extinction (0.0317 [0.0224, 0.0597] sp·myr<sup>-1</sup>). Analyses with time-varying within-regime speciation rates found strong support for a long-term decline, with 95% CIs showing no overlap between the root and the youngest tips in 41 and 39 out of 48 analyses for speciation and net diversification, respectively. In both cases, overlapping 95% CIs resulted from some but not all analyses based on zero-ρ HeEA and MEA trees, indicating that the estimated slowdown was due to neither taxon sampling nor topology. The mean of mean net diversification rates was 0.0463 sp·myr<sup>-1</sup> (range of means = [0.0307, 0.0567] sp·myr<sup>-1</sup>) at the root and -0.0117 sp·myr<sup>-1</sup> ([−0.0153, −0.0028] sp·myr<sup>-1</sup>) at the youngest tips. The 95% CIs were notably wider at the root (mean

width = 0.0758 sp·myr<sup>-1</sup>) than at the youngest tips (0.0233 sp·myr<sup>-1</sup>) for speciation but not for extinction (root: 0.0196 sp·myr<sup>-1</sup>, youngest tips: 0.0199 sp·myr<sup>-1</sup>).

**PyRate.**—The parameter-rich TPP-age preservation model exhibited the best fit to both site-linked and site-unlinked data (Supplementary Information, Table S5 available on Dryad). The fossil sampling rates estimated under the TPP model were broadly consistent between the bdMCMC and rjMCMC analyses, showing episodes of increased preservation in the Kimmeridgian–Tithonian, the Albian, and the Maastrichtian, with substantial uncertainty through much of the Early and Middle Jurassic (Supplementary Information, Figs. S20–S23 available on Dryad). In analyses subsequent to the model selection step, we experienced substantial convergence issues with the rjMCMC algorithm. In 2 out of the 12 analytical scenarios (site-linked and site-unlinked rjMCMC+HPP), all replicates failed to reach the ESS threshold for >10 parameters, and we thus excluded these scenarios from further comparisons.

The bdMCMC analyses inferred a diversification pattern similar to the results from time-constant BMM, with the exception of a sharp drop in the speciation rate and an uptick in the extinction rate 10–15 myr before the K–Pg boundary. These shifts may represent edge effects, a known artifact in which the clustering of first appearance dates at the beginning of the studied interval and of last appearance dates toward its end biases the estimated rates of speciation and extinction ([Foote 2000](#)). However, the absolute magnitudes of pre-80 Ma rates of speciation (range of time-averaged mean rates: [0.1737, 0.2684] sp·myr<sup>-1</sup>) and extinction ([0.1276, 0.1926] sp·myr<sup>-1</sup>) as well as their difference were an order of magnitude higher for PyRate than for BMM, with the poorly fitting NHPP model favoring slightly lower rates than the other models (Fig. 4). Consistent with the rate drops observed in the RTT plots, all our bdMCMC analyses inferred one shift in speciation and one shift in extinction (Supplementary Information, Table S6 available on Dryad). The second most frequently sampled speciation model involved one additional shift, yielding a slight decrease in the marginal rate around 160 Ma (Fig. 4).

Our rjMCMC analyses found two major periods of elevated speciation and extinction rates in the late Early through Middle Jurassic and in the early Late Cretaceous, characterized by biologically implausible rates (up to >5 sp·myr<sup>-1</sup> for the Jurassic peak) (Fig. 4). The support for all four corresponding shifts (at the beginning and end of either period) was always positive (BF > 2) and often strong (BF > 6) for both speciation (13 out of the 19 rjMCMC analyses that reached convergence) and extinction (14 analyses). The peaks were close to, but did not reach, the edges of the analyzed time interval, and we therefore consider them distinct from the usual edge effects, which we also observed in some analyses (Fig. 4). Between the peaks, the inferred rates were nearly constant and similar to the bdMCMC estimates for speciation (range of time-averaged mean rates between 160

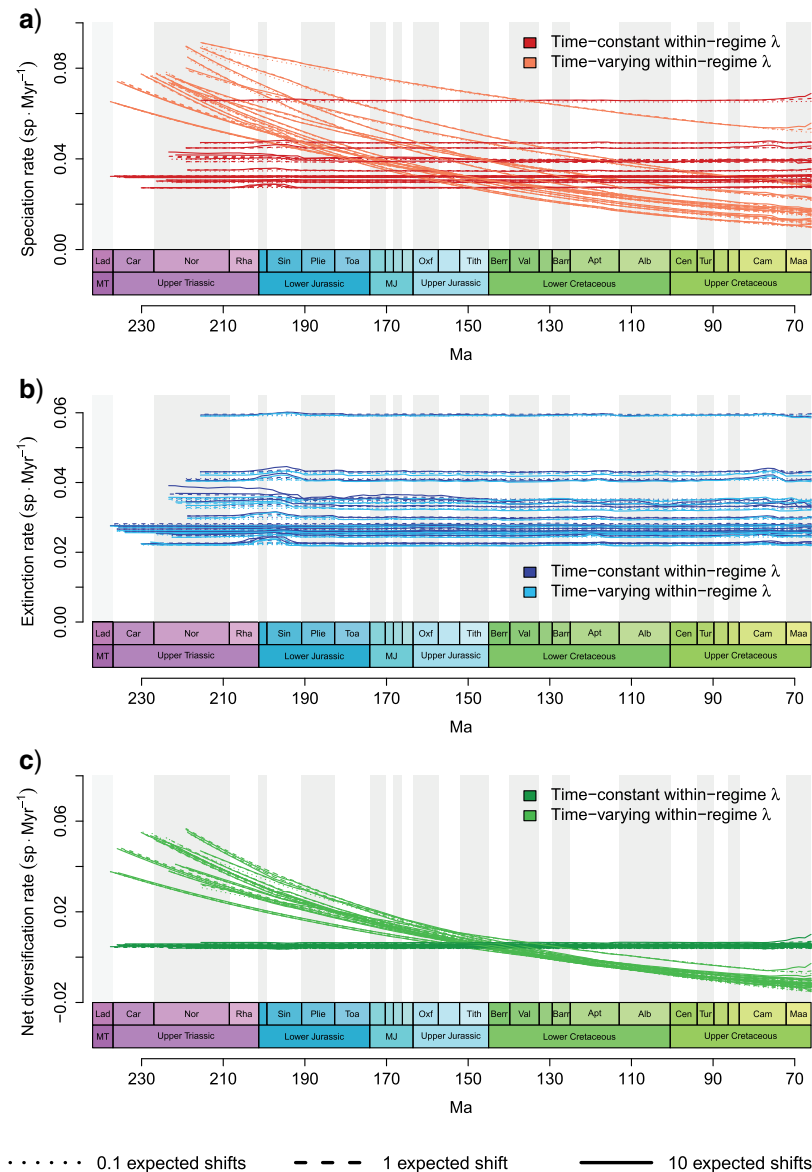


FIGURE 3. Diversification of Ornithischia estimated using Fossil BAMB, with rate-through-time plots (RTT) for a) speciation, b) extinction, and c) net diversification. Each curve represents the posterior mean from one analysis. The absence of support for multiple rate regimes results in net diversification rates that are either close to zero and constant (time-constant analyses), or gradually decay from the root toward the youngest tips (time-varying analyses).

and 90 Ma:  $[0.1329, 0.2684]$   $\text{sp} \cdot \text{myr}^{-1}$ ) but slightly higher for extinction ( $[0.1458, 0.2608]$   $\text{sp} \cdot \text{myr}^{-1}$ ), occasionally producing negative net diversification rates. As in the bdMCMC analyses, the marginal rates were strongly correlated, with the pulses of speciation almost entirely offset in magnitude by pulses of extinction; however, due to slight differences in timing, the net diversification rate obtained by their subtraction displayed a series of short-lived peaks and dips (Fig. 4f). The rjMCMC algorithm consistently supported more rate shifts compared to bdMCMC in both speciation and extinction. The latter algorithm never sampled more than 7 extinction rates, while the rjMCMC analyses favored up to 8 rates for

extinction, and sampled as many as 13 (Supplementary Information, Table S6 available on Dryad).

### Simulations

**Fossil BAMB.**—BAMB analyses consistently inferred a single rate regime under Decreasing Speciation, with only 6 out of 40 analyses including at least one shift in their MAP configuration. The analytical scenarios in which the MAP configurations did contain one or more shifts also tended to produce larger 95% credible shift sets (Supplementary Information, Table S9 available on



