Testing extinction events and temporal shifts in diversification and fossilization rates through the skyline Fossilized Birth-Death (FBD) model: The example of some mid-Permian synapsid extinctions

Gilles Didier, Michel Laurin

Testing extinction events and temporal shifts in diversification and fossilization rates through the skyline FBD model: the example of some mid-Permian synapsid extinctions

Gilles Didier[1] and Michel Laurin[2]

[1]IMAG, Univ Montpellier, CNRS, Montpellier, France

[2]CR2P ("Centre de Recherches sur la Paléobiodiversité et les Paléoenvironnements"; UMR 7207), CNRS/MNHN/UPMC, Sorbonne Université, Muséum National d'Histoire Naturelle, Paris, France

March 7, 2024

Abstract

In the last decade, the fossilized-birth-death process has yielded interesting clues about the evolution of biodiversity through time. To facilitate such studies, we extend our method to compute the probability density of phylogenetic trees of extant and extinct taxa in which the only temporal information is provided by the fossil ages (i.e., without the divergence times) in order to deal with the piecewise constant fossilized-birth-death process, known as the "skyline FBD", which allows rates to change between pre-defined time intervals, as well as modelling extinction events at the bounds of these intervals.

We develop approaches based on this method to assess hypotheses about the diversification process and to answer questions such as "Does a massive extinction occur at this time?" or "Is there a change in the fossilization rate between two given periods?". Our software can also yield Bayesian and maximum likelihood estimates of the parameters of the skyline FBD model under various constraints.

These approaches are applied to a simulated dataset in order to test their ability to answer the questions above.

Finally, we study an updated dataset of Permo-Carboniferous synapsids, to get additional insights into the dynamics of biodiversity change in three clades (Ophiacodontidae, Edaphosauridae, and Sphenacodontidae) in the Pennsylvanian (Late Carboniferous) and Cisuralian (Early Permian), and to assess support for end-Sakmarian (or Artinskian) and end-Cisuralian mass extinction events discussed in previous studies.

Keywords: fossilized-birth-death model, fossil ages, divergence times, probability distribution, Permo-Carboniferous synapsids, mass extinction events

1 Introduction

Discovering the dynamics of taxonomic diversification, in particular identifying phenomena such as evolutionary radiations or mass extinction events, is a central question of evolutionary biology. These events are brief periods, nearly instantaneous in geological terms (less than 1 Ma, often much less, like a few tens of thousands of years), in which the extinction rate is much higher than the background rate (Benton 2003; Arens and West 2008; Burgess *et al.* 2014; Day *et al.* 2015; Tabor *et al.* 2020). Previous paleontological studies on mass extinction events have used two main approaches. The first emphasized a visual assessment of detailed, high-resolution stratigraphic ranges, sometimes compiled from fresh field data (Ward *et al.* 2005; Smith and Botha-Brink 2014; Kammerer *et al.* 2023). These studies can give precise clues about the timing of extinction events, but did not provide a statistical basis to assess their significance. The second type of studies has relied on quantitative analyses of such data (Viglietti *et al.* 2021) based on relatively broad time bins and on counts of taxa that cross time bin boundaries (Foote 2003; Alroy 2015). Such analyses, based on time bins that typically represent a geological stage (about 5 to 10 Ma), and sometimes, a greater timespan, allow fluctuations in extinction rates among discrete time bins (temporal intervals) to be quantitatively assessed, but they are not designed to test hypotheses

about short-term events, such as mass extinctions, most of which (e.g., P/Tr and K/Pg crises) appear to be short in geological times (certainly much shorter than most geological stages); for instance, the P/Tr crisis appears to have lasted a mere 60 ky (Burgess and Bowring 2015). Variations in diversification rates over time can also be studied from the fossil record using these same methods (Foote 2003; Alroy 2015).

Most of the sophisticated numerical approaches to study the evolution of biodiversity over time in a phylogenetic context were developed by systematists who work only (or mostly) on extant taxa (Nee *et al.* 1994). In particular, the few model-based approaches to detect mass extinction events do not take into account fossils (May *et al.* 2016; Culshaw *et al.* 2019), even though simulations show that incorporating fossils helps to estimate extinction rates (Paradis 2004; Didier *et al.* 2012, 2017). These approaches suffer from serious identifiability issues in diversification models of extant taxa (Quental and Marshall 2010; Louca and Pennell 2020; Morlon *et al.* 2022). Even though Louca *et al.* (2021) and Beaulieu and O'Meara (2023) reported identifiability issues with methods incorporating fossils in some situations, our simulation study suggests that the parameters of a skyline FBD model can actually be inferred. We further discuss this point below. Similarly, the question of detecting shifts in diversification rates was intensively studied using timetrees of extant taxa (Slowinski and Guyer 1993; Rabosky 2006; Alfaro *et al.* 2009; Wertheim and Sanderson 2010; Stadler 2011; Morlon 2014; Rabosky 2014; Rabosky *et al.* 2017; Morlon *et al.* 2016).

The models used to handle variable rate shifts from datasets with fossils (as opposed to extant taxa datasets, as in the methods cited above) generally fall within the broader category of Markov switching models, also known as regime switching models (Hamilton 1989). These models are defined from a finite number of diversification regimes, which are in one-to-one correspondence with a set of latent states following a first-order Markov chain. At any time of the diversification process, each lineage is associated to a latent state that defines its current diversification regime and switches to another state/regime following an independent time continuous Markov process or a Poisson process (Mitchell *et al.* 2019; Barido-Sottani *et al.* 2020; Beaulieu and O'Meara 2023). Another research line is provided by Silvestro *et al.* (2014a,b, 2019) who developed a global Bayesian framework based on a slighly different model implemented in the software *PyRate*, to estimate variable fossilization and diversification rates, along with confidence intervals from datasets with fossils. All of these approaches, assuming a constant fossilization rate, enable the estimation of variable speciation and extinction rates through time, which is a different question from the one addressed in this work.

Though we are also interested in estimating (variable) past cladogenesis, extinction and fossilization rates, we focus here on testing paleontological hypotheses about the evolution of taxonomic diversity, notably through possible mass extinction events. Namely, we seek to answer questions such that "Does an extinction event occur at this time?" or "Does the fossilization rate change between these two periods?" etc. To this end, we propose an approach based on the piecewise-constant Fossilized Birth-Death model, called the *skyline FBD* (Stadler *et al.* 2013; Gavryushkina *et al.* 2014; Zhang *et al.* 2016). This model requires to split the diversification time into pre-defined intervals, during which the rates are constant but (i) they may change between the intervals and (ii) the lineages are sampled according to a survival probability at the end of each interval, which allows to model extinction events of variable intensity. Unlike Mitchell *et al.* (2019); Barido-Sottani *et al.* (2020); Beaulieu and O'Meara (2023), lineages do not have specific extinction and speciation (cladogenesis) rates. These rates, as well the fossilization rate, change for all of them simultaneously, which is more suited to assess hypotheses about global changes in the diversification and mass extinctions. Though it can also perform hypothesis testing, a notable difference with PyRate (Silvestro *et al.* 2014b) is that we use a different model which allows us to deal with mass extinction events. Our approach is implemented in a quite versatile framework, in which users can specify various constraints over the parameters, e.g., force the extinction rate to be the same among all the intervals, or set some of the parameters to particular values, e.g., set the survival probability of an interval to 1 (which means assuming that no extinction event occurred). These constraints serve as formal counterparts to diversification hypotheses. In addition to computing both Bayesian and maximum likelihood estimates of the extinction, speciation, and fossilization rates, as well as the extinction intensities at boundaries between time intervals, the method provides AIC and Akaike weights for 'models with constraints'/'diversification hypotheses' selection.

While the skyline FBD model is already available on the BEAST2 and RevBayes platforms (Bouckaert *et al.* 2019; Höhna *et al.* 2016), we are introducing a new implementation based on an original methodological contribution. Specifically, it stands in a method to compute the probability density of tree topologies with fossils and (possibly) extant taxa, in which the only temporal information used is

the fossil ages, along with the start and end times of the diversification process under the skyline FBD model. This stands in contrast to previous approaches that necessitate the inclusion of divergence times. The computational methodology extends our prior work in Didier (2021) about the standard FBD model and is outlined in Appendix A. Through this new method, we eliminate reliance on divergence times, which act as nuisance parameters in our analyses, and which are always poorly constrained; typically, they have much greater associated age uncertainties than fossils (Sterli *et al.* 2013). This significantly reduces the dimension of parameter space to be sample in the MCMC runs and, more generally, improves the accuracy of the quantities computed by replacing stochastic integration by an exact one.

Our approach is first assessed on a simulated dataset, and it is then applied to an empirical dataset compiled for studying the evolution of biodiversity of ophiacodontids, edaphosaurids and sphenacodontids in the Pennsylvanian (Late Carboniferous) and Cisuralian (Early Permian). The extensive synapsid fossil record documents various macroevolutionary events (Kammerer *et al.* 2014; Sidor 2001), such as the appearance of endothermy near the base of Therapsida (Olivier *et al.* 2017), the transformation of the old, so-called reptilian jaw joint into an impedance-matching middle ear (Allin and Hopson 1992; Bramble 1978), and the emergence of mammals (Rowe 1988). All these are documented in a more or less continuous fossil record from the Late Carboniferous through the late Pleistocene (large gaps exist in the geographical coverage, but some synapsids are known from every geological stage), which shows a spectacular taxonomic diversification from a single (inferred) lineage about 330 Ma (Didier and Laurin 2020) to over 7000 lineages today (Kammerer *et al.* 2014). This fossil record is thus well-suited to document the evolution of biodiversity through time, including extinction events of various magnitudes. In the Permian alone, up to four mass extinction events have been recognized in the synapsid fossil record (Lucas 2017). The first event may have occurred in the Artinskian (Benton 1989) or slightly earlier, in the late Sakmarian, and it may have affected all main synapsid clades except edaphosaurids (Brocklehurst *et al.* 2013). A second event, which may correspond to the extinction of three clades (Ophiacodontidae, Edaphosauridae and Sphenacodontidae) and possibly, a reduction in diversity of two other clades (Varanopidae and Caseasauria), arguably occurred around the Kungurian/Roadian boundary (Sahney and Benton 2008; Brocklehurst 2018), which is also the Cisuralian/Guadalupian boundary. Both of these events are studied in this contribution. The next two extinction events, not tackled here but mentioned for the sake of completeness, are better known. The third event, near the end of the Capitanian, affected several synapsid clades and saw the extinction of dinocephalians. It may be linked to the intensive volcanism of the Emeishan Large Igneous Province (Day *et al.* 2015). Last but not least, the end-Permian event is widely known as the most severe mass extinction event of the Phanerozoic (Viglietti *et al.* 2021; Kammerer *et al.* 2023).

The second of the four Permian potential mass extinction events coincides with one of the most spectacular faunal changes of the Phanerozoic, between the Early and Middle Permian, when most Permo-Carboniferous, ectothermic synapsid clades were replaced, to a large extent, by the endothermic Therapsida, a synapsid clade that includes mammals. The amniote fossil record in the first stages of this replacement was studied intensively, in the south-western USA, by Olson (1965, 1968). Sahney and Benton (2008, p. 764) suggested that this event, sometimes called "Olson's extinction", "had a much more severe impact at the community level than the terminal end-Permian event." Our recent reassessment of this possible mass extinction event, using a higher-resolution stratigraphic scheme and a single-rate FBD model, suggests a gradual decline of Ophiacodontidae, Edaphosauridae and Sphenacodontidae. It reveals neither a dramatic mass extinction event around the Kungurian/Roadian boundary, nor an earlier (end-Sakmarian or Artinskian) event (Didier and Laurin 2021). However, the approach developed here should yield more informative and reliable results, given that at least some rates are expected to vary for a portion of the considered time interval (Pennsylvanian to early Roadian).

The software and datasets presented in this work are available at `https://github.com/gilles-didier/PiecewiseFBD`.

2 Methods

2.1 Birth-death-fossil-sampling model

We model here the diversification and fossilization process by using the FBD model (Stadler 2010; Heath *et al.* 2014). This model assumes that the diversification and fossilization process starts with a

single lineage at the origin time, which is a parameter of the model. While the process is running, each lineage is subject to three possible types of events: speciation (cladogenese), which gives birth to a new lineage, extinction, resulting in its death, and fossilization, meaning that we found a fossil of the lineage dated at the time of the event, each type of event occurring at respective constant rates λ, μ and ψ. At the end of the process, the lineages alive are uniformly sampled with probability ρ.

2.2 Variable rates

Following Gavryushkina *et al.* (2014) and Zhang *et al.* (2016), we allow the diversification and fossilization rates to change a fixed number of times along the process. Namely, we divide the time into intervals in which the rates are constant, but they (may) vary among the intervals. Still following Gavryushkina *et al.* (2014) and Zhang *et al.* (2016), the lineages alive at the end of an interval are uniformly sampled with a probability associated with this interval. Note that the sampling probability has to be understood differently for the interval ending at the present time and the other ones. Namely, the sampling probability at time 0 (the present) accounts for the proportion of lineages/species which are known (or incorporated into a study) while the sampling probabilities for earlier dates are survival probabilities. Figure 1a displays an example of a skyline FBD model in which the time is divided into three intervals ($[350, 250]$, $[250, 200]$ and $[200, 0]$ Ma), each with specific rates and sampling probability at its end. In particular, the sampling probability $\rho = 0.2$ at 250 Ma implies that a lineage alive at this date has only 20% probability to survive beyond that time, which corresponds to a (mass) extinction event at this date. The model of Figure 1a starts with a period of slow diversification ending with a massive extinction, which is followed by a short phase of fast diversification, itself followed by a stable period until the the present time. This model is vaguely inspired from the diversification of crown-tetrapods, which started about 350 Ma ago (Laurin 2010), with the end-Permian event 250 Ma ago, with a new diversification rate in the Triassic, and a third time slice for the last interval (Jurassic-present), but we could also have selected other time slices and mass extinction event parameters. The rates were set to be realistic and in general less favorable than our empirical dataset. Thus, we simulated a fossilization rate of only 0.02, which is much less than the rate estimated in our previous study (Didier and Laurin 2021) on an earlier version of our empirical dataset. Similarly, the speciation rate is lower than the one we previously estimated, so the amount of information provided by the simulated datasets should be lower than what the fossil record supports, at least for Permo-Carboniferous synapsids. Only the speciation rate varies between the intervals of the model. This allows us to test the ability of our skyline FBD model implementation to correctly detect different speciation rates between the time slices, as well as the ability to detect a mass extinction event and correctly estimate its magnitude.

We shall not consider the whole realizations of the skyline FBD process (Fig. 1a); instead, we focus on their *observable* part, which is that can be reconstructed from the present time using data from extant taxa and from the fossil record as depicted in Figure 1b (Didier and Laurin 2020). Our typical data consists of the evolutionary relationships between lineages/taxa (extant or extinct) which are represented as a tree topology and of the fossil ages, which are the only available temporal data used (extant taxa would also contain time information if present, but the empirical dataset used here focuses on the earliest synapsids and does not consider events after the early Roadian). Namely, all the divergence times of the reconstructed phylogeny are assumed to be unknown.

2.3 Probability density

We adapted the method presented in Didier and Laurin (2020); Didier (2021) to compute the probability density of our typical data, that is a tree topology with the starting and ending times of the diversification process and fossil ages (without the divergence times) under the skyline FBD model. Details are provided in Appendix A.