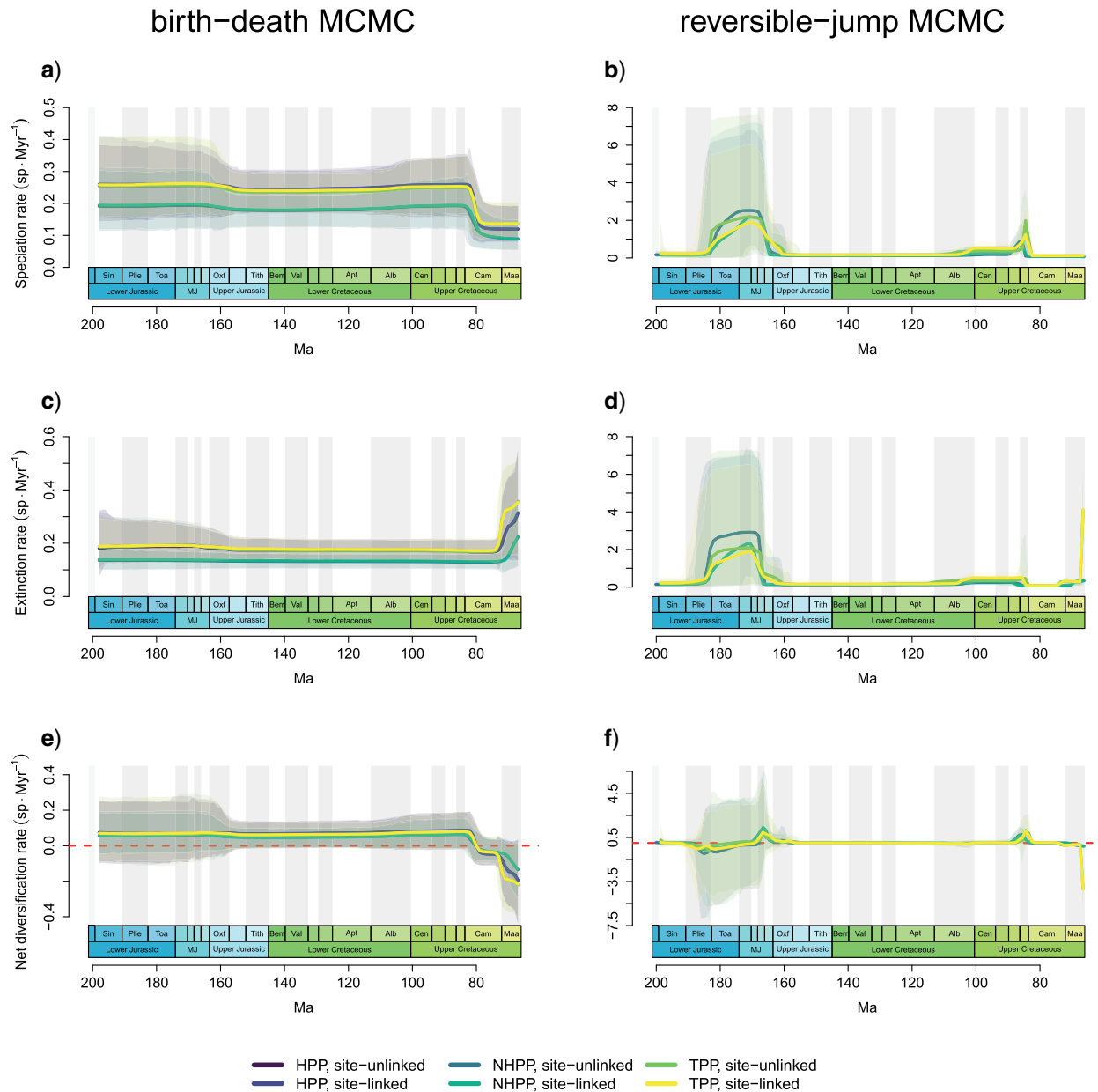


FIGURE 4. Diversification of Ornithischia estimated using PyRate under two different model-averaging algorithms, with RTT plots for (a, b) speciation, (c, d) extinction, and (e, f) net diversification. Each curve represents the posterior mean averaged across all replicates that reached convergence ( $\leq 10$  parameters with effective sample sizes of  $< 200$ ). The shaded areas show the associated 95% credibility intervals. The dashed line highlights the boundary between positive and negative net diversification. Aside from edge effects near the K–Pg boundary, the bdMCMC results resemble those obtained with BAMM, whereas the rjMCMC algorithm yields highly correlated speciation and extinction rates that accelerate to implausible values in the late Early Jurassic and the early Late Cretaceous.

Dryad); however, these never included configurations with more than 6 shifts, even though all Decreasing Speciation trees had 7–13 well-sampled branches crossing the time horizon of the shift (Supplementary Information, Table S7 available on Dryad). When the MAP configuration did include one (five analyses) or two (one analysis) shifts, these were generally associated with accelerated rather than decelerated speciation and postdated the true tree-wide speciation drop (Supplementary Information, Figs. S29–S31 available on

Dryad), suggesting that the method picked up stochastic variation rather than the generating birth–death process. Despite the absence of discrete shifts, the general tendency for the speciation rate to decrease over time was correctly inferred (Table 1). Seventeen out of 20 analyses with time-varying within-regime speciation exhibited non-overlapping 95% CIs about the net diversification rate at the root and at the youngest tips. The rates at the root and at the youngest tips were close to their true values in time-varying analyses, and the mean speciation



TABLE 1. Overview of the various aspects of Fossil BMM and PyRate performance based on the analyses of simulated data

Variable monitored	Fossil BMM		PyRate	
	Decreasing speciation	Decreasing preservation	Decreasing speciation	Decreasing preservation
Number and timing of $\lambda$ shifts	True tree-wide shift approximated by a single; regime with declining $\lambda$ ; occasional spurious shifts	Constant $\lambda$ almost always correctly inferred	Number overestimated (bdMCMC) or correctly inferred (rjMCMC); true shift accurately placed	Constant $\lambda$ correctly inferred
Number and timing of $\mu$ shifts	Constant $\mu$ correctly inferred	Constant $\mu$ correctly inferred	Spurious $\mu$ shifts not always attributable to edge effects	Spurious $\mu$ shifts attributable to edge effects
Number and timing of $\psi$ shifts	Not applicable	Not inferred due to inherent limitations of the method	Time-variable $\psi$ model incorrectly chosen over a time-homogeneous one	$\psi$ model with the true shift correctly chosen
Marginal $\lambda$ rates	Moderate (time-varying BMM) to low (time-constant BMM) accuracy, moderate precision, no bias	High accuracy, moderate precision, slightly biased toward overestimates	Low accuracy, moderate to low precision, consistently overestimated	Low accuracy, moderate to low precision, consistently overestimated
Marginal $\mu$ rates	High accuracy, moderate precision, consistently overestimated	Moderate accuracy, moderate precision, consistently overestimated	Low accuracy, moderate to low precision, consistently overestimated	Low accuracy, moderate to low precision, consistently overestimated
Marginal $\psi$ rates	High accuracy, high precision, poor coverage, consistently overestimated	Intermediate between the two true rates	Correctly inferred as near-constant; occasional spurious shifts	High accuracy, moderate precision, high coverage
Spurious correlation between $\lambda$ and $\psi$	Not applicable	Not applicable	In a minority of analyses	Absent
Spurious correlation between $\lambda$ and $\mu$	Present in marginal RTT plots and spurious shifts	Present in marginal RTT plots	Present and inducing a spurious shift in $\mu$	Present; ( $\lambda - \mu$ ) estimated accurately despite both $\lambda$ and $\mu$ being overestimated
Effect of subsampling	$\lambda$ and $\mu$ estimated with lower accuracy and lower precision; both consistently underestimated	$\lambda$ but not $\mu$ estimated with lower accuracy; both less precise and consistently underestimated	$\lambda$ and $\mu$ estimated with lower precision; more among-replicate variation close to the origin under bdMCMC	$\lambda$ and $\mu$ estimated with lower precision; more among-replicate variation close to the origin under bdMCMC
Effect of analytical settings	Weak for the expected number of shifts, moderate for within-regime speciation (time-constant vs. time-variable)		Weak for the preservation model, substantial for model-averaging algorithm (bdMCMC vs. rjMCMC)	

Note: Accuracy and precision are measured by relative error and the relative width of the 95% credible interval (CI), respectively. Coverage is measured by the proportion of cases in which the true value is included in the 95% CI. "Subsampling" refers to the 70-tip and proportionally subsampled schemes for BMM (Supplementary Information, Fossil BMM available on Dryad) and to the exclusion of sampled ancestors for PyRate.

rates (averaged over the entire history of each tree) inferred by time-constant analyses were intermediate between the two generating rates (Fig. 5). In contrast, subsampled analyses underestimated the speciation rate in both cases (Supplementary Information, Fig. S25 available on Dryad).

The extinction rate was consistently overestimated (Fig. 5), and its high relative precision corresponded to low coverage, with the true value included in the 95% CI in just 4 out of 40 cases both at the root and at the youngest tips. In the time-constant analyses, the 95% CI did not contain the true value at any point throughout the history of the clade. The preservation rate estimates were fairly accurate but biased and overconfident; the

rate was slightly but consistently overestimated (range of means: [0.595, 0.637]  $\text{occ-sp}^{-1} \cdot \text{myr}^{-1}$ ), and the narrow 95% CIs (relative precision: 0.069) never included the true value.

Under Decreasing Preservation, BMM correctly inferred the absence of speciation rate shifts, with only 1 out of 40 analyses containing a shift in its MAP configuration. In time-varying analyses, the 95% CIs about the net diversification rate at the root and at the youngest tips overlapped in 19 out of 20 cases. The marginal speciation rates were estimated with a high degree of accuracy (Fig. 5; Supplementary Information, Table S10 available on Dryad). On average, the 95% CIs were wider at the root (0.888) than at the youngest tips

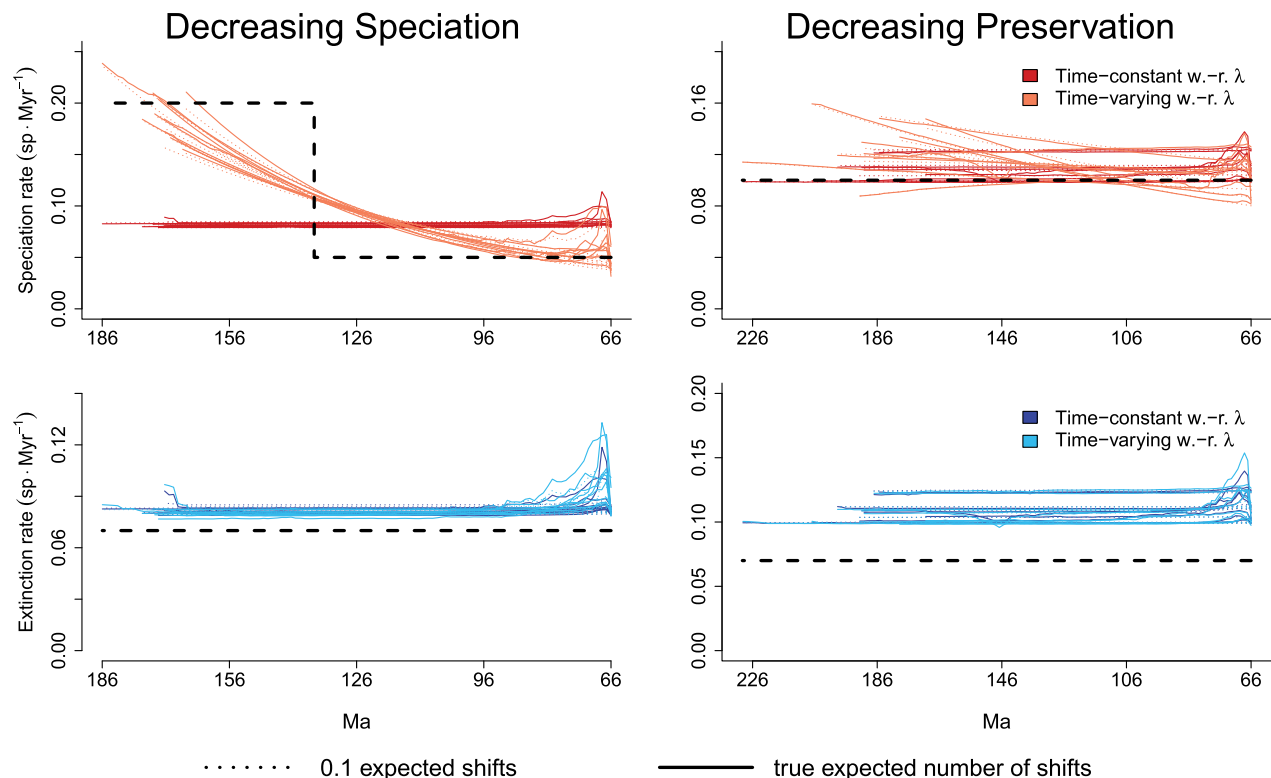


FIGURE 5. Speciation and extinction RTT plots from the Fossil BMM analyses of simulated data. Each curve represents the posterior mean from one analysis; thick dashed lines represent the true values. The shift in the Decreasing Speciation scenario is not identified as a discrete event but approximated by the declining rates in the time-varying analyses.

(0.531), but both values corresponded to comparably high coverage (true value included in 17 and 19 out of 20 cases at the root and at the youngest tips, respectively). Similarly, the true value fell within the CI at least at one point in the history of the clade in 17 out of the 20 time-constant analyses. In the subsampled analyses, accuracy, precision, and coverage were markedly poorer (Supplementary Information, Fossil BMM available on Dryad), suggesting a positive relationship between performance and tree size.

As in the Decreasing Speciation scenario, the extinction rate was systematically overestimated (Fig. 5). In contrast to speciation, the width of the 95% CIs did not differ appreciably between the root (0.310) and the youngest tips (0.341) in the time-varying analyses, resulting in similarly poor coverage in both cases (true value included in 1 and 0 out of 20 cases at the root and at the youngest tips, respectively). Low coverage also characterized the extinction rate estimates from time-constant analyses, where the 95% CIs included the true value at least once in just 2 out of 20 cases. The preservation rate was intermediate between the two true values (range of means: [0.416, 0.602]  $\text{occ-sp}^{-1}\cdot\text{myr}^{-1}$ ).

**PyRate.**—Our PyRate simulations suffered from fewer convergence issues than the empirical analyses. As a result, only 5 replicates out of 122 (3 for bdMCMC+HPP and 2 for bdMCMC+TPP, both in the Decreasing

Speciation scenario) had to be excluded from the following comparisons. Under Decreasing Speciation, model selection failed to identify the true preservation model (HPP), favoring the more parameter-rich TPP-136 and TPP-by10 models in 13 and 7 out of 20 cases, respectively (Supplementary Information, Table S8 available on Dryad). Nevertheless, the marginal preservation rates were mostly flat through time even in the TPP analyses, with most variation restricted to a period of uncertainty early in the history of the clade (Fig. 6). An exception consisted of a small but occasionally significant decrease at the 136 Ma time horizon; in 7 out of the 38 TPP analyses that reached convergence, there was no overlap of the 95% CIs for the time intervals immediately preceding and following the 136 Ma rate shift, indicating possible conflation of speciation and preservation rate changes. Analyses run under the true (HPP) preservation model slightly but systematically overestimated the true rate (range of means: [0.569, 0.642]  $\text{occ-sp}^{-1}\cdot\text{myr}^{-1}$ ), with the 95% CI including the true value in just 2 out of the 37 HPP analyses that reached convergence.

In contrast to the empirical analyses, the rjMCMC algorithm favored models with fewer rate shifts for both speciation and extinction. The rjMCMC analyses correctly inferred two distinct speciation rates under Decreasing Speciation, while the bdMCMC runs supported a third rate (Supplementary Information, Table S11

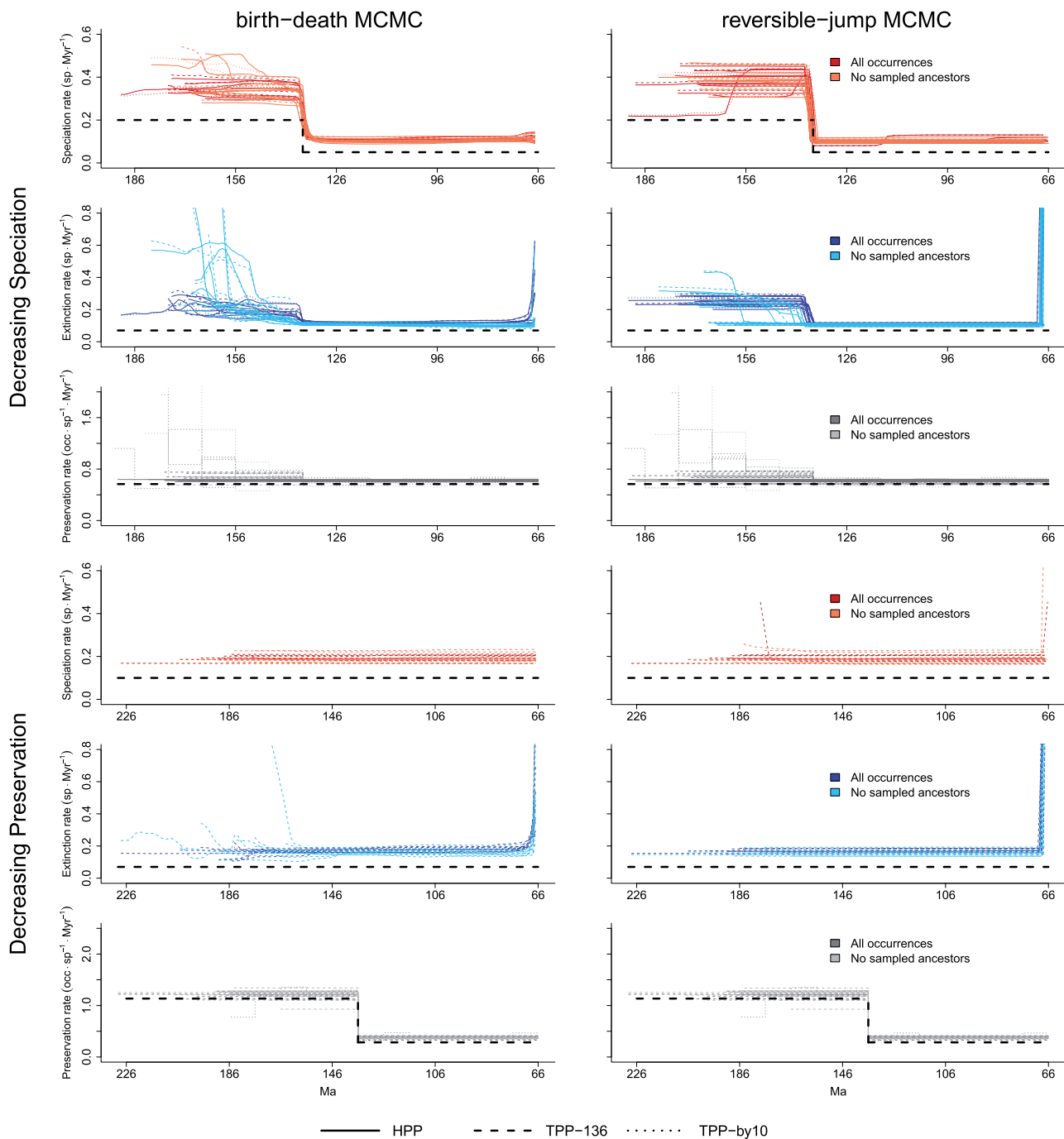


FIGURE 6. Speciation, extinction, and preservation RTT plots from the PyRate analyses of simulated data (cf. Fig. 5). The rate drop in the Decreasing Speciation scenario is correctly inferred but counterbalanced by a spurious drop in the rate of extinction. The rates under Decreasing Preservation are overestimated in absolute terms, but correctly inferred as constant through time despite the presence of a potentially confounding drop in the rate of preservation.

available on Dryad). Both algorithms overestimated the number of extinction rates. The RTT plots accurately reflect the overall speciation dynamics, with a sharp drop at 136 Ma successfully identified by all 80 analyses. In the *rj*MCMC analyses, which allowed estimating marginal posterior rate shift probabilities (Silvestro et al. 2019), the

probability of this drop usually approached 1. Spurious speciation rate shifts were sometimes inferred in the early history of the clade, especially in the analyses excluding sampled ancestors (Fig. 6). This is congruent with the expectation that sampled ancestors should form a higher proportion of lineages close to the origin of a

clade, and their removal should thus disproportionately affect rate inference therein.