2.4 Model choice

A model specification is the given of the time boundaries of a skyline FBD model and of some set of equality constraints on the parameters of the models. This set may include equality constraints between some of the rates (e.g., we may constraint the fossilization rate to be the same among all intervals of times) and/or direct constraints on some rates to be equal to fixed values (e.g., we may constraint the sampling

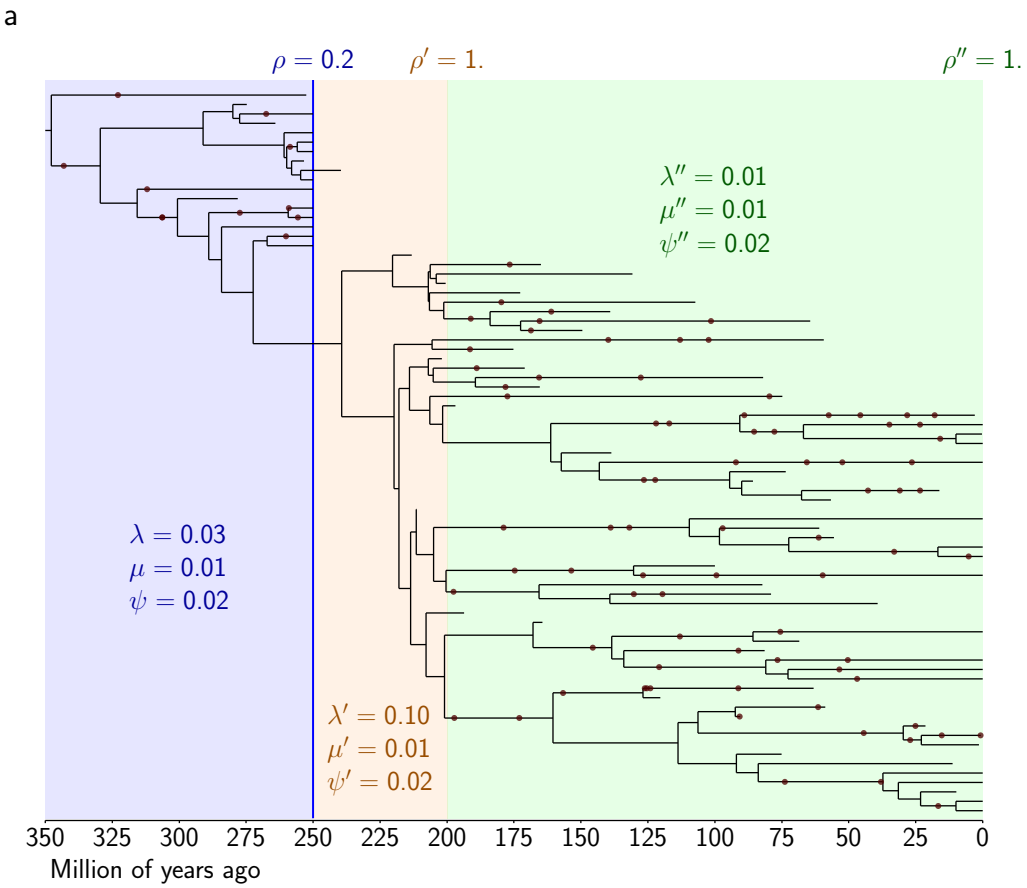

Figure 1: (a) An example of a skyline FBD in which time is divided into three intervals (changes in rates occur at 250 and 200 Ma) and a whole realization of this model where fossils are represented with "●". (b) The observable part of the simulated tree above, in which we replaced the (exact) fossil ages by ranges of constant length, here 5 Ma, uniformly (and randomly) drawn around them to get our simulated dataset.

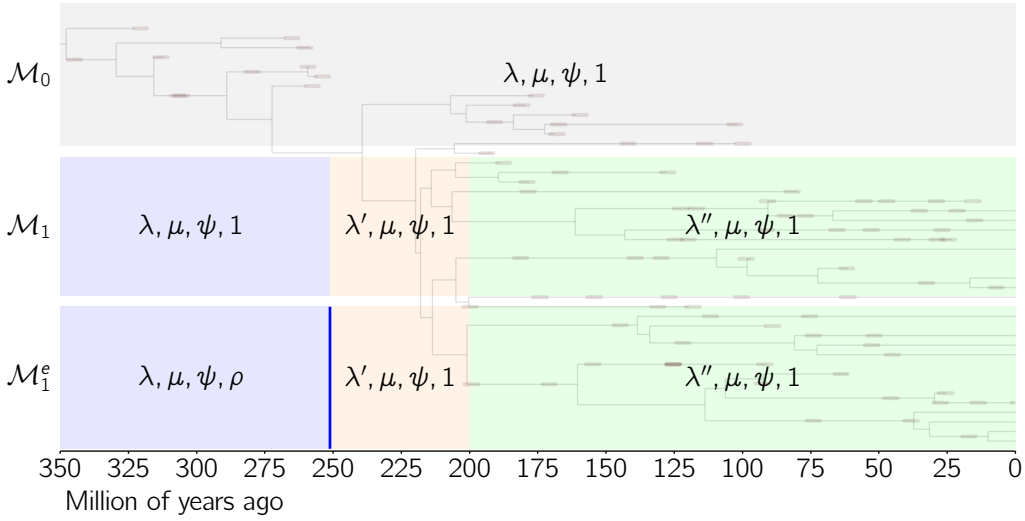

Figure 2: The 3 model specifications considered on the simulated dataset.

probability at the end of intervals to be equal to 1). The constraints included in a model specification basically correspond to a set of assumptions about the diversification process, e.g., constraining the sampling probability to 1 at the end of an interval is the same as assuming no mass extinction event at this time; constraining the extinction rate to be equal between two successive intervals is the same as assuming that it does not change etc. Conversely, various types of hypotheses about the diversification process can be translated into constraints of a model specification.

The parameters associated to a model specification are those that need to be set to a value to get a fully specified skyline FBD model. Figure 2 displays the three model specifications we shall consider in our simulation study. The model specification $\mathcal{M}_0$ has a single time interval $[320, 0]$ with full sampling at the end while the model specifications $\mathcal{M}_1$ and $\mathcal{M}_1^e$ contain three intervals $[350, 250]$, $[250, 200]$ and $[200, 0]$, each with an individual speciation rate but all with the same extinction and fossilization rates. Both for $\mathcal{M}_1$ and $\mathcal{M}_1^e$, the ending survival/sampling probabilities of the two last intervals are set to 1, while the first is to be estimated in $\mathcal{M}_1^e$ and set to 1 in $\mathcal{M}_1$ (Fig. 2). Note that $\mathcal{M}_1$ and $\mathcal{M}_1^e$ only differ from each other in the possibility of having a sampling probability smaller than one (which represents a possible mass extinction event) for $\mathcal{M}_1^e$.

The model specification $\mathcal{M}_1^e$ has 6 parameters λ, λ', λ'', μ, ψ and ρ ($\mathcal{M}_0$ and $\mathcal{M}_1$ have 3 and 5 parameters, respectively). We shall say that a skyline FBD model *satisfies the constraints* of the model specification $\mathcal{M}$ if there is a setting of the parameters of $\mathcal{M}$ which leads to this FBD model. For instance, the model displayed in Figure 1 satisfies the constraints of $\mathcal{M}_1^e$ since by setting $\lambda = 0.03$, $\lambda' = 0.1$, $\lambda'' = 0.01$, $\mu = 0.01$, $\psi = 0.02$, $\rho = 0.2$, we get this model, but it does not satisfy the constraints of neither $\mathcal{M}_0$ nor $\mathcal{M}_1$, since no setting of these model specifications can lead to this more complex model.

The model specifications $\mathcal{M}_1^e$, $\mathcal{M}_1$ and $\mathcal{M}_0$ are nested in the sense that any model satisfying the constraints of $\mathcal{M}_1$ satisfies the constraints of $\mathcal{M}_1^e$ and that any model satisfying the constraints of $\mathcal{M}_0$ satisfies the constraints of $\mathcal{M}_1$ thus those of $\mathcal{M}_1^e$. Note that this property is not required by our approach.

A key question is to determine among a given set of model specifications, which one is the most relevant with regards to a typical dataset. A natural strategy to measure the goodness of fit of a model specification $\mathcal{M}$ with regard to a dataset consists in considering the maximum likelihood of the dataset among all the skyline FBD models that satisfy the constraints of $\mathcal{M}$. Unfortunately, the likelihood of a skyline FBD model with regards to our typical dataset cannot be directly computed since the fossil ages are not given as exact times but as intervals of time. Computing the likelihood of a skyline FBD model requires to integrate numerically or stochastically over all its possible fossil ages, which is far too much time-consuming in a maximum likelihood framework. We thus consider the maximum likelihood of the dataset among all the skyline FBD models and all the possible fossil ages. In practice, we approximate the maximum likelihood by considering the greatest probability density encountered during a MCMC

run (typically with tens of thousands iterations in which the fossil ages are sampled using an uniform distribution between the bounds of the age intervals), which provides a lower bound of the desired quantity and turns out to be far more reliable than the various non-linear optimization algorithms we tried.

In order to measure the accuracy of a model specification $\mathcal{M}$ with regard to a given dataset, we consider the Akaike information criterion (AIC) associated to $\mathcal{M}$, which is defined as $AIC = 2k - 2\ln(L)$, where L the maximum likelihood of the dataset obtained among all the skyline FBD models that satisfy the constraints of $\mathcal{M}$ and all the possible fossil ages, and k is the sum of the number of parameters of the model specification and of the number of fossils in the dataset (Akaike 1998).

2.5 Parameter estimation

By being given such a model specification, we use the MCMC framework presented in Didier and Laurin (2020) to sample both the fossil ages among their stratigraphic ranges and the parameters of the model specification with (improper) uniform distributions under the constraints specified by this model specification. As in Didier and Laurin (2021), this allows us to get the Bayesian posterior distribution of the parameters of a model specification as well as those of the fossil ages.

In addition to their Bayesian posterior distribution, we approximate the maximum likelihood estimation of the parameters of a model specification by the parameters associated to the greatest likelihood encountered during a long MCMC runs, as we did to compute the AIC in Section 2.4. Namely, we maximize the likelihood both on the model specification parameters and on the fossil ages.

2.6 Simulated datasets

The simulated datasets used in this work were obtained in two stages. The first stage simulates a complete realization of a skyline FBD model (i.e., including the parts that cannot be reconstructed from present time using data from extant taxa and from the fossil record), in which the fossil ages are exactly known, like that displayed in Figure 1a. In the second stage, all the parts of this realization that are not observable from the present time (either as extant tips or through their fossil record) are discarded, and so are the divergence times. We also replace each fossil age by an interval of time of constant length uniformly drawn around it (namely, the lower bound of the fossil interval is drawn in such way that the simulated fossil age follows a uniform distribution in this interval, which is what is assumed on the empirical datasets), in order to obtain datasets similar to the empirical ones (Fig. 1b).

All the simulated datasets below were obtained with fossil time interval lengths of 5 Ma, which corresponds to the average length of the stratigraphic ranges of the empirical dataset below. Simulated datasets are also filtered with regards to their size for technical reasons, mainly related to computational times. We kept only datasets with tree that included between 100 and 500 nodes (ancestral and tips).

2.7 Empirical data on Permo-Carboniferous eupelycosaurs

We updated the ages of all fossils of ophiacodontids, edaphosaurids and sphenacodontids included in our recent study on this theme (Didier and Laurin 2021). To update the ages of these 50 terminal taxa, we used a new, integrated dating scheme. This was necessary because the age of various fossiliferous horizons is both uncertain and debated, and in some circumstances, simply accepting at face value the geological age (or geological stage assignment) given in the descriptive literature might have resulted in a conflict between the relative age of the fossils obtained from the literature and the lithostratigraphic relationships. Thus, we used the literature (especially recent papers) to produce a single, hopefully coherent stratigraphic scheme that includes all the formations that yielded the fossils recorded in our Permo-Carboniferous synapsid database. For this task, the careful temporal calibration of the German Rotliegend by Menning *et al.* (2022) proved extremely useful, for two main reasons. First, this study used several criteria to assess relative and absolute ages. Second, it provided correlations with the Texas Cisuralian formations and with the global geological stages. Figure 23 from Menning *et al.* (2022) formed the core of our stratigraphic correlations, but we updated the absolute ages to reflect a more recent (2020, rather than 2016) geological time scale (Gradstein and Ogg 2020). Other conflicts in the literature not addressed in Menning *et al.* (2022) or in Laurin and Hook (2022) were resolved through a literature search and critical reading of the relevant papers, by comparing supporting evidence and the analyses

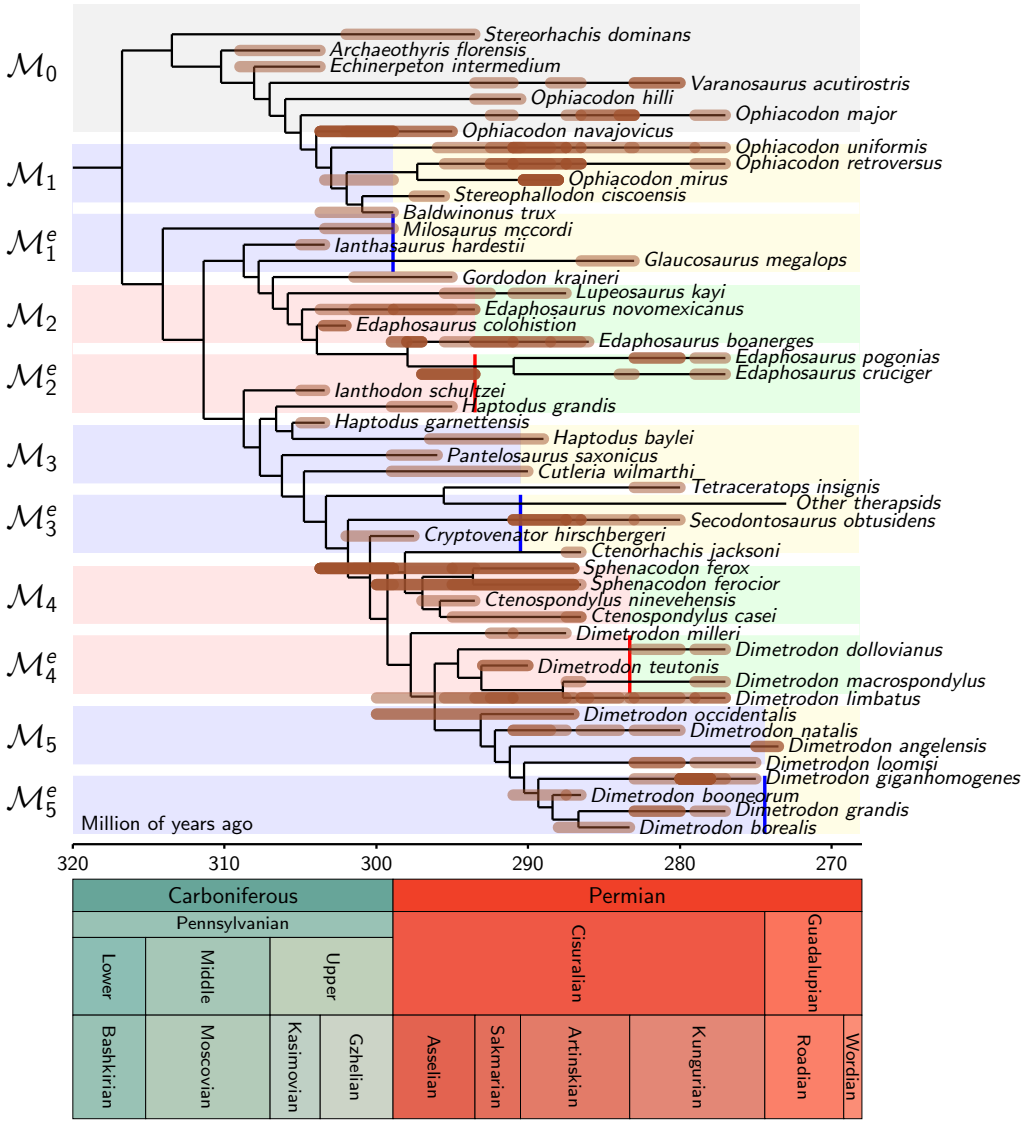

Figure 3: One of the 100 equiparsimonious trees of the empirical dataset with the fossil stratigraphic ages and the 11 model specifications considered in this study. All model specifications have two intervals with specific rates: λ, μ, ψ and λ', μ', ψ'. Model specifications with extinction, i.e. $\mathcal{M}_{1-5}^e$, have an additional parameter, the survival/sampling probability ρ at the end of the first interval, which is set to 1 for the other model specifications.

that were used to date various formations. In general, recent studies were preferred over older ones, and those based on extensive datasets were preferred over studies based on smaller datasets.

For the Texas Cisuralian formations, equivalency between the old nomenclature used in the paleontological literature through the 1980s and the new nomenclature was obtained from figure 3 from Hentz (1988). Thus, for instance, the Putnam Formation in the old nomenclature is equivalent to roughly the last third of the Archer City Formation. The age of the last North American Permian strata that yielded continental vertebrates, namely the San Angelo Formation (Pease River Group, Texas) and the Chickasha Formation (El Reno Group, Oklahoma) has long been controversial. While Olson (1962, 1965) placed these formations in the Guadalupian, this age assignment was controversial, partly because Olson based his interpretations on fragmentary fossils whose taxonomic affinities were difficult to assess. Partly because of this, Lucas (2004) and Schneider *et al.* (2020) argued that the San Angelo and Chickasha formations were of Cisuralian age. However, a recent study dedicated to this topic and based on several criteria concluded that the San Angelo was most likely basal Roadian (possibly its base extended down into the latest Kungurian, the San Angelo thus straddling both stages) and the Chickasha is somewhat higher in the Roadian, but the vertebrate-bearing levels are probably in the early Roadian (Laurin and Hook 2022). This reassessment of the age of these two formations excluded the contentious fossils that Olson interpreted as therapsids, which are still in need of a thorough reassessment, but it considered the presence of *Macroleter agilis* (among many other vertebrates) in the Chickasha Formation (Reisz and Laurin 2001, 2002).

We tried to account for all sources of uncertainty in the age of the fossils included in our database, but some constraints forced us to sometimes provide fairly narrow time intervals for some fossils. Namely, in a given sedimentary basin in which several synapsid-bearing formations are present, the age of a fossil found in a given formation must necessarily be older than that of a fossil from an overlying formation. Thus, for instance, *Ctenorhachis jacksoni* comes from the Petrolia Formation (Hook and Hotton III 1991), which occurs between the fossiliferous Nocona and Waggoner Ranch formations. Thus, we used the age for the Petrolia (288.5 to 286.5 Ma) suggested by Menning *et al.* (2022, fig. 23); the 2016 and 2020 GSTs give similar ages for the relevant (Artinskian) age boundaries. However, given that under the older stratigraphic nomenclature, *Ctenorhachis jacksoni* comes from the Belle Plains Formation, and that this corresponds approximately to the top half of the Petrolia (Hentz 1988), we constrained the age of *Ctenorhachis jacksoni* to 287.5 to 286.5 Ma. The simultaneous use of the old and new stratigraphic nomenclature thus allows greater precision in assessing the relative age of fossils. Ignoring the old stratigraphic nomenclature would have discarded stratigraphic information because fossils formerly assigned to the Putnam and Moran formations, now both considered part of Archer City Formation, would erroneously have been assigned to the same temporal interval. Of course, this probably gives an illusory impression of precision of the age of some strata, but in such an analysis, it is presumably more important to have correct relative ages. Future analyses can be refined as the geological time scale improves. Also, the use of our mathematical modeling and statistical approach should mitigate this effect, to the extent that we compute probability densities for all cladogenesis and extinction times, rather than try to report exact values that would give an illusory impression of precision.

We used the same set of 100 equiparsimonious trees as in Didier and Laurin (2021). These provided only topological information. Nodal ages (divergence times) can be computed through our method, but this is not needed to compute the FBD model parameters.

The hypotheses in the literature determine which models are tested (Fig. 3). Given that extinction events have been proposed to have occurred in the Sakmarian (Brocklehurst *et al.* 2013) or Artinskian (Benton 1989), but probably not both (these suggestions probably reflect uncertainty in the timing of the event, rather than the possibility of two events occurring in close succession), we devised model specifications in which rates change and mass extinction events may occur at the boundaries between these stages. However, our previous study (Didier and Laurin 2021) suggested that extinctions occurred later (and failed to provide evidence for a Sakmarian or Artinskian event), so we also produced models designed to test a more recent slow-down in origination rates and a possible extinction event at the end of the Kungurian, which is just after the last occurrences of Ophiacodontidae and Edaphosauridae are documented (Didier and Laurin 2021). Finally, we included a model specification to test if an extinction event or the deceleration in origination rates occurred across the Permo/Carboniferous boundary because this is the highest-order geochronological boundary included in the studied interval, and many such boundaries (e.g., P/Tr; K/Pg) coincide with mass extinction events, or at least, major faunal and floral turnover. These yield five model specifications differing by when the boundary is placed between the

model specification	$\mathcal{M}_0$	$\mathcal{M}_1$	$\mathcal{M}_1^e$
number of parameters	3	5	6
log max likelihood	-600.14	-592.95	-584.13
AIC	1386.28	1375.89	<u>1360.27</u>
w(AIC)	0.00	0.00	<u>1.00</u>

Table 1: Fit of the models on the simulated dataset displayed in Figure 1b (the AIC computation takes into account its number of fossils, i.e. 90). The model used to simulated the dataset indeed has the greatest Akaike weight.

two time slices. For each model specification, we tested variants with, and without, an extinction event (of a magnitude to be evaluated by our software) between both time slices, which yielded ten models.

Note that while the skyline FBD model can be used to estimate model parameters in several (rather than two, as done here) time slices, we believe that estimating parameters in two time slices is a good compromise between getting an idea of how diversification changed over time (the more time slices, the better), and having enough data to estimate the parameters accurately and being able to discriminate between models. In addition, the literature and our previous findings (Didier and Laurin 2021) suggest a simple, two-phase history of diversification of synapsids in the Late Carboniferous and Early Permian: an evolutionary radiation, followed by a decline in diversity. In this respect, nothing would justify recognizing more than two time slices. In addition, we included the single time-slice FBD model used in our previous publications, for a total of eleven tested model specifications (Fig. 3).

In the discussion, we briefly consider another model specification, which is not part of our main analysis, in order to test an additional hypothesis, somehow suggested by the results obtained from the model specifications above. This model specification has the same time slices as $\mathcal{M}_2$ but with a mass extinction a bit before the end of the Kungurian (this required recognizing a third time slice, whose rates were constrained to be equal to those of the previous time slice, because the thin, additional time slice contains a single taxon, *Dimetrodon angelensis*, and this would not have been sufficient to estimate the FBD model parameters).

Since it cannot lead to any confusion, we use the same notations for the model specifications of the simulated and the empirical datasets. Namely, $\mathcal{M}_0$, $\mathcal{M}_1$ and $\mathcal{M}_1^e$ refer to the model specifications displayed in Figure 2 in the simulated dataset context and to those displayed in Figure 3 in the empirical study.

3 Results

3.1 Simulated dataset

3.1.1 Model choice

We applied the approach presented in Section 2.4 to the simulated dataset of Figure 1b in order to select the most accurate model specification among $\mathcal{M}_0$, $\mathcal{M}_1$ and $\mathcal{M}_1^e$. Results are presented in Table 1, where we display the AIC of each model as well as their Akaike weights (Burnham and Anderson 1998; Wagenmakers and Farrell 2004). The smallest (best) AIC is achieved by model $\mathcal{M}_1^e$ with an Akaike weight of 1 (actually almost 1, but weights are rounded to two decimal places).

Under our model selection procedure, model $\mathcal{M}_1^e$ is without ambiguity the best fitting one for the dataset of Figure 1b, which is perfectly consistent with the way it was simulated (Fig. 1a).

In order to assess in what extent our model selection procedure was able to distinguish between model specifications $\mathcal{M}_0$, $\mathcal{M}_1$ and $\mathcal{M}_1^e$, we simulated (a) 500 datasets under the model presented in Figure 1a and used to simulated the dataset of Figure 1b, which satisfies the constraints of $\mathcal{M}_1^e$, (b) 500 datasets under the skyline FBD model maximizing the likelihood of the simulated dataset and satisfying the constraints of $\mathcal{M}_1$ and (c) 500 datasets under the skyline FBD model maximizing the likelihood of the simulated dataset and satisfying the constraints of $\mathcal{M}_0$.

The distributions of the Akaike weights of models $\mathcal{M}_1^e$, $\mathcal{M}_1$ and $\mathcal{M}_0$ are plotted in Figure 4 for the three sets of 500 simulated datasets (a), (b) and (c).

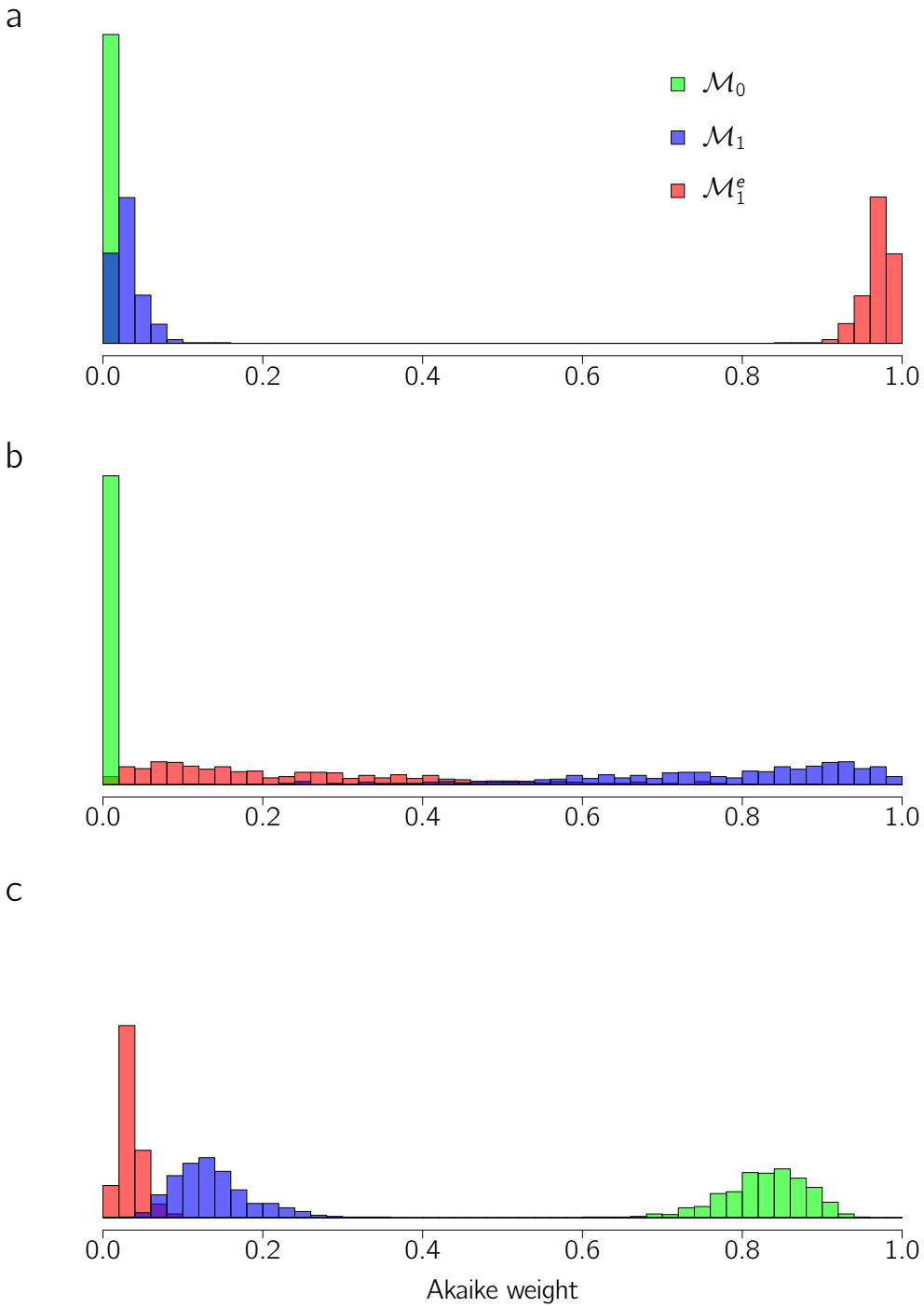

Figure 4: Akaike weight distributions of the three model specifications considered for the simulated dataset (distributions of $\mathcal{M}_0$, $\mathcal{M}_1$ and $\mathcal{M}_1^e$ are displayed in green, blue and red, respectively) obtained from 500 trees simulated (a) under the model displayed in Figure 1, which satisfies the constraints of $\mathcal{M}_1^e$, (b) under the maximum likelihood model satisfying the constraints of $\mathcal{M}_1$ and (c) under the maximum likelihood model satisfying the constraints of $\mathcal{M}_0$.

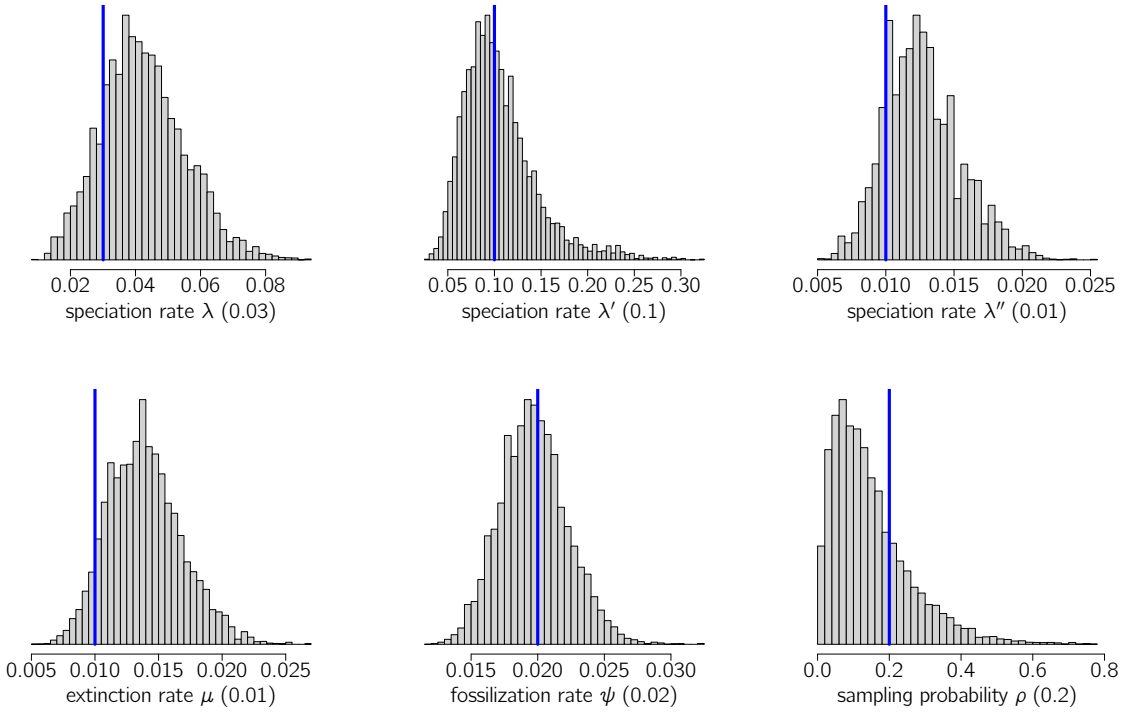

Figure 5: Posterior distributions of the parameters of the model specification $\mathcal{M}_1^e$ obtained from the simulated dataset displayed in Figure 1b.

We observe that for the datasets simulated under the constraints of $\mathcal{M}_1^e$, the Akaike weight of $\mathcal{M}_1^e$ is always the greatest while that of $\mathcal{M}_0$ is negligible and that $\mathcal{M}_1$ may have a small support (Fig. 4a).

The Akaike weight of $\mathcal{M}_0$ is still negligible for the datasets simulated under the constraints of $\mathcal{M}_1$. For most of the simulations (more than 90%), the Akaike weight of $\mathcal{M}_1$ is the greatest and always has a substantial support, but the weight of $\mathcal{M}_1^e$, which is "$\mathcal{M}_1$ with extinction", may be close and sometimes greater, in a small proportion of the simulated datasets (less than 10%).

Last, for the datasets simulated under the model satisfying the constraints of $\mathcal{M}_0$, the greatest Akaike weight is always that of $\mathcal{M}_0$, well above those of $\mathcal{M}_1^e$ and $\mathcal{M}_1$, which are small but not negligible (Fig. 4c).

In conclusion, our model selection procedure perfectly distinguishes the model specification associated to the model used to simulated the data, except in the case where they were simulated using $\mathcal{M}_1$, where $\mathcal{M}_1$ and $\mathcal{M}_1^e$ may have similar supports in a few cases.