While the number of spurious speciation shifts was relatively low, a spurious drop in the extinction rate temporally coinciding with that in the rate of speciation appeared in most of the 80 analyses. In the rjMCMC analyses, its posterior probability was similar to that of the speciation shift (i.e., close to 1) when sampled ancestors were included but lower ( $<0.6$ ) when they were excluded. The analyses without this spurious drop yielded other erroneous inferences about the extinction dynamics, including drops predating the 136 Ma time horizon, gradual decline between the origin of the clade and the 136 Ma boundary, or multiple shifts (Fig. 6). The magnitudes of the speciation rates both before and after the shift were systematically overestimated, as was the extinction rate. Interestingly, the highly correlated and overestimated rates combined to produce remarkably accurate estimates of the net diversification rate (Table 1), which were generally very close to the true values both before ( $0.13 \text{ sp} \cdot \text{myr}^{-1}$ ) and after ( $-0.02 \text{ sp} \cdot \text{myr}^{-1}$ ) the shift (Supplementary Information, Fig. S27 available on Dryad). However, because the temporal correlation between the speciation and extinction rate shifts was imperfect, and because of edge effects as well as the spurious extinction rate fluctuations, even the net diversification RTT plots usually failed to capture the true clade dynamics. Typical spurious results included a brief period of negative rates following the origin of the clade (common in bdMCMC analyses without sampled ancestors), and an uptick immediately preceding the drop at 136 Ma (common in rjMCMC analyses).

Under Decreasing Preservation, PyRate favored the true (TPP-136) model in 19 out of 20 cases; in one case, the more parameter-rich TPP-by10 model was preferred. The simulated drop in the preservation rate at 136 Ma was accurately detected in all cases (Fig. 6), with none of the 42 analyses showing an overlap between the 95% CIs before and after the shift. When the true model was used, the true value was usually included in the 95% CI about the pre-shift rate (30 out of 40 cases) but never in the 95% CI about the post-shift rate, which was slightly but consistently overestimated (range of means:  $[0.325, 0.418] \text{ occ} \cdot \text{sp}^{-1} \cdot \text{myr}^{-1}$ ). In the two TPP-by10 analyses, the 95% CIs contained the true values in all five pre-shift time periods, but only in the first of the seven post-shift periods.

Both the bdMCMC and rjMCMC algorithms correctly favored the model with a single speciation rate, and both again overestimated the number of distinct extinction rates (Supplementary Information, Table S11 available on Dryad). The extinction model misspecification was caused by an edge effect consisting of a sharp rise in the extinction rate shortly before the end of the observed time interval. Spurious shifts in the speciation rate were rare and again attributable to edge effects at the beginning or the end of the clade's lifespan (Fig. 6). Aside from these, no shift had a posterior

probability of  $>0.25$  in the rjMCMC analyses (for which this metric was available), and no analysis placed a speciation rate drop at 136 Ma, indicating that there was no conflation of speciation with preservation in this scenario. Similarly, the extinction rates were correctly inferred as flat throughout much of the clade's history, with rare exceptions limited to a short period following the origin of the clade and bdMCMC analyses without sampled ancestors (Fig. 6).

In absolute terms, the rates of speciation and extinction again displayed a bias toward higher values. While errors in the inference of the overall extinction dynamics were less serious under Decreasing Preservation than under Decreasing Speciation, they ultimately resulted in even lower levels of accuracy when quantified using the relative error (Supplementary Information, Table S12 available on Dryad). The accuracy of the speciation rate estimates was slightly higher and comparable between the two scenarios; however, their higher precision under Decreasing Preservation resulted in especially poor coverage, with only 2 out of 42 replicates containing the true value in their 95% CIs at any point in the history of the clade. As with Decreasing Speciation, the biases in the estimates of both rates largely canceled out when the two were subtracted to obtain the net diversification rate, whose true value ( $0.03 \text{ sp} \cdot \text{myr}^{-1}$ ) was inferred with high accuracy except when subject to edge effects.

## DISCUSSION

We inferred the first time trees for Ornithischia based on three large morphological data sets, and combined them with curated occurrence data to infer the dynamics of ornithischian diversification using Fossil BMM (Mitchell et al. 2018) and PyRate (Silvestro et al. 2014a,b, 2019). We found results that were markedly inconsistent between both methods as well as between different implementations of each approach (time-constant vs. time-varying within-regime speciation in BMM, bdMCMC vs. rjMCMC algorithms in PyRate). To investigate the causes of these discrepancies, we simulated phylogenies and fossil records approximating the ornithischian data under two simple scenarios. These were designed to test the ability of both methods to disentangle time-varying speciation, extinction, and preservation rates under conditions typical of data sets used in vertebrate paleontology. Analyses of the synthetic data led to the overestimation of the number of rate shifts (PyRate) or their smearing across the history of the clade (Fossil BMM), and revealed cases of inaccurate or spuriously correlated marginal rate estimates (Table 1). Our results suggest that the data sets available for many extinct clades may not be sufficient to provide reliable inferences about macroevolutionary dynamics when analyzed using existing methods, and offer insights into how the applicability of such methods to fossil data may be improved.



### Implications for Ornithischian Diversification

Previous analyses of ornithischian macroevolutionary dynamics have mostly focused on absolute species richness rather than speciation or net diversification rates (Barrett et al. 2009; Lloyd 2012; Brusatte et al. 2015; Chiarenza et al. 2019), hindering comparisons of the present results with earlier literature. Multiple studies have hypothesized that Ornithischia and other Mesozoic dinosaurs entered a period of decline prior to the K–Pg mass extinction (Fastovsky et al. 2004; Wang and Dodson 2006; Chiarenza et al. 2019). This period is often regarded as restricted to the Maastrichtian (Archibald 1996, 2011; Barrett et al. 2009), though a longer-term decline spanning much of the Cretaceous has been suggested (Lloyd 2012) and supported by one of the few attempts to directly estimate ornithischian diversification rates (Sakamoto et al. 2016). The scenario of a diversity diminution restricted to the last 10 myr of the Cretaceous is broadly compatible with our bdMCMC PyRate results, which show negative net diversification rates in this time interval. In contrast to the gradual waning proposed by earlier studies (Archibald 1996, 2011), our bdMCMC PyRate analyses support a two-stage decline comprising a discrete drop in speciation in the early Campanian, followed by a rising extinction rate in the Maastrichtian (Fig. 4). However, given their proximity to the K–Pg boundary, these rate changes possibly represent edge effects and should not be over-interpreted. The analogous drop in net diversification observed in the rjMCMC analyses was driven solely by extinction, while speciation held steady through the Campanian–Maastrichtian (Fig. 4).

The results of our time-varying Fossil BMM analyses do not support a discrete pre-K–Pg speciation drop or extinction spike, but are congruent with the longer-term decline inferred by Sakamoto et al. (2016) using a regression-based framework. Both methods show that the ornithischians crossed the zero net diversification line in the Early Cretaceous and continued to decline until the K–Pg boundary (Fig. 3). The absolute net diversification rate estimates are also similar between the two approaches, with maximum rates of  $\sim 0.06$  sp·myr<sup>−1</sup> attained early in the history of the clade (Fig. 3; Sakamoto et al. 2016). However, the results of Sakamoto et al. (2016) were recently interpreted as a possible artifact of failing to sample divergences that narrowly predate the K–Pg boundary (Wagner 2019). Fossil BMM, which also relies on the temporal distribution of divergences throughout a phylogeny, may be subject to the same issue, suggesting that an explicit parameterization of mass extinctions may be needed to capture the diversification dynamics of clades like Ornithischia.

Compared to Fossil BMM and PyRate with the bdMCMC algorithm, the rjMCMC PyRate analyses revealed a much more complex diversification history for Ornithischia, marked by multiple correlated shifts in speciation and extinction (Fig. 4). However, the magnitudes of the shifted rates estimated by rjMCMC

cast doubt on the reliability of the results. The mean speciation rates of 2–3 sp·myr<sup>−1</sup> inferred for the Early–Middle Jurassic spike (Fig. 6) rival the fastest rates ever recorded for vertebrates, including those of the white-eyes (1.95–2.63 sp·myr<sup>−1</sup>; Moyle et al. 2009), *Todiramphus* kingfishers (2.01–4.49 sp·myr<sup>−1</sup>; Andersen et al. 2015), and haplochromine cichlids (tip rates of up to 3.13–3.42 sp·myr<sup>−1</sup>; Chang et al. 2019); and the values included in the 95% CI (up to 7.5 sp·myr<sup>−1</sup>) substantially exceed them. The speciation rates during the Coniacian–Santonian spike are only slightly lower, and the extinction rates are similarly high for both peaks (Fig. 4).

The strong intra-method and inter-method incongruence calls into question the reliability of the macroevolutionary inferences obtained from our data. In particular, the question of a pre-K–Pg decline in speciation remains unresolved, with flat rates, a two-stage drop, or prolonged slowdown all receiving support depending on the method used. The small data sets employed here ( $\sim 50$ – $70$  tips, 1200 occurrences) may not contain a signal of the clade's diversification history that is strong enough to be independently picked up under the different modeling assumptions adopted by the two methods. Given the unexceptional quality of the ornithischian fossil record, and the fact that the ornithischian time trees generated for this study are fairly typical of paleontological phylogenies in terms of the number of tips, this conclusion is likely generalizable to numerous other extinct clades. We thus recommend that whenever possible, phylogeny-based inferences about macroevolutionary rates in extinct clades should be corroborated using multiple analytical frameworks with distinct modeling assumptions.

### Rate Shift Models vs. Model-Averaged Rates

One of the attractive features of Fossil BMM and PyRate consists in their ability to fit models of varying complexity and average the estimates of shared parameters across the different models, each weighted by its posterior probability (Rabosky 2014; Silvestro et al. 2019). Therefore, both the relative probabilities of different models and the model-averaged parameter values are of interest. Using simulations, we found that PyRate frequently failed to assign the highest posterior probability to the model containing the true number of speciation and extinction rate shifts, while Fossil BMM accounted for the tree-wide shift in the generating model by fitting a single regime and smearing the rate decline across its duration (Fig. 5), a tendency also noted by Mitchell et al. (2018). The failure of the tree-wide shift to induce a large number of simultaneous clade-specific shifts in BMM analyses is unsurprising given the previously noted difficulty of detecting small rate regimes (Mitchell and Rabosky 2017; Mitchell et al. 2018), with the little-discussed bias of BMM toward preferentially inferring shifts to regimes with increased rather than decreased rates (cf. Mitchell et al. 2018: Fig. 7b) as another

potential contributing factor. Arguably, the tree-wide decline in speciation inferred by BAMM is preferable to selectively mapping discrete rate drops onto sufficiently speciose clades only, and provides an imprecise but basically correct assessment of the overall diversification dynamics.