3.1.2 Parameter estimates

As in Didier (2021), we compute the Bayesian posterior distribution of the parameters of the model specification $\mathcal{M}_1^e$ with regard to the dataset of Figure 1b (since $\mathcal{M}_1^e$ has the greatest Akaike weight for this dataset). These distributions are displayed in Figure 5a.

For all the parameters of $\mathcal{M}_1^e$, except the sampling probability ρ, we do observe that the corresponding posterior distribution are quite close to the parameter value used to simulated the dataset.

Maximum likelihood estimation of the $\mathcal{M}_1^e$ parameters gives $\lambda = 0.045215$, $\lambda' = 0.104704$, $\lambda'' = 0.014057$, $\mu = 0.013830$, $\psi = 0.019330$ and $\rho = 0.038154$, which is fairly consistent with both with the posterior distributions of Figure 5a and with the parameters used to simulated the dataset, except for the sampling/survival probability ρ, which is underestimated both in its Bayesian posterior density and in its maximum likelihood estimate.

Maximum likelihood estimation of the parameters of the model specification $\mathcal{M}_0$ and $\mathcal{M}_1$, used to simulated the datasets on which the Akaike weight distributions of Figures 4b,c were computed,

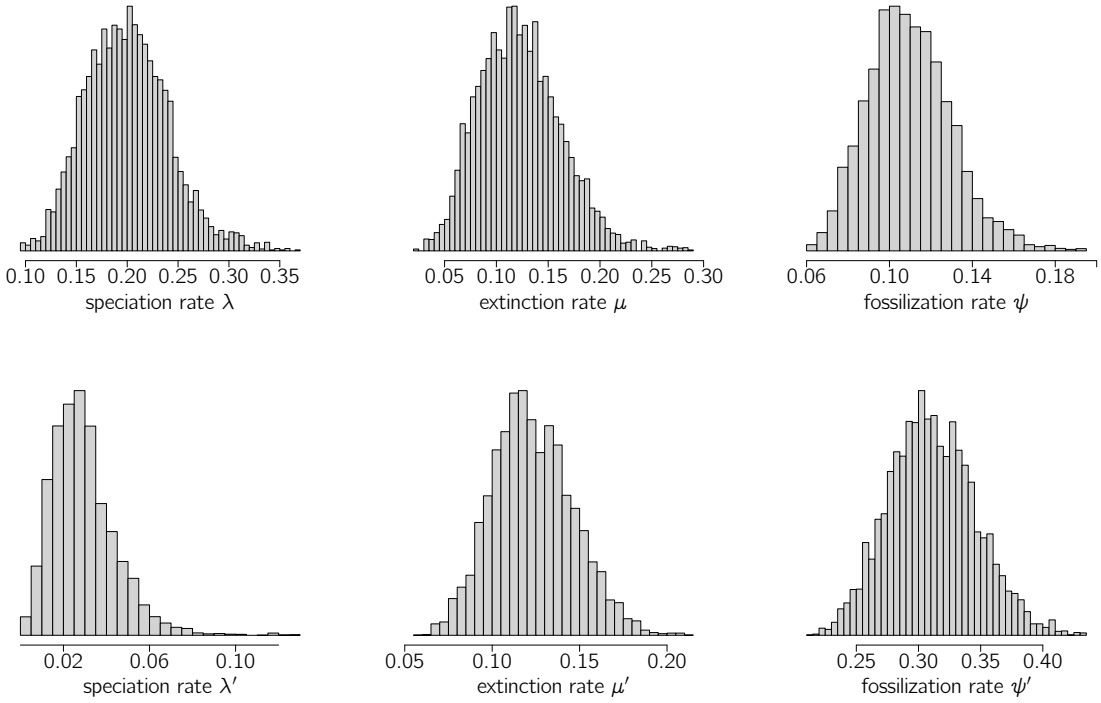

Figure 6: Posterior distributions of the parameters of the model specification $\mathcal{M}_2$ (see Fig. 3) obtained from the 100 trees of the empirical dataset (see Didier and Laurin (2021) for details about the Bayesian MCMC framework). Upper row: model parameters for the first time interval ($[320, 293.5]$ Ma). Lower row: model parameters for the second time interval ($[293.5, 0]$ Ma).

gives $\lambda = 0.020079$, $\mu = 0.015619$ and $\psi = 0.020400$ (for $\mathcal{M}_0$) and $\lambda = 0.027559$, $\lambda' = 0.062272$, $\lambda'' = 0.013098$, $\mu = 0.014278$ and $\psi = 0.018138$ (for $\mathcal{M}_1$). The maximum likelihood estimates of all the model specifications considered on the simulated dataset are provided in Appendix B.

3.2 Empirical data on Permo-Carboniferous eupelycosaurs

The best model specification ($\mathcal{M}_2$; see Table 2) has two time slices delimited by the Asselian/Sakmarian boundary and lacks a mass extinction event. Its maximum likelihood and AIC are only moderately better than the model specifications with 2 time slices without extinction with boundaries between slices at the Gzhelian/Asselian ($\mathcal{M}_1$) and Sakmarian/Artinskian ($\mathcal{M}_3$) boundaries, respectively. However, all other model specifications, including those with the same time slices but with an extinction ($\mathcal{M}_1^e$ and $\mathcal{M}_2^e$), have substantially worse support.

In order to assess the identifiability of the model specifications considered for the empirical dataset, we simulated datasets under the maximum likelihood (ML) skyline FBD models estimated under model specifications $\mathcal{M}_0$, $\mathcal{M}_1$, $\mathcal{M}_1^e$, $\mathcal{M}_2$, $\mathcal{M}_2^e$, $\mathcal{M}_3$ and $\mathcal{M}_3^e$ from the empirical dataset (100 simulations for each model). The distributions of Akaike weights of all the model specifications of Figure 3 computed from these datasets are displayed in Figure 7. We do not displays results obtained from datasets simulated by using ML estimates under $\mathcal{M}_4$, $\mathcal{M}_4^e$, $\mathcal{M}_5$ and $\mathcal{M}_5^e$ in order to not overload this figure and since their behavior is not very different from the other ones. By examining the columns of the empirical distributions of Figure 7, we observe that the model specification associated to the model used to simulated the dataset have Akaike weights greater than 10% in most replicates, and that the correct model is little represented in the first column, which provides the proportion of Akaike weights lower than 10%. Conversely, the model specifications not used to simulate the datasets have mostly low Akaike weights, except those which are close to the simulated model, notably those with the same interval bounds which

model specification	$\mathcal{M}_0$	$\mathcal{M}_1$	$\mathcal{M}_1^e$	$\mathcal{M}_2$	$\mathcal{M}_2^e$	$\mathcal{M}_3$	$\mathcal{M}_3^e$
number of parameters	3	6	7	6	7	6	7
log average likelihood	-731.64	-705.28	-706.27	-704.22	-708.26	-706.19	-707.74
AIC	1959.27	1912.56	1916.55	1910.43	1920.52	1914.37	1919.48
w(AIC)	0.00	0.22	0.03	0.65	0.00	0.09	0.01

model specification	$\mathcal{M}_4$	$\mathcal{M}_4^e$	$\mathcal{M}_5$	$\mathcal{M}_5^e$
number of parameters	6	7	6	7
log max average likelihood	-718.45	-721.71	-726.50	-725.41
AIC	1938.91	1947.41	1955.01	1954.82
w(AIC)	0.00	0.00	0.00	0.00

Table 2: Fit of the models displayed in Figure 3 on the empirical dataset (the AIC computation takes into account its number of fossils, i.e. 245).

differ only in the presence/absence of a mass extinction.

The maximum likelihood estimate of the parameters of the best model specification ($\mathcal{M}_2$) gives that the rates of the first interval ($[320, 293.5]$) are $\lambda = 0.205211$, $\mu = 0.146072$ and $\psi = 0.109980$ and those of the second one ($[293.5, 0]$) are $\lambda' = 0.021534$, $\mu' = 0.098555$ and $\psi' = 0.352608$ (the maximum likelihood estimated of all the model specifications considered on the empirical dataset are provided in Appendix B). The posterior distributions of these rates are displayed in Figure 6.

Both of these posterior distributions and the maximum likelihood estimates show that both time slices differ most strongly from each other by their speciation (cladogenesis) rates, which decrease nearly ten-fold between the first (older, at 0.2 originations per lineage per Ma) and second (more recent, at 0.02 originations per lineage per Ma) time slice, and the distribution of the parameter values over the 100 trees and random sampling of the ages of all fossils within the bounds of their plausible ages yields a similar, though slightly less pectacular, pattern (Fig. 6). This general pattern is similar in the other two models that have moderate support ($\mathcal{M}_1$, with an Akaike weight of 0.22 and $\mathcal{M}_3$, with an Akaike weight of 0.09). Thus, our results indicate that sometime in the Cisuralian, most likely in the Asselian or Sakmarian, cladogenetic rates decreased drastically in eupelycosaurs. However, the posterior distribution of parameter estimates do not suggest an increase in extinction rates (Fig. 6), and maximul likelihood estimates even suggest a moderate decrease (by about a third) in this rate. The drastic reduction in standing diversity of eupelycosaurs between the early Kungurian and early Roadian appears to result from lack of replacement of lineages that continued to become extinct at a similar rate as before (or even at a slightly slower rate).

4 Discussion

The simulation study based on the evolutionary model shown in Figure 1a demonstrates the effectiveness of our model selection in distinguishing between $\mathcal{M}_0$, $\mathcal{M}_1$ and $\mathcal{M}_1^e$ (shown in Figure 2). Despite their close resemblance to each other (with differences only in the speciation rate and the survival/sampling probability), the Akaike weights obtained from our approach accurately highlight the correct model specification. The only situation where the model choice may be ambiguous, and even possibly wrong, is when applied on datasets simulated from models satisfying the constraints of $\mathcal{M}_1$, where a strong support for model $\mathcal{M}_1^e$ may be observed in some replicates. This finding is moderately surprising, considering that (i) $\mathcal{M}_1^e$ has only one more parameter than $\mathcal{M}_1$, and thus its AIC is not heavily penalized with regards to that of $\mathcal{M}_1$ and (ii) though mass extinction events are not generated in the simulated datasets, modeling an extinction with a small intensity at the end of the first interval may increase the likelihood of some of the simulated datasets. However, this problem is observed only in a small proportion of the simulated datasets (less than 10%).

Similarly, the parameters estimated from the simulated datasets are mostly consistent with those used to generate them, both considering their maximum likelihood estimates and their posterior distributions. However, there is an exception for the sampling probability at the end of the first interval, which is underestimated both in its Bayesian posterior distribution, showing a relatively wide range, and by the

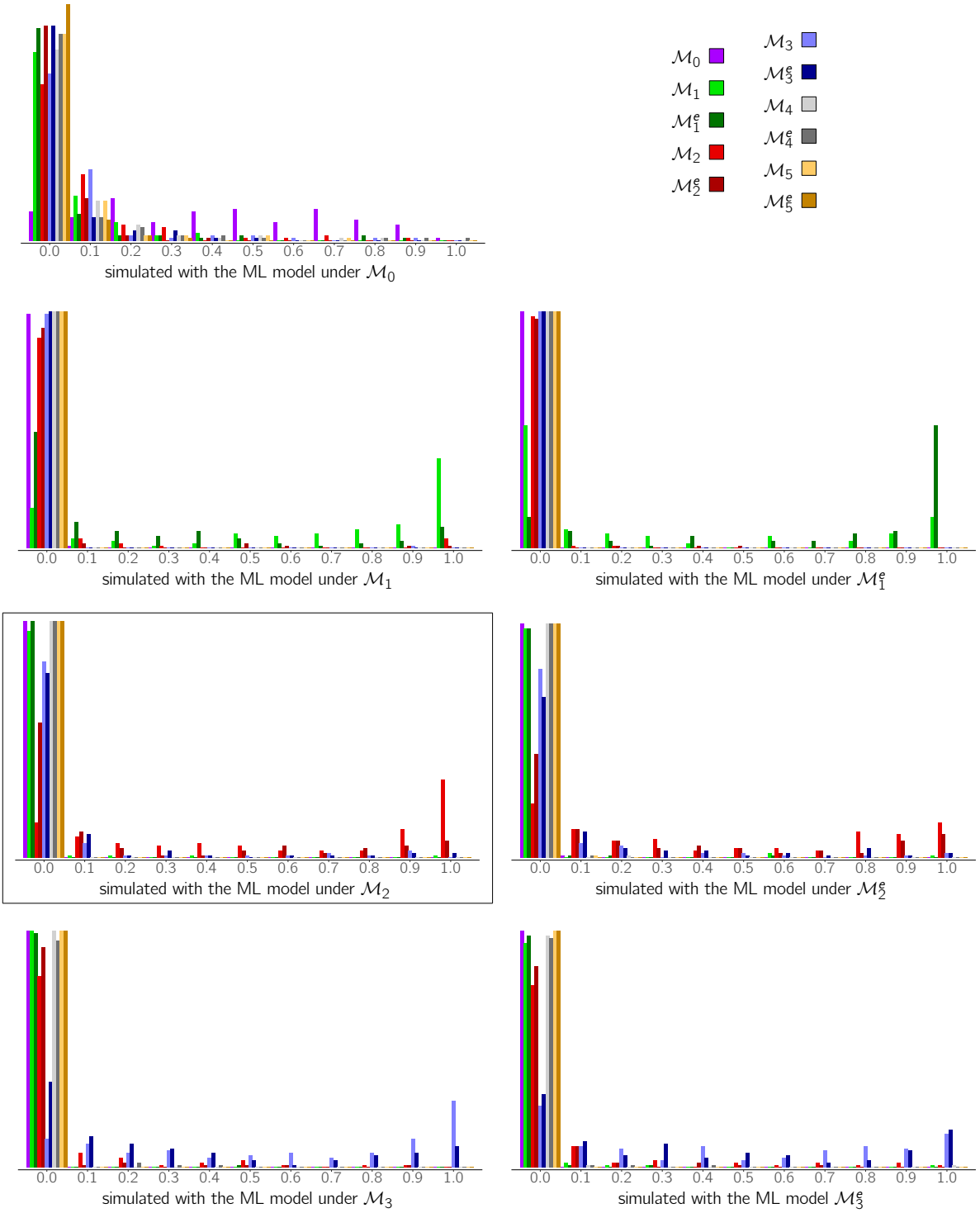

Figure 7: Distributions of the Akaike weights obtained from datasets simulated from the maximum likelihood (ML) models estimated under model specifications $\mathcal{M}_0$, $\mathcal{M}_1$, $\mathcal{M}_1^e$, $\mathcal{M}_2$, $\mathcal{M}_2^e$, $\mathcal{M}_3$ and $\mathcal{M}_3^e$ from the empirical dataset. The plot obtained from $\mathcal{M}_2$, the best supported model specification, is framed.

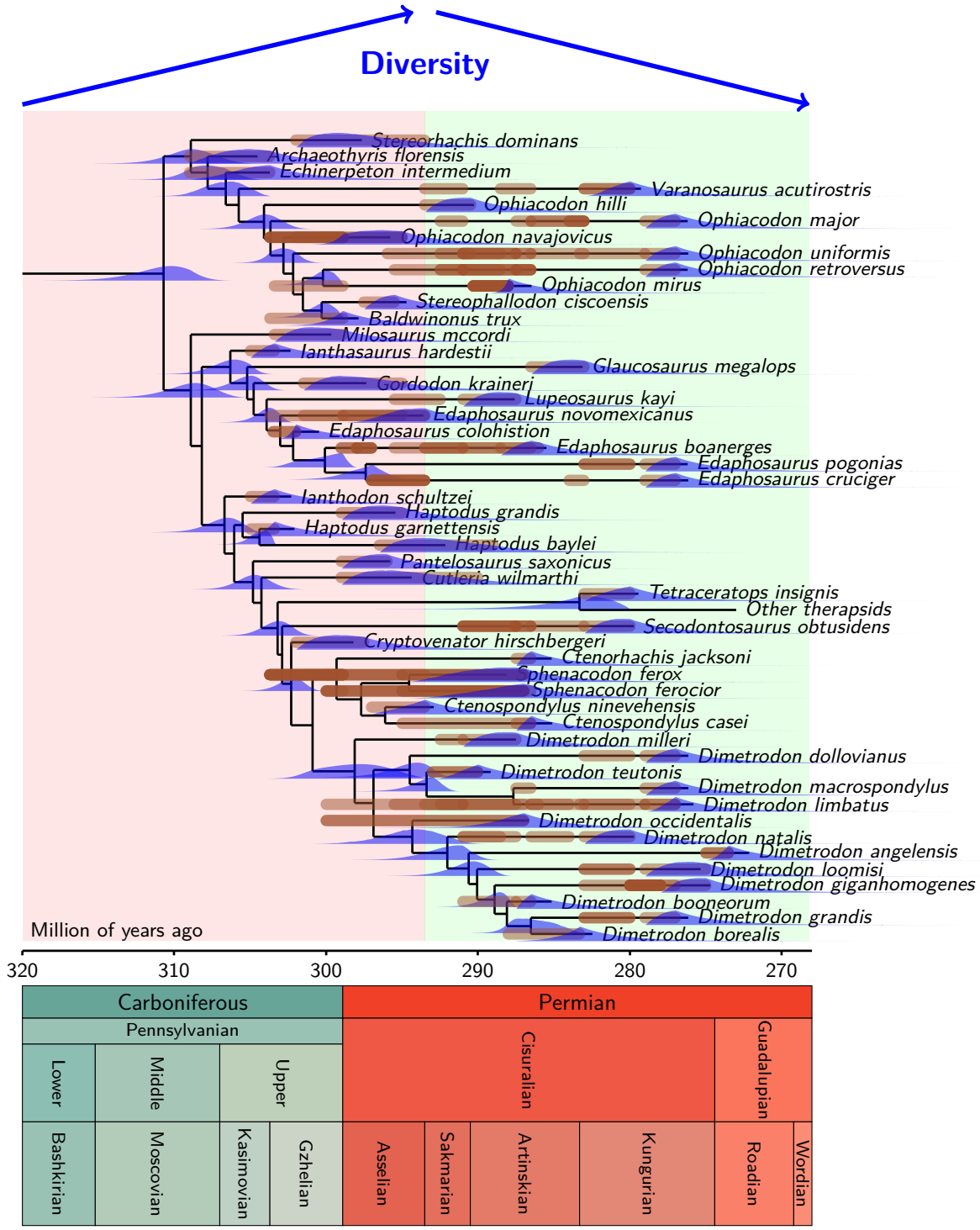

Figure 8: The model with the best fit, with a limit between the phases of evolutionary radiation and dwingling biodiversity at the Asselian/Sakmarian boundary. The distributions of divergence and extinction times were computed by using the sofware presented in Didier and Laurin (2020, 2021). The brown bars represent age uncertainty associated with each fossil occurrence; darker bars represent two or more overlapping stratigraphic intervals.

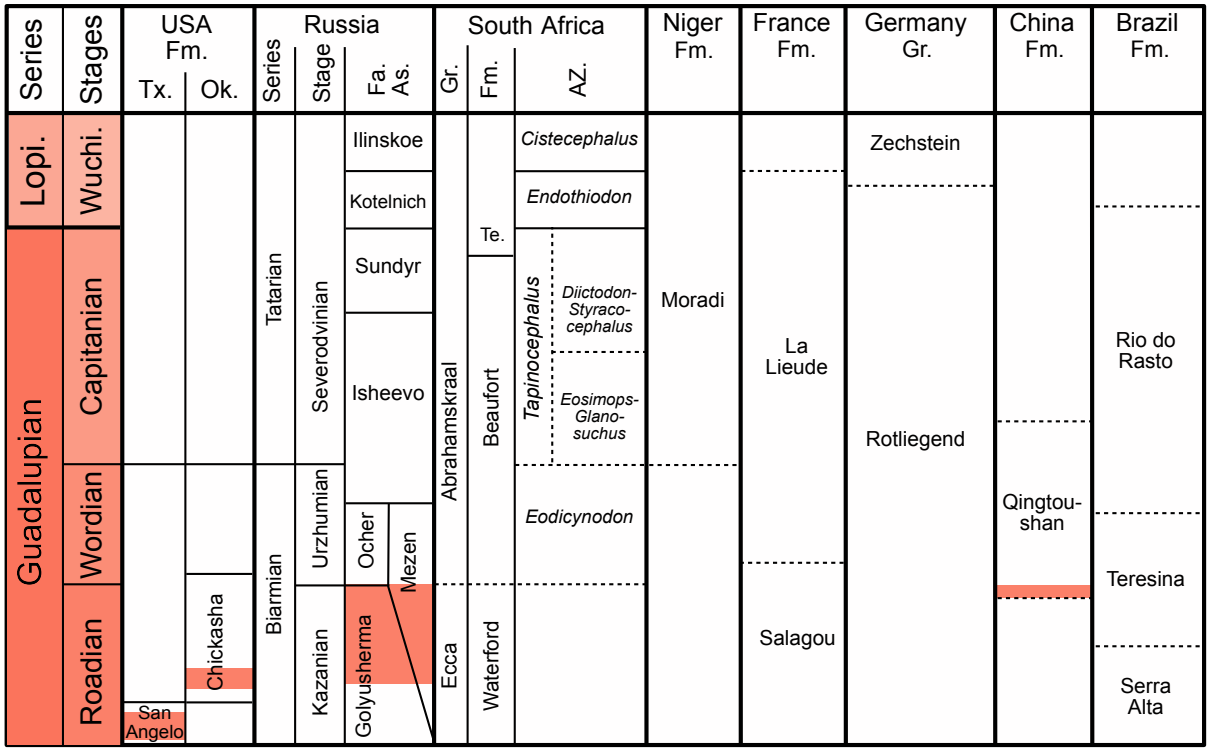

Figure 9: Stratigraphic correlations between Middle Permian strata from various parts of the world showing that several formations are known in the Roadian (in what has often been called Olson's gap), and some have yielded many amniote fossils. The best records occur in Russia, and more limited records are known from North America (San Angelo and Chickasha Formations). Levels that have yielded Roadian stegocephalian body fossils are shaded in orange. The Qingtoushan Formation, formerly misnamed Xidagou Formation (Liu 2018), has yielded amniote body fossils of a few taxa, including therapsids (Duhamel *et al.* 2021), but the age of this formation is poorly constrained (Roadian or Wordian).

maximum likelihood method. Let us first note that posterior distributions are not expected to peak exactly at the corresponding parameter value due to stochastic fluctuations, especially on relatively small datasets, such as the one depicted in Figure 1b. Estimating the sampling/survival probability was anticipated to be challenging because, unlike the diversification and fossilization rates that are estimated from events occurring during (large) intervals of time, the survival probability is estimated from a single punctual event involving mostly extinct (and unobserved) lineages. Considering the intrinsic difficulty of the question, the survival probabilities estimated on the dataset are not absurd in relation to the actual parameter. Estimating the survival probability may be problematic even with access to all the information about the diversification process (as seen in Fig. 1a) especially when the extinction event involves a small number of lineages. For example, in the complete realization that generates the simulated dataset, there are 12 lineages alive at the end of the first interval, of which only 2 survive the mass extinction (Figure 1a). The most accurate survival probability that can be estimated at this time for the simulated dataset is about 0.17, not the probability 0.2 used to generate it (an example of the stochastic fluctuations mentioned above). Additionally, apart from the considerations mentioned earlier, the observed difference between the posterior distribution and the maximum estimate of the survival probability could arise from the fact that we are using an approximation of the maximum likelihood.

As for the simulated dataset, the ability of our approach to distinguish between the various model specifications considered on the empirical dataset is quite well assessed by the simulation study displayed in Figure 7. In particular, the most supported model specification $\mathcal{M}_2$ is by far the most often chosen from datasets simulated under the $\mathcal{M}_2$-maximum likelihood model on the empirical dataset. We observe that the model specification $\mathcal{M}_2$ has significant Akaike weights for datasets simulated under the $\mathcal{M}_2^e$-maximum likelihood model. This was not surprising since (i) $\mathcal{M}_2^e$ is basically "$\mathcal{M}_2$ with extinction" and (ii) its maximum likelihood estimate on the empirical dataset has a survival probability greater than 80% (Appendix B); thus, it can be easily mistaken with a model without extinction. The model specification $\mathcal{M}_2$ shows significant Akaike weights for only a few datasets simulated under other models.

As in our previous study (Didier and Laurin 2021), we found no evidence of a late Sakmarian or Artinskian extinction event, and the extinction event that has often been called "Olson's extinction" (Sahney and Benton 2008; Brocklehurst *et al.* 2017; Brocklehurst 2018) appears to result from a protracted, gradual decline in biodiversity over 20 Ma, rather than a brief mass extinction event (see Table 2). The late Sakmarian or Artinskian do not appear to be times of elevated extinction in eupelycosaurs, but it is conceivable that the decrease in standing biodiversity produced by the dramatic (ten-fold) deceleration in cladogenesis rate that we detect near the Asselian/Sakmarian boundary may have been mistaken as an extinction signal by previous studies (Benton 1989; Brocklehurst *et al.* 2013). Our new analyses reveal that the slow erosion of eupelycosaur diversity in the Cisuralian results from this spectacular drop in origination rate, rather than a rise in extinction rate. Biodiversity drops through low origination rates (rather than elevated extinction rates) over fairly long periods have been documented for earlier parts of the Paleozoic by Bambach *et al.* (2004). Our failure to detect a late Kungurian mass extinction event (or such events at other Cisuralian stage boundaries) cannot be attributed to the Signor-Lipps effect (Signor *et al.* 1982) because our approach takes into consideration the incompleteness of the fossil record of all lineages (as shown in Figure 8).

Sahney and Benton (2008, p. 760) concluded that Olson's extinction was "an extended period of low diversity when worldwide two-thirds of terrestrial vertebrate life was lost"; they indicated that this low-diversity period was in the Roadian and Wordian, but their figure 1 shows that diversity was already low in the Kungurian. Sahney and Benton (2008, p. 761) reported high standing diversity in the Artinskian, and this is compatible with our findings, which suggest that diversification rates fell around the Asselian/Sakamrian transition, but extinction had not yet reduced standing diversity much by the Artinskian (Fig. 8). In any case, some discrepancies are to be expected between the results of our study and Sahney and Benton (2008), because the taxonomic samples and analytical methods used in both studies differ. More importantly, Sahney and Benton (2008) quantified taxonomic diversity through taxon counts at the family level, whereas we worked on nominal species, interpreted as evolutionary lineages. Despite this, in the Kungurian, both studies suggest a lower diversity than in the Artinskian. This is true even though the three main clades in our sample (Ophiacodontidae, Edaphosauridae, and Sphenacodontidae) persist at least into the late Kungurian (but with fewer included species); a simple family count would not have perceived a drop in biodiversity in the Kungurian in our dataset.

Brocklehurst *et al.* (2017) argued that the the temperate Inta Kungurian fauna (Golubev 2015) from the Russian Platform was more similar to the coeval equatorial fauna from North America than to the

Kazanian fauna. From this, they concluded that this implied a fairly sudden mass extinction event near the Kungurian/Roadian boundary. We disagree with this reasoning, for two reasons.

First, similarities in faunal compositions have at best indirect implications on extinction rates. Indeed, our results show that for the three eupelycosaur clades that we sampled, this event is actually a slow decline, which is qualitatively different from mass extinction events as they are usually understood, as catastrophic phenomena that occur over a short time (geologically speaking), typically about 1 Ma or less (Burgess *et al.* 2014). Some crises may imply a slow decline followed by a brief extinction climax (Ward *et al.* 2005), and our data cannot completely rule out such a pattern, given that a visual inspection suggests that extinction rates might possibly be higher a bit before the end of the Kungurian than earlier in the cisuralian (Fig. 8). This is why we tested the twelfth model, modified from the optimal model ($\mathcal{M}_2$; see Table 2) to allow a mass extinction event at -275 Ma, a little before the end of the Kungurian. The low support of this model, only about a third of the AIC weight of the optimal model ($\mathcal{M}_2$), refutes the hypothesis of a crisis near the end of the Kungurian. However, this question deserves further investigation with a larger taxonomic sample.

Second, the Inta fauna is poorly known; it is represented, according to Brocklehurst *et al.* (2017), by 14 specimens representing only eight nominal species, which is hardly a good basis for making statistical comparisons. This fauna is best known from its aquatic component, which is dominated by temnospondyls and an eogyrinid embolomere (Golubev 2000). The more terrestrial component is known only from a captorhinid and a second reptile, which may be a captorhinid or a bolosaurid (Brocklehurst *et al.* 2017). No skeletal remains of synapsids have been reported so far. Brocklehurst *et al.* (2017) stated that "the ichnotaxa *Dromopus* and *Dimetropus* are known from the area". While *Dimetropus* was probably made by Permo-Carboniferous (non-therapsid) synapsids (Sacchi *et al.* 2014), Brocklehurst *et al.* (2017) overstated the significance of these ichnofossils. While Inta is near the western foothills of the northern Urals, the locality that yielded the ichnofossils mentioned by Brocklehurst *et al.* (2017) is from the Russian Caucasus (Lucas *et al.* 1999), about 2500 km away! This is hardly the same area. Furthermore, while Inta is fairly well-constrained biostratigraphically to the Kungurian (Golubev 2015), Lucas *et al.* (1999) stated that the age of the *Dimetropus* and *Dromopus* localities in the Caucasus "is Early Permian, probably pre-late Sakmarian." Thus, these ichnofossil from the Caucasus are close to Inta neither geographically nor chronologically and do not contribute to our knowledge of the Inta faunal assemblage. To conclude, we think that the Inta fauna is too poorly known to draw firm conclusions about the pace of faunal turnover in synapsids (and probably, in reptiles as well) around the Kungurian/Roadian boundary in the northern Laurasian areas of the Russian platform. To a lesser extent, this sample size limitation extends to the Roadian equatorial amniote fossil record, which is limited to the rather scant records of the San Angelo and Chickasha Formations (Laurin and Hook 2022). Some of the most intriguing taxa of these two formations, which were interpreted as therapsids by Olson (Olson 1962, 1965), need to be re-examined. After Olson stopped working on this material, only a conference abstract has been published on this material (Sidor and Hopson 1995). That study concluded that all the San Angelo specimens interpreted as theraspids by Olson belong to more basal synapsid clades, but did not mention the Chickasha specimens, which are mentioned by (Olson 1962) as being from the Flower Pot, but these strata, in current stratigraphic nomenclature, are part of the Chickasha Formation (Laurin and Hook 2022, p. 6).

In the literature, "Olson's extinction" has been opposed to "Olson's gap" (Lucas 2001), as if they were mutually exclusive phenomena, and that the absence of one implied the presence of the other (Brocklehurst *et al.* 2017; Brocklehurst 2020). However, we argue that it is misleading to call the Roadian "Olson's gap" because it is simply a time during which the fossil record of amniotes is less dense and less geographically widespread than before or after (Fig. 9). A true gap would lack any record, but the Roadian amniote fossil record is represented in North America by the San Angelo and the Chickasha formations, most likely in the early Roadian (Laurin and Hook 2022), whereas a much greater (and probably slightly more recent) portion of the Roadian is represented in the Russian Kazanian, especially the Golyusherma and Mezen assemblages (Golubev 2015; Ivakhnenko 2015). This was recognized by other authors before (Benton 2013; Duhamel *et al.* 2021; Brocklehurst 2020), some of whom even called this period "Olson's bridge" (Lozovsky 2005). This bridge may be narrow right now, but it might improve through additional field work. For instance, no stegocephalian body fossils have been found so far in the Teresina Formation as far as we know, but it has yielded xenacanthiform shark remains (Richter 2005), which are associated with stegocephalians in many Permo-Carboniferous localities (Huttenlocker *et al.* 2018; Johnson 2018). Better dating of some formations might also reveal other Roadian amniote

records. Thus, the Qingtoushan Formation (formerly misnamed Xidagou), which has yielded therapsids (Duhamel *et al.* 2021), may, at least its lower part, date from the Roadian; so far, this remains uncertain (Liu 2018). Similarly, the Cala del Vino Formation (Sardinia, Italy), which has yielded the giant caseid *Alierasaurus ronchii*, is late Kungurian to Roadian in age (Romano *et al.* 2017).