The overestimation of the number of rate shifts by PyRate appears to be due in part to edge artifacts, the causes of which are well-understood and can potentially be corrected for (Pires et al. 2018). Like Silvestro et al. (2019), we find rjMCMC to outperform bdMCMC when estimating the number of distinct speciation rates. However, in contrast to the findings of Silvestro et al. (2019), this is because bdMCMC overestimates rather than underestimates rate heterogeneity through time (Supplementary Information, Table S11 available on Dryad). In PyRate, the number of shifts is assigned a Poisson prior with rate parameter  $r$ , where  $r$  is fixed under bdMCMC and estimated under rjMCMC (Silvestro et al. 2014b, 2019). Since the rjMCMC prior adapts to the data, while the bdMCMC one does not, the difference in findings likely reflects the smaller number of shifts in our generating model compared to those of Silvestro et al. (2019). Interestingly, our simulation results also contrast with our empirical bdMCMC analyses, which tended to include fewer shifts than the corresponding rjMCMC runs (Supplementary Information, Table S6 available on Dryad). This may indicate genuine support for multiple shifts in our data that were better picked up by the flexible rjMCMC prior. However, it is also conceivable that when the ability of the data to accurately infer  $r$  is low, bdMCMC analyses that fix  $r$  to a plausible value can outperform the rjMCMC analyses that estimate it. The previously hypothesized conflation of macroevolutionary and preservation rates (Condamine et al. 2016) may represent another factor contributing to the appearance of spurious shifts. This effect was negligible in our simulations (Table 1; Fig. 6), but the highest-probability speciation and extinction rate shifts inferred by the rjMCMC analyses of the empirical data corresponded to an uptick in the number of geological sampling proxies (Supplementary Information, Fig. S2 available on Dryad) and a concomitant cluster of first appearance dates (Supplementary Information, Fig. S1 available on Dryad).

We corroborate the previously reported “paradox” of Bayesian diversification rate inference estimating marginal rates more successfully than the exact number of rate shifts (Mitchell et al. 2018). Our model-averaged rate estimates are robust to most sources of variation introduced into the simulations. The marginal rates inferred by PyRate are largely unaffected by the choice of the preservation model and the exclusion of sampled ancestors (Fig. 6). The latter is particularly remarkable, since the removal of sampled ancestors not only violates PyRate’s assumption that every species has been sampled at least once (Warnock et al. 2020), but does so in a nonrandom manner. For BAMM

we find that, similar to model probabilities (Mitchell and Rabosky 2017; Rabosky et al. 2017), the model-averaged rates are insensitive to the prior on the expected number of shifts (Figs. 3 and 5). Similar to rate variance (Rabosky 2019) and the power to detect rate shifts (Rabosky et al. 2017; Kodandaramaiah and Murali 2018), the accuracy of marginal rates depends more on tree size, as indicated by analyses performed on artificially subsampled trees of 23–70 tips (Supplementary Information, Fossil BAMM available on Dryad). However, this relationship is not straightforward, as shown by the extinction rate inference under Decreasing Preservation (cf. Fig. 5 and Supplementary Information, Fig. S25 available on Dryad), and should be explored over a wider range of tree sizes. More worryingly, we find evidence for spurious correlations between the speciation and extinction rates inferred by both PyRate and BAMM (Table 1), hinting at a possibility that only quantities that confound both rates, such as net diversification or the extinction fraction, may be estimable from low-information data sets (Supplementary Information, Results and Discussion available on Dryad).

#### *Challenges to Estimating the Diversification Dynamics of Extinct Clades*

The diversification dynamics of extinct groups, marked by symmetric phases of diversity expansion and contraction (Foote 2007), may differ from that of extant clades, which appear to steadily accumulate diversity from their origin to the present (Rabosky et al. 2012; Burin et al. 2019; but see Quental and Marshall 2010). The fact that BAMM and PyRate have been developed or predominantly used for the analysis of extant groups raises the question of whether their performance in the present study is not simply an effect of stretching their application beyond the use cases they were intended for. However, this is arguably not the case. The ornithischian dinosaurs, used here as an empirical system, were extirpated in the near-instantaneous K–Pg extinction event (Brusatte et al. 2015) at high standing diversity (a phenomenon commonly observed in mass extinctions; Jablonski 1986), and it is unclear from this study whether they entered a period of decline prior to their demise (see also Wang and Dodson 2006; Barrett et al. 2009; Sakamoto et al. 2016). Similarly, our synthetic data sets were generated under the birth–death process, which is invariant to translation in time. The trees were conditioned on a given number of “extant” tips and shifted into the past by an arbitrary number of time units. In the case of Fossil BAMM analyses, this shift was purely formal: the method does not independently account for the ages of the root and the youngest known occurrence, but only for their difference (“observationTime”; Mitchell et al. 2018), which remained unaffected by the shift.

We suggest that rather than exhibiting fundamentally different macroevolutionary dynamics, extinct clades suffer from problems related to both data quantity

and quality that render them ill-suited to analyses using currently available diversification rate estimation methods. The inferior performance of Fossil BMM on subsampled phylogenies ([Supplementary Information, Table S10](#) available on Dryad) suggests the need for large trees that are rare in paleontology; indeed, we are aware of only one matrix ([Hartman et al. 2019](#)) in which the number of taxa exceeds the average number of sampled tips in our simulated trees ( $n=291$ ). Moreover, morphological character matrices are often designed to resolve relationships at specific taxonomic levels, with comprehensive sampling often limited to narrow taxonomic scales. In contrast, data sets covering broader groups (such as Ornithischia) tend to adopt a diversified sampling strategy ([Höhna et al. 2011](#)), in which each subclade is represented by a small sample of well-preserved species that may not reflect its overall diversity. Accordingly, the absence of shifts from our Fossil BMM analyses is possibly explained by the fact that the character matrices used for this study included a wealth of early-diverging taxa relevant to the higher-order ornithischian phylogeny but few or no representatives of Hadrosauriformes and Ceratopsidae, the two deeply nested clades previously associated with diversification rate shifts ([Lloyd et al. 2008](#); [Sakamoto et al. 2016](#)). While the number of tips does not account for the ambiguous inferences obtained with PyRate, which does not use phylogenies as its input, data quality is also of concern. Compared to the rhinocerotid and marine mammal data sets previously used to explore the performance of PyRate ([Silvestro et al. 2014b, 2019](#)), our data included fewer occurrences (1240 vs. 2463 and 4740) spanning a longer period of time (135 myr vs. 45 and 66 myr), a lower average number of occurrences per taxonomic unit (3.12 vs. 15.02 and 8.86), and a much higher proportion of singletons (0.61 vs. 0.21 and 0.38). If the rich record of Cenozoic fossil mammals is the exception rather than the rule ([Marshall 2017](#)), our results suggest that PyRate may not perform well in settings representative of a typical vertebrate fossil record.

### Best Practices

We recommend that for extinct taxa, the use of Fossil BMM be limited to the basic assessment of overall clade dynamics. Our simulations suggest that the method can reliably detect declining speciation rates and accurately infer the magnitude of the decline, but its focus on clade-specific events makes Fossil BMM unable, at present, to distinguish between gradual and sharp transitions. This precludes inference of the onset and duration of any tree-wide decline, rendering the method unsuited for addressing many questions of paleobiological interest, including that of a possible pre-extinction drop in ornithischian diversification investigated here. While the ability of BMM to infer clade-specific shifts allows valuable insights into the macroevolutionary dynamics of extant taxa ([Friedman et al. 2019](#); [Henao Diaz et al. 2019](#)), the species-poor phylogenetic data sets available for extinct clades are unlikely to afford BMM sufficient power to

detect lineage-specific events. This difficulty may be further compounded by diversified sampling ([Höhna et al. 2011](#)), which obscures the differences between species-rich and species-poor clades and remains a common feature of paleontological data sets ([Simões et al. 2020](#)). If Fossil BMM analyses are conducted on such data sets, the lack of rate shifts should not be interpreted as positive evidence for uniform diversification.

The epoch model of diversification implemented in PyRate is more appropriate when coordinated shifts in speciation or extinction rates are expected to have occurred across multiple lineages, as may be the case during mass extinctions and rapid radiations typical of post-extinction recovery. Our results suggest that real shifts can be accurately inferred ([Fig. 6](#)), but the converse is not true: inferred shifts may not always be real. We thus suggest that PyRate analyses be coupled with (1) a careful re-examination of the original occurrence data to determine whether the inferred shifts coincide with conspicuous clusters of first or last appearance dates (cf. [Supplementary Information, Fig. S1](#) available on Dryad), and (2) independent assessments of the rate of preservation to gain insight into the extent to which such clusters reflect preservational rather than macroevolutionary dynamics. Such assessments may include simple geological proxies (e.g., number of rock units or localities per time bin; [Butler et al. 2012](#); [Holland 2016](#)), or estimates from more traditional paleobiological methods such as the capture-mark-recapture ([Connolly and Miller 2001](#)) or three-timer ([Alroy 2008](#)) techniques. We further suggest exploring (3) the influence of the model-averaging algorithm (bdMCMC vs. rjMCMC) and (4) the effect of the Poisson prior (or the corresponding hyperprior in the case of rjMCMC) on the number of shifts. The issue of the potential sensitivity of the inferred number of shifts to the corresponding prior has received attention in the context of BMM analyses of extant data ([Moore et al. 2016](#); [Rabosky et al. 2017](#)), and its importance may be even higher for the low-power data sets typically available for fossil vertebrates.