Sahney *et al.* (2010) suggested that Carboniferous rainforest collapse in the earliest Kasimovian time (ca. 305 Ma) stopped diversification of anamniotic stegocephalians, but that amniotes fared better, perhaps because they were better adapted to dryer conditions. Our findings are compatible with this scenario, without providing strong support for it. A rainforest collapse giving a competitive edge to early amniotes (over other stegocephalians) starting in the early Kasimovian could explain the early phase of amniote diversification that seems to last until the early Cisuralian. However, this scenario does not explain why eupelycosaur diversification apparently slowed down dramatically 10 to 15 My later, in the Asselian or Sakmarian. Further comparisons are hampered by the poor fossil record of eupelycosaurs and other amniotes before the Kasimovian, but it would be interesting to perform similar analyses with a time-sliced (skyline) FBD model on other stegocephalian clades that have a more extensive Mississippian and early Pennsylvanian fossil record to test the scenario proposed by Sahney *et al.* (2010).

Our understanding of the faunal replacement around the Kungurian/Roadian boundary remains scanty. Only a small portion of this phenomenon, the extinction of Ophiacodontidae, Edaphosauridae, and Sphenacodontidae, now appears to be reasonably well-dated. However, the cause of these extinctions remains unclear, partly because there is no obvious external cause for this crisis (i.e., no major phase of intensive volcanism, no large meteorite impact). A simple climatic change (warming and aridification) over a long time might be involved (Giles *et al.* 2013; Foster *et al.* 2014), but the evidence for this, namely the evaporite-rich sediments in the Midland, Palo Duro and Anadarko basins, could also simply be explained by the filling of the Midland Basin (Laurin and Hook 2022, p. 20). In fact, even in the presence of putative causes, untangling the complex web of environmental effects on taxa is difficult to assess, even though progress is being made in this direction (Foster *et al.* 2022). The evolution of captorhinids around that time is also fairly well-understood, but captorhinids do not show particularly spectacular taxonomic diversification (or extinction) patterns at that time (Brocklehurst 2017). The component of this faunal turnover that remains poorly dated but might be highly relevant is the rise of therapsids, given their likely origin in the Late Carboniferous (Fig. 8), as suggested by their long-accepted sister-goup relationship with Sphenacodontidae (Sidor 2001; Amson and Laurin 2011). However, they are currently unknown in the Carboniferous, except for some fragmentary remains that have been suggested to be therapsids (Spindler 2014), and their presence remains fairly cryptic in the Cisuralian, with a single putative therapsid (Amson and Laurin 2011), neither of which is universally accepted as such (Duhamel *et al.* 2021). Their spectacular evolutionary radiation in the Roadian and Wordian thus occurs dozens of millions of years after their likely time of origin.

Our new (skyline) implementation of the FBD model with rich stratigraphic and phylogenetic data associated with fossils provides new insights that were unavailable in our previous FBD implementation (Didier and Laurin 2021). While we did not find evidence for a mass extinction event (notably at the Kungurian/Roadian boundary), this model identified important rate shifts between two broad time slices. In addition to the great decrease in cladogenesis rate, the model shows that fossilization potential actually improved (it more than tripled) in the second time slice. This probably reflects the greater exposures of Cisuralian continental fossiliferous strata than the equivalent Carboniferous strata, especially in the South-Western USA, where the fossil record of Permo-Carboniferous synapsids is the richest. This suggests that the slow attrition in the diversity of ophiacodontids, edaphosaurids, and sphenacodontids throughout much of the Cisuralian (after the Asselian) is unlikely to reflect a taphonomic artefact. These findings also illustrate the macro-taphonomic potential of our new skyline FBD model implementation.

The new insights provided by our analysis of the Permo-Carboniferous fossil record contrast sharply with the pessimistic assertions of Beaulieu and O'Meara (2023p. 60), who concluded that "Taken together, our simulations provide us with little hope for the utility of fossils in realworld applications, at least with regard to estimating diversification rate heterogeneity across a tree." This is a surprising claim, which raises doubts about the relevance of the FBD model itself. Beaulieu and O'Meara (2023) reached these conclusions after performing fours series of analyses of simulated data. The simulations were complex and designed to challenge the methods (perhaps too much) because they included four regimes of turnover and extinction fraction spread randomly in the tree, as well as two mass extinction events and a biased sampling of extant lineages of 75 percent. Beaulieu and O'Meara (2023) then tested various models (which do not match the models used in simulating the data) to see how well they would estimate the

FBD model parameters. In these simulations, Beaulieu and O'Meara (2023) distinguished between fossils that have no descendants whatsoever (either other fossils, including those belonging to the same lineage or nominal species, or extant taxa), which they call "childless fossils", and those, called "parent fossils", that left some descendants (including more representatives of the same lineage or nominal species). Their four series of analyses of simulated data are: one with only extant taxa, disregarding fossils; a second one with extant taxa and all fossils; a third with extant taxa and only half of the "parent fossils" (but all the "childless fossils"), and a fourth, with only extant taxa and "childless fossils". Beaulieu and O'Meara (2023) found that the inclusion of all fossils helped to better constrain estimates of the FBD parameters, but had no impact on bias. However, excluding some or all of the "parent fossils" resulted in biased estimates, and Beaulieu and O'Meara (2023) argued that established paleontological practice would most likely result in such a bias. We believe that Beaulieu and O'Meara (2023) misunderstood paleontolocical practice, and this explains the differences between our results and those from Beaulieu and O'Meara (2023).

First, Beaulieu and O'Meara (2023p. 50) was aimed at "understanding diversification given an existing, largely modern tree", which might be illustrated by placental mammals or passerine birds, but many paleontological studies have documented spectacular diversification and extinction patterns that could not possibly be detected based only on extant taxa. Our study provides such an example; we tackled the evolution of Permo-Carboniferous synapsids, and among the 50 terminal taxa represented in the tree, a single one (other therapsids) left descendants (which include all mammals). Obviously, no method, even a very sophisticated one, could study the evolution of biodiversity in a clade that includes descendants of a single lineage from the period of interest, but paleontologists have been tackling such problems for a long time (Raup and Sepkoski 1982; Benton 1985; Sahney and Benton 2008; Brocklehurst *et al.* 2013), and this research program is ongoing (Servais *et al.* 2023).

Second, Beaulieu and O'Meara (2023) confused two related but distinct practical problems raised by the use of fossils in the FBD process. They rightly pointed out that there has been a debate about our ability to identify ancestors in the fossil record, debates that hark back to the works of Hennig (1966, 1981) but that continue today (Foote 1996; Gavryushkina *et al.* 2014; Zeitoun 2015). However, these debates have focused on placing distinct taxa represented by fossils in an ancestral position or on terminal branches, rather than on the possible ancestor-descendant relationships of fossils attributed to the same nominal species but originating from different stratigraphic levels. This means that it is difficult to place fossils on internal branches, as ancestors of other nominal taxa, but this does not necessarily rule out fossils being ancestors of other fossils placed on the same branch and belonging (typically) to the same nominal species. On the contrary, the widespread use of a phenetic species concept in paleontology (see below) has led to the recognition of many putative ancestor-descendant relationships between populations of a presumed lineage sampled in various stratigraphic intervals (Wood *et al.* 2007). This paleontological tradition also underlies the ancient concept of stratigraphic range, which is at the core of biostratigraphy, and which remains relevant today (Marshall 1994). Ancestral fossils are also implied by the chronospecies concept, which was proposed before cladistics (Young 1960), but remains in use in plaeontology (Gheerbrant *et al.* 2021), and which is a particular case of the phenetic species concept (Laurin 2023, p. 173-174). The test of possible bias introduced by excluding fossils that represent ancestors in the fossil record performed by Beaulieu and O'Meara (2023) consisted in removing half or all of the "parent fossils", rather than fossils located on internal branches of the trees. This does not match the constraints imposed by the difficulty in recognizing ancestors of higher taxa among nominal species in the fossil record.

Third, Beaulieu and O'Meara (2023, p. 53) claimed that "Sampled ancestors outnumbering other kinds of fossils are not the pattern typically found with FBD studies", but this seems to be based on examination of "a few papers" (apparently four, judging by their figure 3, and paleontologists appear to have had minimal involvement in these studies). Our datasets, despite the moderately rich fossil record of Permo-Carboniferous synapsids, contains 126 "parent fossils" and only 49 "childless fossils". Thus, the low number of "parent fossils" in a few FBD empirical studies appears to result from lack of collecting more data and is no grounds to dismiss the relevance of fossils in such studies. On the contrary, this shows that more care has to be paid in collecting paleontological data for such analyses. However, inappropriate fossil samples may have affected fewer studies that used the FBD than implied by (Beaulieu and O'Meara 2023), as shown by a brief examination of the three studies cited by Beaulieu and O'Meara (2023) as incorporating very few "parent fossils". The first study, by Zhang *et al.* (2016), includes both simulations and study of an empirical dataset (on Hymenoptera). It includes two simulations, each

analyzed in four different ways, which result in eight cases; in seven of these, the number of estimated ancestral fossils was close to that of the genuinely ancestral fossils (Zhang *et al.* 2016, figs. 4, 5). In the empirical study, seven out of eight analyses found very few ancestral fossils (no more than 2 percent), but another found that a third of the fossils were ancestors (Zhang *et al.* 2016, table 5). However, in this empirical dataset, the true proportion of ancestors is unknown, but Zhang *et al.* (2016, p. 244) expect it to be low for their dataset. Similarly, Didier *et al.* (2017, pp. 977-978) found that only two out of 50 taxa with stratigraphic data included in their analysis could plausibly be interpreted as ancestors (but perhaps neither is), and considering them as such resulted in marginal changes (about 1 to 5 percent of the estimated value) in the estimated FBD model parameters. This is coherent with the theoretical expectations formulated by (Zhang *et al.* 2016, p. 243) that "if the fossilization rate is high and the extinction rate is low, we expect most fossils to sit on branches leading to extant taxa. Conversely, if the fossilization rate is low and the extinction rate high, most fossils are instead likely to represent extinct side branches" because the dataset of Didier *et al.* (2017) indeed yielded a fairly high extinction rate and a low fossilization rate. Thus, a low observed proportion of "parent fossils" in an empirical study is not necessarily problematic. The second study, by Slater *et al.* (2017), incorporated fossils into their analysis to better assess body size evolution in mysticetes (baleen whales). They incorporated a FBD prior in the context of a total evidence phylogenetic analysis, but it is difficult to determine if the topology and branch lengths were determined mostly by their molecular and morphological data of by the FBD model. The FBD parameters on which Beaulieu and O'Meara (2023) focused are not really relevant to Slater *et al.* (2017), which neither presented nor discussed them. The third study (Pyron 2017) is likewise not directly concerned by FBD parameters, even though the FBD is used in the context of a total-evidence phylogenetic analysis of squamates. Also like Slater *et al.* (2017), Pyron (2017) used a sampled-ancestor process, so the status of fossils as "parent fossils" or "childless fossils" is a result of the analysis, rather than an assumption, and certainly does not reflect the inability of cladistics to positively identify ancestors in the fossil record. Pyron (2017) had to enter age ranges for extinct terminal taxa because many of these were scored from several specimens, sometimes of different geological ages. Brief, studies based on the FBD could certainly use more stratigraphic information in the future, but our brief literature survey does not reveal a strong bias against "parent fossils" that would cast doubt on the main conclusions of previous FBD studies.

Fourth, Beaulieu and O'Meara (2023, p. 56) dismissed the improvements brought by fossils in the extinction rate estimates obtained by Mitchell *et al.* (2019) because they claimed that these were visible mostly in simulations with a fossilization rate of 1 fossil per lineage and per million years, while they claimed that the maximal rate that they examined (0.1 fossil per lineage and per million years) was a more reasonable value. This seems overly pessimistic because our own dataset yields values of 0.11 for the relatively poorly known Late Carboniferous and Asselian, and 0.35 for the Sakmarian-Kungurian interval, and Paleozoic vertebrates are by no means exceptionally well-represented in the fossil record. Our study of Neogene Old World camelids yielded values ranging from 1.29 to 1.43, depending on the tree used, despite the fact that camelids are not the best-represented mammals in the fossil record.

Fifth, Beaulieu and O'Meara (2023) implemented a model that misrepresents how paleontological species have been recognized and delimited. Their model considers that no cladogenesis can happen over the stratigraphic range of a nominal species. This is true if we conceptualize species as a branch (internode) of the Tree of Life. However, paleontologists have named new species based on diagnostic differences, most often (and nearly always, in studies published before the 1970s) without doing a phylogenetic analysis. Thus, given that the overwhelming majority of paleontologists effectively use a phenetic species concept, even though many of them probably accept, in principle, the biological species concept (Laurin 2023), cladogeneses can happen within nominal species(Marshall 1994, fig. 2). In fact, this long-recognized phenomenon (Bloch *et al.* 2001, fig. 12) has been called "budding" speciation, as opposed to "bifurcating (or symmetric)" speciation and "anagenetic speciation", all three of which have been implemented in some versions of the FBD (Stadler *et al.* 2018).

To sum up, while Beaulieu and O'Meara (2023) showed the genuine dangers in integrating only a limited, biased portion of the fossil record (namely, only one fossil occurrence per terminal taxon) in FBD analyses, their study neither demonstrates basic flaws in the FBD, nor the lack of usefulness of the fossil record to study temporal variations in the rates of cladogenesis and extinction. Indeed, our results are encouraging, and we hope to see further developments of the FBD and more empirical studies in this promising field.

References

Akaike, H. (1998). Information theory and an extension of the maximum likelihood principle. In K. T. E. Parzen and G. Kitagawa, editors, *Selected papers of Hirotugu Akaike*, pages 199–213. Springer.

Alfaro, M. E., Santini, F., Brock, C., Alamillo, H., Dornburg, A., Rabosky, D. L., Carnevale, G., and Harmon, L. J. (2009). Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. *Proceedings of the National Academy of Sciences*, **106**(32), 13410–13414.

Allin, E. F. and Hopson, J. A. (1992). Evolution of the auditory system in Synapsida ("mammal-like reptiles" and primitive mammals) as seen in the fossil record. In *The evolutionary biology of hearing*, pages 587–614. Springer.

Alroy, J. (2015). A more precise speciation and extinction rate estimator. *Paleobiology*, **41**(4), 633–639.

Amson, E. and Laurin, M. (2011). On the affinities of *Tetraceratops insignis*, an Early Permian synapsid. *Acta Palaeontologica Polonica*, **56**(2), 301–312.

Arens, N. C. and West, I. D. (2008). Press-pulse: a general theory of mass extinction? *Paleobiology*, **34**(4), 456–471.

Bambach, R. K., Knoll, A. H., and Wang, S. C. (2004). Origination, extinction, and mass depletions of marine diversity. *Paleobiology*, **30**(4), 522–542.

Barido-Sottani, J., Vaughan, T. G., and Stadler, T. (2020). A multitype birth-death model for Bayesian inference of lineage-specific birth and death rates. *Systematic Biology*, **69**(5), 973–986.

Beaulieu, J. M. and O'Meara, B. C. (2023). Fossils do not substantially improve, and may even harm, estimates of diversification rate heterogeneity. *Systematic Biology*, **72**(1), 50–61.

Benton, M. J. (1985). Mass extinction among non-marine tetrapods. *Nature*, **316**(6031), 811–814.

Benton, M. J. (1989). Mass extinctions among tetrapods and the quality of the fossil record. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, **325**(1228), 369–386.

Benton, M. J. (2003). *When life nearly died: the greatest mass extinction of all time*. Thames & Hudson London.

Benton, M. J. (2013). No gap in the Middle Permian record of terrestrial vertebrates: Reply. *Geology*, **41**(9), e294–e294.

Bloch, J. I., Fisher, D. C., Rose, K. D., and Gingerich, P. D. (2001). Stratocladistic analysis of Paleocene Carpolestidae (Mammalia, Plesiadapiformes) with description of a new late Tiffanian genus. *Journal of Vertebrate Paleontology*, **21**(1), 119–131.