The ability of Fossil BMM and PyRate to provide checks on each other's results is limited by their focus on fundamentally different types of rate shifts (clade-specific vs. tree-wide, respectively). However, we still suggest that both methods may be profitably employed together. This is particularly true when marginal rates are of interest, as BMM estimates speciation rates with moderate to high accuracy ([Fig. 5](#); [Supplementary Information, Table S10](#) available on Dryad), while PyRate performs better at estimating the rate of net diversification ([Supplementary Information, Fig. S27](#) available on Dryad), suggesting possible complementarity of the two approaches. The preservation rates estimated by Fossil BMM and PyRate also allow useful comparisons, and evidence for preservational heterogeneity from the latter method, such as the preference for a TPP model and non-overlapping credibility intervals for interval-specific rates, can reveal cases where the time-constant preservation model employed by BMM may be misspecified, potentially causing erroneous inferences



about diversification. Additionally, while Fossil BAMM should not be expected to replicate the tree-wide shifts inferred by PyRate, it can capture the dynamics of a shift (or a monotonic series of shifts) in the form of non-overlapping credibility intervals for the rates at the root and at the tips, providing for a limited but important range of diversification scenarios that can be tested by assessing congruence between the two methods.

### *Future Directions*

The single empirical test case and simple simulation scenarios explored in this study do not exhaust the range of conditions likely to be encountered in diversification rate analyses of extinct clades. Our synthetic data were generated under a bifurcating model of speciation, following the complete absence of ancestor-descendant pairs from our ornithischian time trees. However, budding and anagenetic speciation are common features of the fossil record (Raup 1985; Foote 1996; Hopkins 2011; Parins-Fukuchi 2021), and can be simulated using the same tools we employed here (Barido-Sottani et al. 2019b). Exploring the sensitivity of diversification rate estimates to the mode of speciation can thus represent a fruitful subject for future research. Another problem left unaddressed by the current study is the amount of data needed to distinguish discrete tree-wide shifts from prolonged rate changes. While our simulations show that Fossil BAMM tends to approximate tree-wide speciation rate shifts by fitting a single exponentially varying rate rather than multiple clade-specific shifts, it may be worthwhile to fit both models and record their likelihood difference as a function of tree size and rate change magnitude, so as to estimate the amount of evidence required for Fossil BAMM to detect simultaneous shifts along multiple branches. Valuable insights may also be obtained from more complex simulations in which clade-specific shifts are clustered in a short period of time rather than perfectly simultaneous, or in which the shift magnitude varies across branches in a probabilistic fashion. Besides representing a more realistic description of adaptive radiation-like phenomena, such simulations would have the added advantage of being able to tease apart the relative contributions of low power and model violation to the Fossil BAMM performance issues reported here.

The model-based framework adopted by Fossil BAMM and PyRate lends itself easily to empirically motivated extensions, holding out promise that continued model development may mitigate the issues reported here and improve the applicability of both methods to entirely extinct taxa. We suggest that high priority should be given to the development of approaches that accommodate both tree-wide and clade-specific shifts in a single framework (cf. Laurent et al. 2015), possibly including mass extinctions as a third and separate class of events (Höhna 2015; May et al. 2016). It may be argued that true tree-wide shifts should virtually never be expected, since events acting across the tree, like mass extinctions and subsequent recoveries, should

not change the diversification rates of all lineages to exactly the same extent; stochastic variation should be expected even in the absence of life history differences or other factors that render some lineages more prone to speciation or extinction than others. However, we argue that while they may not perfectly capture the underlying diversification dynamics, discrete tree-wide shifts represent a useful compromise between precision (which compares favorably to smearing a shift across the entire duration of a rate regime) and the generally low power to detect lineage-specific shifts subtending tip-poor clades (Rabosky et al. 2017)—a point borne out by the multiple methods already dedicated to inferring such events (Rabosky 2006; Morlon et al. 2011; Stadler 2011b; May et al. 2016). A further argument could be made for treating mass extinctions as fundamentally different from rate shifts of any kind; we refer readers to May et al. (2016) for a defense of this approach.

When analyzing extinct taxa, fossil preservation has to be modeled with the same care as lineage diversification (Foote 2003, 2007; Liow and Finarelli 2014). Currently, PyRate holds an advantage over Fossil BAMM in allowing greater flexibility in its preservation models, the most general of which can account for variation both through time and among lineages (Silvestro et al. 2019). Fossil BAMM may benefit from adopting similarly flexible models, although it is notable that our simulations did not find its performance to be seriously compromised by the assumption of rate constancy (Fig. 5), and that the preservation rate inferred by BAMM for the ornithischians was closer to an independently derived estimate than the PyRate results based on a more complex model (Supplementary Information, Preservation Rates available on Dryad). The latter finding points to the importance of balancing the realism of a model against the ability to estimate all of its parameters with sufficient accuracy, and indicates that model selection tools may be as important as the models themselves. PyRate currently relies on the relatively liberal corrected Akaike information criterion to avoid the high computational costs associated with full BF comparisons (Silvestro et al. 2019), underscoring the importance of developing more efficient ways of estimating marginal likelihoods (Fourment et al. 2020). While the risk of overparameterization remains salient, occurrence-rich data sets may benefit from preservation models even more complex than those currently implemented in PyRate. Such models could account for geographic and environmental rate variation (Wagner and Marcot 2013; Holland 2016) or use empirically informed descriptions of how preservation potential changes over the lifespan of a lineage (Marshall 2019). Alternatively, fossil preservation could be treated in the same manner as diversification, allowing for both temporal (tree-wide) rate shifts and lineage-specific events, such as the colonization of environments conferring increased preservation potential.

We encourage researchers to take advantage of the parametric framework common to the methods evaluated here and continue the development of models that better account for the idiosyncrasies of fossil data,



as well as the exploration of their performance under paleobiologically realistic conditions. In light of the issues highlighted here, it is imperative that these steps be taken before the widespread application of existing methods to extinct clades. It is equally important to recognize the limitations inherent to most paleontological phylogenetic data sets, which were typically not assembled with diversification rate estimation in mind, and to interrogate their suitability to address the question under examination. Combined with the flexibility of Bayesian model-averaging approaches, such restricted but informed use of fossil data can help shed light on macroevolutionary dynamics throughout the tree of life (Silvestro et al. 2018; Louca and Pennell 2020; Lloyd and Slater 2021).

#### SUPPLEMENTARY MATERIAL

Data available from the Dryad Digital Repository: <http://dx.doi.org/10.5061/dryad.sbcc2fr4x>.

#### ACKNOWLEDGMENTS

We thank Matt Friedman, Peter J. Wagner, and three anonymous reviewers for helpful comments and constructive criticism. We benefited from PBDB data compiled by numerous enterers and authorizers, especially M.T. Carrano, P.D. Mannion, A.K. Behrensmeier, G.T. Lloyd, and J. Alroy. D.Č. further thanks the instructors and participants of the Taming the BEAST, eh! Workshop (11–16 August 2019, Squamish, British Columbia), Jonathan Chang, and Michael Foote for valuable advice, and April Wright for stimulating discussions. Michael Foote and Rossy Natale provided helpful comments on an early draft of the article. All simulations were run on the Midway2 Research Computing Cluster at the University of Chicago. This is Paleobiology Database publication number 401.

#### FUNDING

This work was supported by the National Science Centre, Poland (2020/37/B/NZ8/01321 to D.M.).

#### COMPETING INTERESTS

There are no conflicts of interest to disclose.

#### REFERENCES

- Agnolín F.L., Rozadilla S. 2018. Phylogenetic reassessment of *Pisano-saurus mertii* Casamiquela, 1967, a basal dinosauriform from the Late Triassic of Argentina. *J. Syst. Palaeont.* 16:853–879.
- Alfaro M.E., Santini F., Bock C., Alamillo H., Dornburg A., Rabosky D.L., Carnevale G., Harmon L.J. 2009. Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. *Proc. Natl Acad. Sci. USA.* 106:13410–13414.
- Ali R.H., Bark M., Miró J., Muhammad S.A., Sjöstrand J., Zubair S.M., Abbas R.M., Arvestad L. 2017. VMCMC: a graphical and statistical analysis tool for Markov chain Monte Carlo traces. *BMC Bioinformatics.* 18:97.
- Alroy J. 2008. Dynamics of origination and extinction in the marine fossil record. *Proc. Natl Acad. Sci. USA.* 105:11536–11542.
- Alroy J. 2014. Accurate and precise estimates of origination and extinction rates. *Paleobiology.* 40(3):374–397.
- Andersen M.J., Shult H.T., Cibois A., Thibault J.-C., Filardi C.E., Moyle R.G. 2015. Rapid diversification and secondary sympatry in Australo-Pacific kingfishers (Aves: Alcedinidae: *Todiramphus*). *R. Soc. Open Sci.* 2:140375.
- Archibald J.D. 1996. Dinosaur extinction and the end of an era: what the fossils say. New York: Columbia University Press.
- Archibald J.D. 2011. Extinction and radiation: how the fall of the dinosaurs led to the rise of the mammals. Baltimore: The Johns Hopkins University Press.
- Archibald J.D., Fastovsky D. 2004. Dinosaur extinction. In: Weishampel D.B., Dodson P., Osmólska H., editors. *The Dinosauria*. 2nd ed. Berkeley: University of California Press. p. 672–684.
- 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., Aguirre-Fernández G., Hopkins M.J., Stadler T., Warnock R.C.M. 2019a. Ignoring stratigraphic age uncertainty leads to erroneous estimates of species divergence times under the fossilized birth–death process. *Proc. R. Soc. B.* 286:20190685.
- Barido-Sottani J., Pett W., O'Reilly J.E., Warnock R.C.M. 2019b. FossilSim: an R package for simulating fossil occurrence data under mechanistic models of preservation and recovery. *Methods Ecol. Evol.* 10:835–840.
- Baron M.G. 2019. *Pisano-saurus mertii* and the Triassic ornithischian crisis: could phylogeny offer a solution? *Hist. Biol.* 31:967–981.
- Baron M.G., Norman D.B., Barrett P.M. 2017. A new hypothesis of dinosaur relationships and early dinosaur evolution. *Nature.* 543:501–506.
- Barrett P.M., McGowan A.J., Page V. 2009. Dinosaur diversity and the rock record. *Proc. R. Soc. B.* 276(1667):2667–2674.
- Billaud O., Moen D.S., Parsons T.L., Morlon H. 2020. Estimating diversity through time using molecular phylogenies: old and species-poor frog families are the remnants of a diverse past. *Syst. Biol.* 69(2):363–383.
- Bonsor J.A., Barrett P.M., Raven T.J., Cooper N. 2020. Dinosaur diversification rates were not in decline prior to the K-Pg boundary. *R. Soc. Open Sci.* 7:201195.
- 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., Poppinga 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 Comp. Biol.* 15(4):e1006650.
- Boyd C.A. 2015. The systematic relationships and biogeographic history of ornithischian dinosaurs. *PeerJ.* 3:e1523.
- Brayard A., Escarguel G., Bucher H., Monnet C., Brühwiler T., Goudemand N., Galfetti T., Guex J. 2009. Good genes and good luck: ammonoid diversity and the end-Permian mass extinction. *Science.* 325:1118–1121.
- Brusatte S.L., Butler R.J., Barrett P.M., Carrano M.T., Evans D.C., Lloyd G.T., Mannion P.D., Norell M.A., Peppe D.J., Upchurch P., Williamson T.E. 2015. The extinction of the dinosaurs. *Biol. Rev.* 90:628–642.
- Burin G., Alencar L.R.V., Chang J., Alfaro M.E., Quental T.B. 2019. How well can we estimate diversity dynamics for clades in diversity decline? *Syst. Biol.* 68(1):47–62.
- Butler R.J., Brusatte S.L., Andres B., Benson R.B.J. 2012. How do geological sampling biases affect studies of morphological evolution in deep time? A case study of pterosaur (Reptilia: Archosauria) disparity. *Evolution.* 66(1):147–162.
- Chang J., Rabosky D.L., Alfaro M.E. 2020. Estimating diversification rates on incompletely sampled phylogenies: theoretical concerns and practical solutions. *Syst. Biol.* 69(3):602–611.
- Chang J., Rabosky D.L., Smith S.A., Alfaro M.E. 2019. An R package and online resource for macroevolutionary studies using the ray-finned fish tree of life. *Methods Ecol. Evol.* 10(7):1118–1124.

- Chiarenza A.A., Mannion P.D., Lunt D.J., Farnsworth A., Jones L.A., Kelland S.-J., Allison P.A. 2019. Ecological niche modelling does not support climatically-driven dinosaur diversity decline before the Cretaceous/Paleogene mass extinction. *Nat. Commun.* 10:1091.
- Condamine F.L., Clapham M.E., Kergoat G.J. 2016. Global patterns of insect diversification: towards a reconciliation of fossil and molecular evidence? *Sci. Rep.* 6:19208.
- Connolly S.R., Miller A.I. 2001. Joint estimation of sampling and turnover rates from fossil databases: capture-mark-recapture methods revisited. *Paleobiology*. 27(4):751–767.
- Crouch N.M.A. 2020. Extinction rates of non-avian dinosaur species are uncorrelated with the rate of evolution of phylogenetically informative characters. *Biol. Lett.* 16(6):20200231.
- Crouch N.M.A., Ramanauskas K., Igić B. 2019. Tip-dating and the origin of Telluraves. *Mol. Phylogenet. Evol.* 131:55–63.
- Dieudonné P.-E., Cruzado-Caballero P., Godefroit P., Tortosa T. 2020. A new phylogeny of cerapodan dinosaurs. *Hist. Biol.* doi:10.1080/08912963.2020.1793979.
- Etienne R.S., Haegeman B., Stadler T., Aze T., Pearson P.N., Purvis A., Phillimore A.B. 2012. Diversity-dependence brings molecular phylogenies closer to agreement with the fossil record. *Proc. R. Soc. B.* 279:1300–1309.
- Fastovsky D.E., Huang Y., Hsu J., Martin-McNaughton J., Sheehan P.M., Weishampel D.B. 2004. Shape of Mesozoic dinosaur richness. *Geology*. 32(10):877–880.
- Foote M. 1996. On the probability of ancestors in the fossil record. *Paleobiology*. 22(2):141–151.
- Foote M. 2000. Origination and extinction components of taxonomic diversity: general problems. *Paleobiology*. 26(S4):74–102.
- Foote M. 2003. Origination and extinction through the Phanerozoic: a new approach. *J. Geol.* 111(2):125–148.
- Foote M. 2007. Symmetric waxing and waning of invertebrate genera. *Paleobiology*. 33(4):517–529.
- Foote M., Cooper R.A., Crampton J.S., Sadler P.M. 2018. Diversity-dependent evolutionary rates in early Palaeozoic zooplankton. *Proc. R. Soc. B.* 285:20180122.
- Foote M., Sadler P.M., Cooper R.A., Crampton J.S. 2019. Completeness of the known graptoloid palaeontological record. *J. Geol. Soc.* 176(6):1038–1055.
- Fourment M., Magee A.F., Whidden C., Bilge A., Matsen F.A., Minin V.N. 2020. 19 dubious ways to compute the marginal likelihood of a phylogenetic tree topology. *Syst. Biol.* 69(2):209–220.
- Friedman M., Feilich K.L., Beckett H.T., Alfaro M.E., Faircloth B.C., Černý D., Miya M., Near T.J., Harrington R.C. 2019. A phylogenomic framework for pelagiarian fishes (Acanthomorpha: Percomorpha) highlights mosaic radiation in the open ocean. *Proc. R. Soc. B.* 286:20191502.
- Gauthier J.A. 1986. Saurischian monophyly and the origin of birds. *Mem. Calif. Acad. Sci.* 8:1–55.
- Gavryushkina A., Welch D., Stadler T., Drummond A.J. 2014. Bayesian inference of sampled ancestor trees for epidemiology and fossil calibration. *PLoS Comp. Biol.* 10(12):e1003919.
- Han F., Forster C.A., Xu X., Clark J.M. 2018. Postcranial anatomy of *Yinlong downsi* (Dinosauria: Ceratopsia) from the Upper Jurassic Shishugou Formation of China and the phylogeny of basal ornithischians. *J. Syst. Palaeont.* 16(14):1159–1187.
- Hartman S., Mortimer M., Wahl W.R., Lomax D.R., Lippincott J., Lovelace D.M. 2019. A new paravian dinosaur from the Late Jurassic of North America supports a late acquisition of avian flight. *PeerJ*. 7:e7247.
- Harvey P.H., May R.M., Nee S. 1994. Phylogenies without fossils. *Evolution*. 48:523–529.
- Henao Diaz L.F., Harmon L.J., Sugawara M.T.C., Miller E.T., Pennell M.W. 2019. Macroevolutionary diversification rates show time dependency. *Proc. Natl Acad. Sci. USA*. 116:7403–7408.
- Herne M.C., Nair J.P., Evans A.R., Tait A.M. 2019. New small-bodied ornithopods (Dinosauria, Neornithischia) from the Early Cretaceous Wonthaggi Formation (Strzelecki Group) of the Australian–Antarctic rift system, with revision of *Qantassaurus intrepidus* Rich and Vickers-Rich, 1999. *J. Paleont.* 93(3):543–584.
- Höhna S. 2015. The time-dependent reconstructed evolutionary process with a key-role for mass-extinction events. *J. Theor. Biol.* 380:321–331.
- Höhna S., Stadler T., Ronquist F., Britton T. 2011. Inferring speciation and extinction rates under different sampling schemes. *Mol. Biol. Evol.* 28:2577–2589.
- Holland S.M. 2016. The non-uniformity of fossil preservation. *Phil. Trans. R. Soc. B.* 371(1699): 20150130.
- Hopkins M.J. 2011. Species-level phylogenetic analysis of ptercephaliids (Trilobita, Cambrian) from the Great Basin, western USA. *J. Paleont.* 85(6):1128–1153.
- Hunt G., Slater G.J. 2016. Integrating paleontological and phylogenetic approaches to macroevolution. *Ann. Rev. Ecol. Evol. Syst.* 47:189–213.
- Jablonski D. 1986. Background and mass extinctions: the alternation of macroevolutionary regimes. *Science*. 231(4734):129–133.
- King B., Rücklin M. 2020. Tip dating with fossil sites and stratigraphic sequences. *PeerJ*. 8:e9368.
- Kodandaramaiah U., Murali G. 2018. What affects power to estimate speciation rate shifts? *PeerJ*. 6:e5495.
- Lanfear R., Hua X., Warren D.L. 2016. Estimating the effective sample size of tree topologies from Bayesian phylogenetic analyses. *Genome Biol. Evol.* 8(8):2319–2332.
- Langer M.C., Ezcurra M.D., Rauhut O.W.M., Benton M.J., Knoll F., McPhee B.W., Novas F.E., Pol D., Brusatte S.L. 2017. Untangling the dinosaur family tree. *Nature*. 551:E1–E3.
- Laurent S., Robinson-Rechavi M., Salamin N. 2015. Detecting patterns of species diversification in the presence of both rate shifts and mass extinctions. *BMC Evol. Biol.* 15:157.
- Liow L.H., Finarelli J.A. 2014. A dynamic global equilibrium in carnivoran diversification over 20 million years. *Proc. R. Soc. B.* 281:20132312.
- Liow L.H., Quental T.B., Marshall C.R. 2010. When can decreasing diversification rates be detected with molecular phylogenies and the fossil record? *Syst. Biol.* 59(6):646–659.
- Lloyd G.T. 2012. A refined modelling approach to assess the influence of sampling on palaeobiodiversity curves: new support for declining Cretaceous dinosaur richness. *Biol. Lett.* 8(1):123–126.
- Lloyd G.T., Davis K.E., Pisani D., Tarver J.E., Ruta M., Sakamoto M., Hone D.W.E., Jennings R., Benton M.J. 2008. Dinosaurs and the Cretaceous Terrestrial Revolution. *Proc. R. Soc. B.* 275(1650):2483–2490.
- Lloyd G.T., Slater G.J. 2021. A total-group phylogenetic metatree for Cetacea and the importance of fossil data in diversification analyses. *Syst. Biol.* doi:10.1093/sysbio/syab002.
- Louca S., Pennell M.W. 2020. Extant timetrees are consistent with a myriad of diversification histories. *Nature*. 580(7804):502–505.
- Madzia D., Boyd C.A., Mazuch M. 2018. A basal ornithopod dinosaur from the Cenomanian of the Czech Republic. *J. Syst. Palaeont.* 16(11):967–979.
- Malliet O., Hartig F., Morlon H. 2019. A model with many small shifts for estimating species-specific diversification rates. *Nature Ecol. Evol.* 3:1086–1092.
- Marshall C.R. 2017. Five palaeobiological laws needed to understand the evolution of the living biota. *Nature Ecol. Evol.* 1:0165.
- Marshall C.R. 2019. Using the fossil record to evaluate timetree timescales. *Front. Genet.* 10:1049.
- May M.R., Höhna S., Moore B.R. 2016. A Bayesian approach for detecting the impact of mass-extinction events on molecular phylogenies when rates of lineage diversification may vary. *Methods Ecol. Evol.* 7:947–959.
- Mitchell J.S., Etienne R.S., Rabosky D.L. 2018. Inferring diversification rate variation from phylogenies with fossils. *Syst. Biol.* 68(1):1–18.
- Mitchell J.S., Rabosky D.L. 2017. Bayesian model selection with BAMM: effects of the model prior on the inferred number of diversification shifts. *Methods Ecol. Evol.* 8:37–46.
- Moore B.R., Höhna S., May M.R., Rannala B., Huelsenbeck J.P. 2016. Critically evaluating the theory and performance of Bayesian analysis of macroevolutionary mixtures. *Proc. Natl Acad. Sci. USA*. 113:9569–9574.
- Morlon H. 2014. Phylogenetic approaches for studying diversification. *Ecol. Lett.* 17:508–525.
- Morlon H., Parsons T.L., Plotkin J.B. 2011. Reconciling molecular phylogenies with the fossil record. *Proc. Natl Acad. Sci. USA*. 108:16327–16332.