Bouckaert, R., Vaughan, T. G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., Heled, J., Jones, G., Kühnert, D., De Maio, N., Matschiner, M., Mendes, F. K., Müller, N. F., Ogilvie, H. A., du Plessis, L., Popinga, A., Rambaut, A., Rasmussen, D., Siveroni, I., Suchard, M. A., Wu, C.-H., Xie, D., Zhang, C., Stadler, T., and Drummond, A. J. (2019). BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLOS Computational Biology*, **15**(4), 1–28.

Bramble, D. M. (1978). Origin of the mammalian feeding complex: models and mechanisms. *Paleobiology*, **4**(3), 271–301.

Brocklehurst, N. (2017). Rates of morphological evolution in Captorhinidae: an adaptive radiation of Permian herbivores. *PeerJ*, **5**, e3200.

Brocklehurst, N. (2018). An examination of the impact of Olson's extinction on tetrapods from Texas. *PeerJ*, **6**, e4767.

Brocklehurst, N. (2020). Olson's Gap or Olson's Extinction? a Bayesian tip-dating approach to resolving stratigraphic uncertainty. *Proceedings of the Royal Society B*, **287**(1928), 20200154.

Brocklehurst, N., Kammerer, C. F., and Fröbisch, J. (2013). The early evolution of synapsids, and the influence of sampling on their fossil record. *Paleobiology*, **39**(3), 470–490.

Brocklehurst, N., Day, M. O., Rubidge, B. S., and Fröbisch, J. (2017). Olson's Extinction and the latitudinal biodiversity gradient of tetrapods in the Permian. *Proceedings of the Royal Society B: Biological Sciences*, **284**(1852), 20170231.

Burgess, S. D. and Bowring, S. A. (2015). High-precision geochronology confirms voluminous magmatism before, during, and after Earth's most severe extinction. *Science advances*, **1**(7), e1500470.

Burgess, S. D., Bowring, S., and Shen, S.-z. (2014). High-precision timeline for Earth's most severe extinction. *Proceedings of the National Academy of Sciences*, **111**(9), 3316–3321.

Burnham, K. P. and Anderson, D. R. (1998). *Model Selection and Inference – A Practical Information-Theoretic Approach*. Springer, New-York.

Culshaw, V., Stadler, T., and Sanmartìn, I. (2019). Exploring the power of Bayesian birth-death skyline models to detect mass extinction events from phylogenies with only extant taxa. *Evolution*, **73**(6), 1133–1150.

Day, M. O., Ramezani, J., Bowring, S. A., Sadler, P. M., Erwin, D. H., Abdala, F., and Rubidge, B. S. (2015). When and how did the terrestrial mid-Permian mass extinction occur? Evidence from the tetrapod record of the Karoo Basin, South Africa. *Proceedings of the Royal Society B: Biological Sciences*, **282**(1811), 20150834.

Didier, G. (2021). Probabilities of tree topologies with temporal constraints and diversification shifts. *Peer Community Journal*, **1**, e65.

Didier, G. and Laurin, M. (2020). Exact distribution of divergence times from fossil ages and tree topologies. *Systematic Biology*, **69**(6), 1068–1087.

Didier, G. and Laurin, M. (2021). Distributions of extinction times from fossil ages and tree topologies: the example of mid-Permian synapsid extinctions. *PeerJ*, **9**, e12577.

Didier, G., Royer-Carenzi, M., and Laurin, M. (2012). The reconstructed evolutionary process with the fossil record. *Journal of Theoretical Biology*, **315**(0), 26–37.

Didier, G., Fau, M., and Laurin, M. (2017). Likelihood of tree topologies with fossils and diversification rate estimation. *Systematic Biology*, **66**(6), 964–987.

Duhamel, A., Benoit, J., Rubidge, B., and Liu, J. (2021). A re-assessment of the oldest therapsid *Raranimus* confirms its status as a basal member of the clade and fills Olson's gap. *The Science of Nature*, **108**(4), 1–12.

Foote, M. (1996). On the probability of ancestors in the fossil record. *Paleobiology*, **22**(2), 141–151.

Foote, M. (2003). Origination and extinction through the Phanerozoic: a new approach. *The Journal of Geology*, **111**(2), 125–148.

Foster, T. M., Soreghan, G. S., Soreghan, M. J., Benison, K. C., and Elmore, R. D. (2014). Climatic and paleogeographic significance of eolian sediment in the Middle Permian Dog Creek Shale (Midcontinent US). *Palaeogeography, Palaeoclimatology, Palaeoecology*, **402**, 12–29.

Foster, W. J., Ayzel, G., Münchmeyer, J., Rettelbach, T., Kitzmann, N. H., Isson, T. T., Mutti, M., and Aberhan, M. (2022). Machine learning identifies ecological selectivity patterns across the end-Permian mass extinction. *Paleobiology*, **48**(3), 357–371.

Gavryushkina, A., Welch, D., Stadler, T., and Drummond, A. J. (2014). Bayesian inference of sampled ancestor trees for epidemiology and fossil calibration. *PLoS computational biology*, **10**(12), e1003919.

Gheerbrant, E., Khaldoune, F., Schmitt, A., and Tabuce, R. (2021). Earliest embrithopod mammals (Afrotheria, Tethytheria) from the Early Eocene of Morocco: anatomy, systematics and phylogenetic significance. *Journal of Mammalian Evolution*, **28**, 245–283.

Giles, J. M., Soreghan, M. J., Benison, K. C., Soreghan, G. S., and Hasiotis, S. T. (2013). Lakes, loess, and paleosols in the Permian Wellington Formation of Oklahoma, USA: Implications for paleoclimate and paleogeography of the Midcontinent. *Journal of Sedimentary Research*, **83**(10), 825–846.

Golubev, V. (2000). The faunal assemblages of Permian terrestrial vertebrates from Eastern Europe. *Paleontological Journal*, **34**(SUPP/2), S211–S224.

Golubev, V. (2015). Dinocephalian stage in the history of the Permian tetrapod fauna of Eastern Europe. *Paleontological Journal*, **49**(12), 1346–1352.

Gradstein, F. M. and Ogg, J. G. (2020). The chronostratigraphic scale. In *Geologic time scale 2020*, pages 21–32. Elsevier.

Hamilton, J. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, **57**(2), 357–84.

Harding, E. F. (1971). The probabilities of rooted tree-shapes generated by random bifurcation. *Advances in Applied Probability*, **3**(1), 44–77.

Heath, T. A., Huelsenbeck, J. P., and Stadler, T. (2014). The fossilized birth-death process for coherent calibration of divergence-time estimates. *Proceedings of the National Academy of Sciences*, **111**(29), E2957–E2966.

Hennig, W. (1966). *Phylogenetic systematics*. University of Illinois Press, Chicago.

Hennig, W. (1981). *Insect phylogeny*. John Wiley & Sons, New York.

Hentz, T. F. (1988). *Lithostratigraphy and paleoenvironments of Upper Paleozoic continental red beds, North-central Texas–Bowie (new) and Wichita (revised) Groups*, volume 170. Bureau of Economic Geology, University of Texas at Austin.

Höhna, S., Landis, M. J., Heath, T. A., Boussau, B., Lartillot, N., Moore, B. R., Huelsenbeck, J. P., and Ronquist, F. (2016). RevBayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification language. *Systematic Biology*, **65**(4), 726–736.

Hook, R. W. and Hotton III, N. (1991). A new sphenacodontid pelycosaur (Synapsida) from the Wichita Group, Lower Permian of north-central Texas. *Journal of Vertebrate Paleontology*, **11**(1), 37–44.

Huttenlocker, A. K., Henrici, A., Nelson, W. J., Elrick, S., Berman, D. S., Schlotterbeck, T., and Sumida, S. S. (2018). A multitaxic bonebed near the Carboniferous–Permian boundary (Halgaito Formation, Cutler Group) in Valley of the Gods, Utah, USA: vertebrate paleontology and taphonomy. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **499**, 72–92.

Ivakhnenko, M. (2015). Patterns of changes in theromorph taxa of permian terrestrial communities of Eastern Europe. *Paleontological Journal*, **49**, 70–78.

Johnson, G. D. (2018). *Orthacanthus platypternus* (Cope, 1883)(Chondrichthyes: Xenacanthiformes) teeth andother isolated vertebrate remains from a single horizon in the early Permian (Artinskian) Craddock Bonebed, lower Clear Fork Group, Baylor County, Texas, USA. *Acta Geologica Polonica*, pages 421–436.

Kammerer, C. F., Angielczyk, K. D., and Fröbisch, J. (2014). *Early evolutionary history of the Synapsida*. Springer.

Kammerer, C. F., Viglietti, P. A., Butler, E., and Botha, J. (2023). Rapid turnover of top predators in african terrestrial faunas around the Permian-Triassic mass extinction. *Current Biology*, **33**(11), 2283–2290.

Laurin, M. (2010). *How vertebrates left the water*. Univ of California Press.

Laurin, M. (2023). *The Advent of PhyloCode: The Continuing Evolution of Biological Nomenclature*. CRC Press.

Laurin, M. and Hook, R. W. (2022). The age of North America's youngest Paleozoic continental vertebrates: a review of data from the Middle Permian Pease River (Texas) and El Reno (Oklahoma) Groups. *Bulletin de la Société Géologique de France*, **193**(1).

Liu, J. (2018). New progress on the correlation of Chinese terrestrial Permo-Triassic strata. *Vertebrata Palasiatica*, **56**(4), 327.

Louca, S. and Pennell, M. W. (2020). Extant timetrees are consistent with a myriad of diversification histories. *Nature*, **580**(7804), 502–505.

Louca, S., McLaughlin, A., MacPherson, A., Joy, J. B., and Pennell, M. W. (2021). Fundamental identifiability limits in molecular epidemiology. *Molecular Biology and Evolution*, **38**(9), 4010–4024.

Lozovsky, V. R. (2005). Olson's gap or Olson's bridge, that is the question. *New Mexico Museum of Natural History and Science Bulletin*, **30**, 179–184.

Lucas, S. G. (2001). A global hiatus in the Middle Permian tetrapod fossil record. *Permophiles*, **38**, 24–27.

Lucas, S. G. (2004). A global hiatus in the Middle Permian tetrapod fossil record. *Stratigraphy*, **1**, 47–64.

Lucas, S. G. (2017). Permian tetrapod extinction events. *Earth-Science Reviews*, **170**, 31–60.

Lucas, S. G., Lozovsky, V. R., and Shishkin, M. A. (1999). Tetrapod footprints from Early Permian redbeds of the northern Caucasus, Russia. *Ichnos*, **6**(4), 277–281.

Marshall, C. R. (1994). Confidence intervals on stratigraphic ranges: Partial relaxation of the assumption of randomly distributed fossil horizons. *Paleobiology*, **20**(4), 459–469.

May, M. R., Höhna, S., and 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 in Ecology and Evolution*, **7**(8), 947–959.

Menning, M., Glodny, J., Boy, J., Gast, R., Kowalczyk, G., Martens, T., Rößler, R., Schindler, T., von Seckendorff, V., and Voigt, S. (2022). The Rotliegend in the Stratigraphic Table of Germany 2016 (STG 2016). *Zeitschrift der Deutschen Gesellschaft für Geowissenschaften*, **173**, 3–139.

Mitchell, J. S., Etienne, R. S., and Rabosky, D. L. (2019). Inferring diversification rate variation from phylogenies with fossils. *Systematic Biology*, **68**(1), 1–18.

Morlon, H. (2014). Phylogenetic approaches for studying diversification. *Ecology Letters*, **17**(4), 508–525.

Morlon, H., Lewitus, E., Condamine, F. L., Manceau, M., Clavel, J., and Drury, J. (2016). RPANDA: an R package for macroevolutionary analyses on phylogenetic trees. *Methods in Ecology and Evolution*, **7**(5), 589–597.

Morlon, H., Robin, S., and Hartig, F. (2022). Studying speciation and extinction dynamics from phylogenies: addressing identifiability issues. *Trends in Ecology & Evolution*, **37**(6), 497–506.

Nee, S., Holmes, E., May, R., and Harvey, P. (1994). Extinction rates can be estimated from molecular phylogenies. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, **344**(1307), 77–82.

Olivier, C., Houssaye, A., Jalil, N.-E., and Cubo, J. (2017). First palaeohistological inference of resting metabolic rate in an extinct synapsid, *Moghreberia nmachouensis* (Therapsida: Anomodontia). *Biological Journal of the Linnean Society*, **121**(2), 409–419.

Olson, E. C. (1962). Late Permian terrestrial vertebrates, USA and USSR. *Transactions of the American Philosophical Society*, **52**(2), 1–224.

Olson, E. C. (1965). New Permian vertebrates from the Chickasha Formation in Oklahoma. *Oklahoma Geological Survey, Circular*, **70**, 1–70.

Olson, E. C. (1968). The family Caseidae. *Fieldiana Geology*, **17**, 225–349.

Paradis, E. (2004). Can extinction rates be estimated without fossils? *Journal of theoretical biology*, **229**(1), 19–30.

Pyron, R. A. (2017). Novel approaches for phylogenetic inference from morphological data and total-evidence dating in squamate reptiles (lizards, snakes, and amphisbaenians). *Systematic Biology*, **66**(1), 38–56.

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

Rabosky, D. L. (2006). Likelihood methods for detecting temporal shifts in diversification rates. *Evolution*, **60**(6), 1152–1164.

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., Mitchell, J. S., and Chang, J. (2017). Is BAMM Flawed? Theoretical and practical concerns in the analysis of multi-rate diversification models. *Systematic Biology*, **66**(4), 477–498.

Raup, D. M. and Sepkoski, J. J. (1982). Mass extinctions in the marine fossil record. *Science*, **215**(4539), 1501–1503.

Reisz, R. R. and Laurin, M. (2001). The reptile *Macroleter*: First vertebrate evidence for correlation of Upper Permian continental strata of North America and Russia. *Geological Society of America Bulletin*, **113**(9), 1229–1233.

Reisz, R. R. and Laurin, M. (2002). Discussion and reply: The reptile *Macroleter*: First vertebrate evidence for correlation of Upper Permian continental strata of North America and Russia: Discussion. *Geological Society of America Bulletin*, **114**(9), 1176–1177.

Richter, M. (2005). A new xenacanthid shark (Chondrichthyes) from the Teresina Formation, Permian of the Paraná basin, southern Brazil. *Revista Brasileira de Paleontologia*, **8**(2), 149–158.

Romano, M., Ronchi, A., Maganuco, S., and Nicosia, U. (2017). New material of *Alierasaurus ronchii* (Synapsida, Caseidae) from the Permian of Sardinia (Italy), and its phylogenetic affinities. *Palaeontologia Electronica*, **20**, 1–27.

Rowe, T. (1988). Definition, diagnosis, and origin of Mammalia. *Journal of vertebrate Paleontology*, **8**(3), 241–264.

Sacchi, E., Cifelli, R., Citton, P., Nicosia, U., and Romano, M. (2014). *Dimetropus osageorum* n. isp. from the Early Permian of Oklahoma (USA): a trace and its trackmaker. *Ichnos*, **21**(3), 175–192.

Sahney, S. and Benton, M. J. (2008). Recovery from the most profound mass extinction of all time. *Proceedings of the Royal Society B: Biological Sciences*, **275**(1636), 759–765.

Sahney, S., Benton, M. J., and Falcon-Lang, H. J. (2010). Rainforest collapse triggered Carboniferous tetrapod diversification in Euramerica. *Geology*, **38**(12), 1079–1082.

Schneider, J. W., Lucas, S. G., Scholze, F., Voigt, S., Marchetti, L., Klein, H., Opluštil, S., Werneburg, R., Golubev, V. K., Barrick, J. E., *et al.* (2020). Late Paleozoic–early Mesozoic continental biostratigraphy–links to the standard global chronostratigraphic scale. *Palaeoworld*, **29**(2), 186–238.

Servais, T., Cascales-Miñana, B., Harper, D. A., Lefebvre, B., Munnecke, A., Wang, W., and Zhang, Y. (2023). No (Cambrian) explosion and no (Ordovician) event: A single long-term radiation in the early Palaeozoic. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **623**, 111592.

Sidor, C. A. (2001). Simplification as a trend in synapsid cranial evolution. *Evolution*, **55**(7), 1419–1442.