- Moyle R.G., Filardi C.E., Smith C.E., Diamond J. 2009. Explosive Pleistocene diversification and hemispheric expansion of a “great speciator”. *Proc. Natl Acad. Sci. USA*. 106(6):1863–1868.
- Nee S., May R.M., Harvey P.H. 1994. The reconstructed evolutionary process. *Philos. Trans. R. Soc. B*. 344(1309):305–311.
- Pannetier T., Martinez C., Bunnefeld L., Etienne R.S. 2021. Branching patterns in phylogenies cannot distinguish diversity-dependent diversification from time-dependent diversification. *Evolution*. 75(1):25–38.
- Parins-Fukuchi C. 2021. Morphological and phylogeographic evidence for budding speciation: an example in hominins. *Biol. Lett.* 17(1):20200754.
- Paterson J.R., Edgecombe G.D., Lee M.S.Y. 2019. Trilobite evolutionary rates constrain the duration of the Cambrian explosion. *Proc. Natl Acad. Sci. USA*. 116(10):4394–4399.
- Pires M.M., Rankin B.D., Silvestro D., Quental T.B. 2018. Diversification dynamics of mammalian clades during the K–Pg mass extinction. *Biol. Lett.* 14:20180458.
- Plummer M., Best N., Cowles K., Vines K. 2006. CODA: convergence diagnosis and output analysis for MCMC. *R News*. 6:7–11.
- Pol D., Rauhut O.W.M., Becerra M. 2011. A Middle Jurassic heterodontosaurid dinosaur from Patagonia and the evolution of heterodontosaurids. *Naturwiss.* 98:369–379.
- Pybus O.G., Harvey P.H. 2000. Testing macro-evolutionary models using incomplete molecular phylogenies. *Proc. R. Soc. B*. 267(1459):2267–2272.
- Quental T.B., Marshall C.R. 2009. Extinction during evolutionary radiations: reconciling the fossil record with molecular phylogenies. *Evolution*. 63:3158–3167.
- Quental T.B., Marshall C.R. 2010. Diversity dynamics: molecular phylogenies need the fossil record. *Trends Ecol. Evol.* 25:434–441.
- Quental T.B., Marshall C.R. 2011. The molecular phylogenetic signature of clades in decline. *PLoS One*. 6(10):e25780.
- R Core Team. 2019. R: a language and environment for statistical computing. v3.5.3. Vienna (Austria): R Foundation for Statistical Computing. Available from: <http://www.R-project.org/>.
- Rabosky D.L. 2006. Likelihood methods for detecting temporal shifts in diversification rates. *Evolution*. 60(6):1152–1164.
- Rabosky D.L. 2010. Extinction rates should not be estimated from molecular phylogenies. *Evolution*. 64:1816–1824.
- Rabosky D.L. 2014. Automatic detection of key innovations, rate shifts, and diversity-dependence on phylogenetic trees. *PLoS One*. 9(2):e89543.
- Rabosky D.L. 2016. Challenges in the estimation of extinction from molecular phylogenies: a response to Beaulieu and O’Meara. *Evolution*. 70:218–228.
- Rabosky D.L. 2019. Phylogenies and diversification rates: variance cannot be ignored. *Syst. Biol.* 68(3):538–550.
- Rabosky D.L., Grundler M.C., Anderson C.J.R., Title P.O., Shi J.J., Brown J.W., Huang H., Larson J.G. 2014. BAMMtools: an R package for the analysis of evolutionary dynamics on phylogenetic trees. *Methods Ecol. Evol.* 5:701–707.
- Rabosky D.L., Mitchell J.S., Chang J. 2017. Is BAMM flawed? Theoretical and practical concerns in the analysis of multi-rate diversification models. *Syst. Biol.* 66(4):477–498.
- Rabosky D.L., Slater, G.J., Alfaro M.E. 2012. Clade age and species richness are decoupled across the eukaryotic tree of life. *PLoS Biol.* 10(8):e1001381.
- Raia P., Carotenuto F., Mondanaro A., Castiglione S., Passaro F., Saggese F., Melchionna M., Serio C., Alessio L., Silvestro D., Fortelius M. 2016. Progress to extinction: increased specialisation causes the demise of animal clades. *Sci. Rep.* 6:30965.
- Raup D.M. 1985. Mathematical models of cladogenesis. *Paleobiology*. 11(1):42–52.
- Sakamoto M., Benton M.J., Venditti C. 2016. Dinosaurs in decline tens of millions of years before their final extinction. *Proc. Natl Acad. Sci. USA*. 113:5036–5040.
- Schluter D., Pennell M.W. 2017. Speciation gradients and the distribution of biodiversity. *Nature*. 546:48–55.
- Sepkoski J.J. 1975. Stratigraphic biases in the analysis of taxonomic survivorship. *Paleobiology*. 1(4):343–355.
- Sepkoski J.J. 1998. Rates of speciation in the fossil record. *Philos. Trans. R. Soc. B*. 353(1366):315–326.
- Silvestro D., Cascales-Miñana B., Bacon C.D., Antonelli A. 2015. Revisiting the origin and diversification of vascular plants through a comprehensive Bayesian analysis of the fossil record. *New Phytol.* 207(2):425–436.
- Silvestro D., Salamin N., Antonelli A., Meyer X. 2019. Improved estimation of macroevolutionary rates from fossil data using a Bayesian framework. *Paleobiology*. 45(4):546–570.
- Silvestro D., Salamin N., Schnitzler J. 2014a. PyRate: a new program to estimate speciation and extinction rates from incomplete fossil data. *Methods Ecol. Evol.* 5(10):1126–1131.
- Silvestro D., Schnitzler J., Liow L.H., Antonelli A., Salamin N. 2014b. Bayesian estimation of speciation and extinction from incomplete fossil occurrence data. *Syst. Biol.* 63(3):349–367.
- Silvestro D., Warnock R.C.M., Gavryushkina A., Stadler T. 2018. Closing the gap between palaeontological and neontological speciation and extinction rate estimates. *Nat. Commun.* 9:5237.
- Simões T.R., Caldwell M.W., Pierce S.E. 2020. Sphenodontian phylogeny and the impact of model choice in Bayesian morphological clock estimates of divergence times and evolutionary rates. *BMC Biol.* 18(1):191.
- Smiley T.M. 2018. Detecting diversification rates in relation to preservation and tectonic history from simulated fossil records. *Paleobiology*. 44(1):1–24.
- Stadler T. 2011a. Simulating trees with a fixed number of extant species. *Syst. Biol.* 60(5):676–684.
- Stadler T. 2011b. Mammalian phylogeny reveals recent diversification rate shifts. *Proc. Natl Acad. Sci. USA*. 108:6187–6192.
- Stamatakis A. 2014. RAXML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 30:1312–1313.
- Stanley S.M. 2007. An analysis of the history of marine animal diversity. *Paleobiology*. 33(54):1–55.
- Starrfelt J., Liow L.H. 2016. How many dinosaur species were there? Fossil bias and true richness estimated using a Poisson sampling model. *Philos. Trans. R. Soc. B*. 371:20150219.
- Tennant J.P., Chiarenza A.A., Baron M. 2018. How has our knowledge of dinosaur diversity through geologic time changed through research history? *PeerJ*. 6:e4417.
- Wagner P.J. 2019. On the probabilities of branch durations and stratigraphic gaps in phylogenies of fossil taxa when rates of diversification and sampling vary over time. *Paleobiology*. 45(1):30–55.
- Wagner P.J., Marcot J.D. 2013. Modelling distributions of fossil sampling rates over time, space and taxa: assessment and implications for macroevolutionary studies. *Methods Ecol. Evol.* 4:703–713.
- Wang S.C., Dodson P. 2006. Estimating the diversity of dinosaurs. *Proc. Natl Acad. Sci. USA*. 103:13601–13605.
- Warnock R.C.M., Heath T.A., Stadler T. 2020. Assessing the impact of incomplete species sampling on estimates of speciation and extinction rates. *Paleobiology*. 46(2):137–157.