Sidor, C. A. and Hopson, J. A. (1995). The taxonomic status of the Upper Permian eotheriodont therapsids of the San Angelo Formation (Guadalupian), Texas. *Journal of Vertebrate Paleontology*, **15**(suppl. to 3), 53A.

Signor, P. W., Lipps, J. H., Silver, L., Schultz, P., *et al.* (1982). Sampling bias, gradual extinction patterns, and catastrophes in the fossil record. *Geological implications of impacts of large asteroids and comets on the Earth*, **190**, 291–296.

Silvestro, D., Schnitzler, J., Liow, L. H., Antonelli, A., and Salamin, N. (2014a). Bayesian estimation of speciation and extinction from incomplete fossil occurrence data. *Systematic Biology*, **63**(3), 349–367.

Silvestro, D., Salamin, N., and Schnitzler, J. (2014b). PyRate: a new program to estimate speciation and extinction rates from incomplete fossil data. *Methods in Ecology and Evolution*, **5**(10), 1126–1131.

Silvestro, D., Salamin, N., Antonelli, A., and Meyer, X. (2019). Improved estimation of macroevolutionary rates from fossil data using a Bayesian framework. *Paleobiology*, **45**(4), 546–570.

Slater, G. J., Goldbogen, J. A., and Pyenson, N. D. (2017). Independent evolution of baleen whale gigantism linked to Plio-Pleistocene ocean dynamics. *Proceedings of the Royal Society B: Biological Sciences*, **284**(1855), 20170546.

Slowinski, J. B. and Guyer, C. (1993). Testing whether certain traits have caused amplified diversification: An improved method based on a model of random speciation and extinction. *The American Naturalist*, **142**(6), 1019–1024.

Smith, R. M. and Botha-Brink, J. (2014). Anatomy of a mass extinction: sedimentological and taphonomic evidence for drought-induced die-offs at the Permo-Triassic boundary in the main Karoo Basin, South Africa. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **396**, 99–118.

Spindler, F. (2014). Reviewing the question of the oldest therapsid. *Freiberger Forschungshefte C*, **548**(22), 1–7.

Stadler, T. (2010). Sampling-through-time in birth-death trees. *Journal of Theoretical Biology*, **267**(3), 396–404.

Stadler, T. (2011). Mammalian phylogeny reveals recent diversification rate shifts. *Proceedings of the National Academy of Sciences*, **108**(15), 6187–6192.

Stadler, T., Kühnert, D., Bonhoeffer, S., and Drummond, A. J. (2013). Birth-death skyline plot reveals temporal changes of epidemic spread in HIV and hepatitis C virus (HCV). *Proceedings of the National Academy of Sciences*, **110**(1), 228–233.

Stadler, T., Gavryushkina, A., Warnock, R. C., Drummond, A. J., and Heath, T. A. (2018). The fossilized birth-death model for the analysis of stratigraphic range data under different speciation modes. *Journal of theoretical biology*, **447**, 41–55.

Sterli, J., Pol, D., and Laurin, M. (2013). Incorporating phylogenetic uncertainty on phylogeny-based palaeontological dating and the timing of turtle diversification. *Cladistics*, **29**(3), 233–246.

Tabor, C. R., Bardeen, C. G., Otto-Bliesner, B. L., Garcia, R. R., and Toon, O. B. (2020). Causes and climatic consequences of the impact winter at the Cretaceous-Paleogene boundary. *Geophysical Research Letters*, **47**(3), e60121.

Viglietti, P. A., Benson, R. B., Smith, R. M., Botha, J., Kammerer, C. F., Skosan, Z., Butler, E., Crean, A., Eloff, B., Kaal, S., *et al.* (2021). Evidence from South Africa for a protracted end-Permian extinction on land. *Proceedings of the National Academy of Sciences*, **118**(17).

Wagenmakers, E.-J. and Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review*, **11**(1), 192–196.

Ward, P., Botha, J., Buick, R., De Kock, M., Erwin, D., Garrison, G., Kirschvink, J., and Smith, R. (2005). Abrupt and gradual extinction among Late Permian land vertebrates in the Karoo Basin, South Africa. *Science*, **307**(5710), 709.

Wertheim, J. O. and Sanderson, M. J. (2010). Estimating diversification rates: how useful are divergence times? *Evolution*, **65**(2), 309–320.

Wood, A. R., Zelditch, M. L., Rountrey, A. N., Eiting, T. P., Sheets, H. D., and Gingerich, P. D. (2007). Multivariate stasis in the dental morphology of the Paleocene-Eocene condylarth *Ectocion*. *Paleobiology*, **33**(2), 248–260.

Young, K. (1960). Biostratigraphy and the new paleontology. *Journal of Paleontology*, **34**(2), 347–358.

Zeitoun, V. (2015). Les paléoanthropologues sont-ils en meilleure position que les autres pour se permettre d'ignorer les règles de la systématique? *Un bref historique. Biosystema*, **30**, 155–172.

Zhang, C., Stadler, T., Klopfstein, S., Heath, T. A., and Ronquist, F. (2016). Total-evidence dating under the fossilized birth-death process. *Systematic Biology*, **65**(2), 228–249.

A Probability density of a topology with fossils under the skyline FBD model

Let us start by recalling some useful probabilities, established in Stadler (2010); Didier *et al.* (2017); Didier and Laurin (2020), under the standard "single slice" FBD model with speciation, extinction and fossilization rates λ, μ and ψ, respectively, and sampling probability ρ at the ending time.

The probability that a single lineage starting at time 0 has n descendants at time $t > 0$ without leaving any fossil (i.e., neither from itself nor from any of its descendants) dated between 0 and t is given by

$$\mathbf{P}(0,t) = \frac{\alpha\beta(1-\mathrm{e}^{\omega t})}{\beta-\alpha\mathrm{e}^{\omega t}} \text{ and } \mathbf{P}(n,t) = \frac{(\beta-\alpha)^2\mathrm{e}^{\omega t}\left(1-\mathrm{e}^{\omega t}\right)^{n-1}}{\left(\beta-\alpha\mathrm{e}^{\omega t}\right)^{n+1}} \text{ for all } n>0,$$

where $\alpha < \beta$ are the roots of $-\lambda x^2 + (\lambda+\mu+\psi)x - \mu = 0$, which are always real and are equal to

$$\frac{\lambda+\mu+\psi \pm \sqrt{(\lambda+\mu+\psi)^2 - 4\lambda\mu}}{2\lambda}$$

and $\omega = -\lambda(\beta-\alpha)$.

The probability density that a single lineage starting at time 0 has $n > 0$ descendants at time $t > 0$, one of them leaving a fossil dated exactly at t, without leaving any other fossil dated between 0 and t is $n\psi\mathbf{P}(n,t)$ (Didier *et al.* 2017).

Let us now consider the skyline FBD model with ℓ intervals $[t_0,t_1]$, $[t_1,t_2]$, …, $[t_{\ell-1},t_\ell]$ and the corresponding parameters $(\lambda_1,\mu_1,\psi_1,\rho_1)$, $(\lambda_2,\mu_2,\psi_2,\rho_2)$, …, $(\lambda_\ell,\mu_\ell,\psi_\ell,\rho_\ell)$. Let us define α_1, …, α_ℓ, β_1, …, β_ℓ and ω_1, …, ω_ℓ accordingly and, for all $1 \leq i \leq \ell$,

$$\mathbf{P}_i(0,t) = \frac{\alpha_i\beta_i(1-\mathrm{e}^{\omega_i t})}{\beta_i-\alpha_i\mathrm{e}^{\omega_i t}} \text{ and } \mathbf{P}_i(n,t) = \frac{(\beta_i-\alpha_i)^2\mathrm{e}^{\omega_i t}\left(1-\mathrm{e}^{\omega_i t}\right)^{n-1}}{\left(\beta_i-\alpha_i\mathrm{e}^{\omega_i t}\right)^{n+1}} \text{ for all } n>0,$$

A lineage is said *observable* at time t if itself or one of its descendants left a fossil dated after t or is sampled at the present time. The reconstructed tree of Figure 1b is made of the observable parts of the lineages of the tree of Figure 1a. Let us put c_i for the probability for a lineage present just before t_i, i.e., an infinitesimal time before the sampling, to be not observable at this time. The probability $p^{(i)}(t)$ for a lineage present at $t \in [t_{i-1},t_i]$ to have no descendant at the ending time t_ℓ without leaving any fossil of age posterior to t, i.e., to be not observable, under the skyline FBD model above is

$$p^{(i)}(t) = \mathbf{P}_i(0,t_{i+1}-t) + \sum_{j=1}^{\infty}\mathbf{P}_i(j,t_{i+1}-t)c_i^j = \frac{\alpha_i(\beta_i-c_i)-\beta_i(\alpha_i-c_i)\mathrm{e}^{\omega(t_{i+1}-t)}}{\beta_i-c_i-(\alpha_i-c_i)\mathrm{e}^{\omega(t_{i+1}-t)}}$$

where $c_i = 1-\rho_i+\rho_i p^{(i+1)}(t_{i+1}) = 1-\rho_i(1-p^{(i+1)}(t_{i+1}))$ if $i < \ell$ and $c_\ell = 1-\rho_\ell$ (Gavryushkina *et al.* 2014; Zhang *et al.* 2016). For all $1 \leq i \leq \ell$ and all $t \in [t_{i-1},t_i]$, the probability for a lineage present at t to be observable at t is $\mathbf{P_o}(t) = 1-p^{(i)}(t)$.

Let $t_{i-1} \leq s < e \leq t_i$. The probability of getting $n > 0$ lineages observable at time e by starting with a single lineage at time s without observing any fossil between s and e is

$$\begin{aligned}\mathbf{P}_A(n,s,e) &= \sum_{j=0}^{\infty} \mathbf{P}_i(j+n, e-s)\binom{j+n}{n}\mathbf{P}_{\mathbf{o}}(e)^n\left(1-\mathbf{P}_{\mathbf{o}}(e)\right)^j \\ &= \frac{(\beta_i-\alpha_i)^2 \mathrm{e}^{\omega_i(e-s)}\mathbf{P}_{\mathbf{o}}(e)^n\left(1-\mathrm{e}^{\omega_i(e-s)}\right)^{n-1}}{\left[\beta_i-(1-\mathbf{P}_{\mathbf{o}}(e))-(\alpha_i-(1-\mathbf{P}_{\mathbf{o}}(e)))\,\mathrm{e}^{\omega_i(e-s)}\right]^{n+1}}\end{aligned}$$

Still by assuming that $t_{i-1} \leq s < e \leq t_i$, the probability density of observing a fossil find of a lineage dated at e with $n-1$ other lineages observable at time e by starting with a single lineage at time s without observing any fossil between s and e, is

$$\begin{aligned}\mathbf{P}_B(n,s,e) &= \sum_{j=0}^{\infty} \psi(j+n)\mathbf{P}_i(j+n, e-s)\binom{j+n-1}{n-1}\mathbf{P}_{\mathbf{o}}(e)^{n-1}\left(1-\mathbf{P}_{\mathbf{o}}(e)\right)^j \\ &= \frac{n\psi(\beta_i-\alpha_i)^2 \mathrm{e}^{\omega_i(e-s)}\mathbf{P}_{\mathbf{o}}(e)^{n-1}\left(1-\mathrm{e}^{\omega_i(e-s)}\right)^{n-1}}{\left[\beta_i-(1-\mathbf{P}_{\mathbf{o}}(e))-(\alpha_i-(1-\mathbf{P}_{\mathbf{o}}(e)))\,\mathrm{e}^{\omega_i(e-s)}\right]^{n+1}}\end{aligned}$$

Under a lineage-homogeneous process with no extinction (so is the reconstructed process of the skyline FBD process), the probability of observing a particular topology $\mathcal{T}$ conditioned on its number of tips, is defined as $\mathbf{T}(\mathcal{T}) = 1$ if $|\mathcal{T}| = 1$, i.e., $\mathcal{T}$ is a single lineage. Otherwise, by putting a and b for the two direct descendants of the root of $\mathcal{T}$, we have that

$$\mathbf{T}(\mathcal{T}) = \frac{2|\mathbf{L}_{\mathcal{T}_a}|!|\mathbf{L}_{\mathcal{T}_b}|!}{(|\mathbf{L}_{\mathcal{T}}|-1)|\mathbf{L}_{\mathcal{T}}|!}\mathbf{T}(\mathcal{T}_a)\mathbf{T}(\mathcal{T}_b),$$

where $|\mathbf{L}_{\mathcal{T}}|$ is the number of tips of $\mathcal{T}$. This probability distribution is called the Yule-Harding distribution (Harding 1971; Didier and Laurin 2020).

Our computations are based on two types of subparts of the skyline FBD process, called *patterns A and B*, which both start with a single lineage and end with two different types of configurations (Fig. A.1). Both patterns A and B are given by a 4-tuple $(s, e, \mathcal{T}, n)$ such that

- s and e are both contained in a same interval of the skyline FBD model and verify $s < e$;
- $\mathcal{T}$ is a topology with n tips.

The pattern A given by the 4-tuple $(s, e, \mathcal{T}, n)$ starts with a single lineage at time s and ends with n lineages observable at time e, their evolutionary relationships being represented by the topology $\mathcal{T}$ (without divergence times). Its probability is $\mathbf{P}_A(n, s, e) \times \mathbf{T}(\mathcal{T})$, that is the product of the probability of its ending configuration and of the probability of the topology $\mathcal{T}$ in the Yule-Harding distribution.

The pattern B given by the 4-tuple $(s, e, \mathcal{T}, n)$ starts with a single lineage at time s and ends with n lineages at time e, one of them having a fossil dated at e, the evolutionary relationships between the ending lineages being represented by the topology $\mathcal{T}$. Its probability density is $\mathbf{P}_B(n, s, e) \times \mathbf{T}(\mathcal{T})$, that is the product of the probability density of its ending configuration and of the probability of its topology in the Yule-Harding distribution.

The computation of the probability density of a topology with fossils exploits the fact that the skyline FBD process is Markov, thus that, conditionally to be present at a time t, the fate of a lineage after t is independent from the rest of the process.

A first consequence of the Markov property is that since a fossil ensures that the corresponding lineage was present at its age, we can write the probability density of a topology with fossils as the product of the probability densities of all the "basic topologies with fossils" obtained by splitting the initial topology at each fossil. Note that all the fossils of a basic topology are by construction on its tips. This first stage of the computation is illustrated on Figure A.1b, where a topology with three fossils is split into four basic topologies (the third one being empty).

The second stage is to compute the probability densities of the basic topologies obtained in the first stage. We proceed in a way similar to that used in Didier and Laurin (2020); Didier (2021) and we

just sketch the idea of the computation. We remark that the fossil ages and the interval bounds of the skyline FBD model encompassed by a basic topology are both stopping times of the process. The Markov property ensures that the parts of the process before and after one of these times are independent conditionally to the lineages present at this time. This is in particular true for the oldest time among the fossil ages and the interval bounds encompassed by the basic topology. A difficulty here is that since we don't have the divergence times, we do not know if some parts of the process are before or after this oldest time. In the example of Figure A.1c, the oldest time is the fossil age f_1, which implies that the two ancestral divergences of the corresponding fossil are anterior to this time but the two other divergences may occur before or after f_1. We have to consider all the possible ways of placing these two divergences with regard to f_1. Since these possibilities are mutually exclusive, the probability density of the initial basic topology is the sum of these associated to each possibility (second column of Fig. A.1c). From the Markov property, the probability density of these possibilities, in which all the relative positions of the divergences with regard to f_1 are fixed, is the product of the probability density of the part of the process which occurred before f_1, which is by construction a pattern B fully included in an interval of the skyline FBD model, and of the probability densities of the parts of the process which occurred after f_1, which are (smaller) basic topologies starting from f_1 (third column of Fig. A.1c), which can be recursively computed in the same way (last two columns of Fig. A.1c). In the case where the oldest time is a bound of an interval of the model, we proceed in the same way but the part of the process which is before this time is by construction a pattern of type A (last column of Fig. A.1c). The fact that the basic topologies encountered during the computation go smaller and smaller ensures that the computation eventually terminates.

Note that the number of possibilities to consider in the computation may be exponential but we use the same trick as in Didier (2021) to factorize this computation in order to get a polynomial algorithmic complexity.

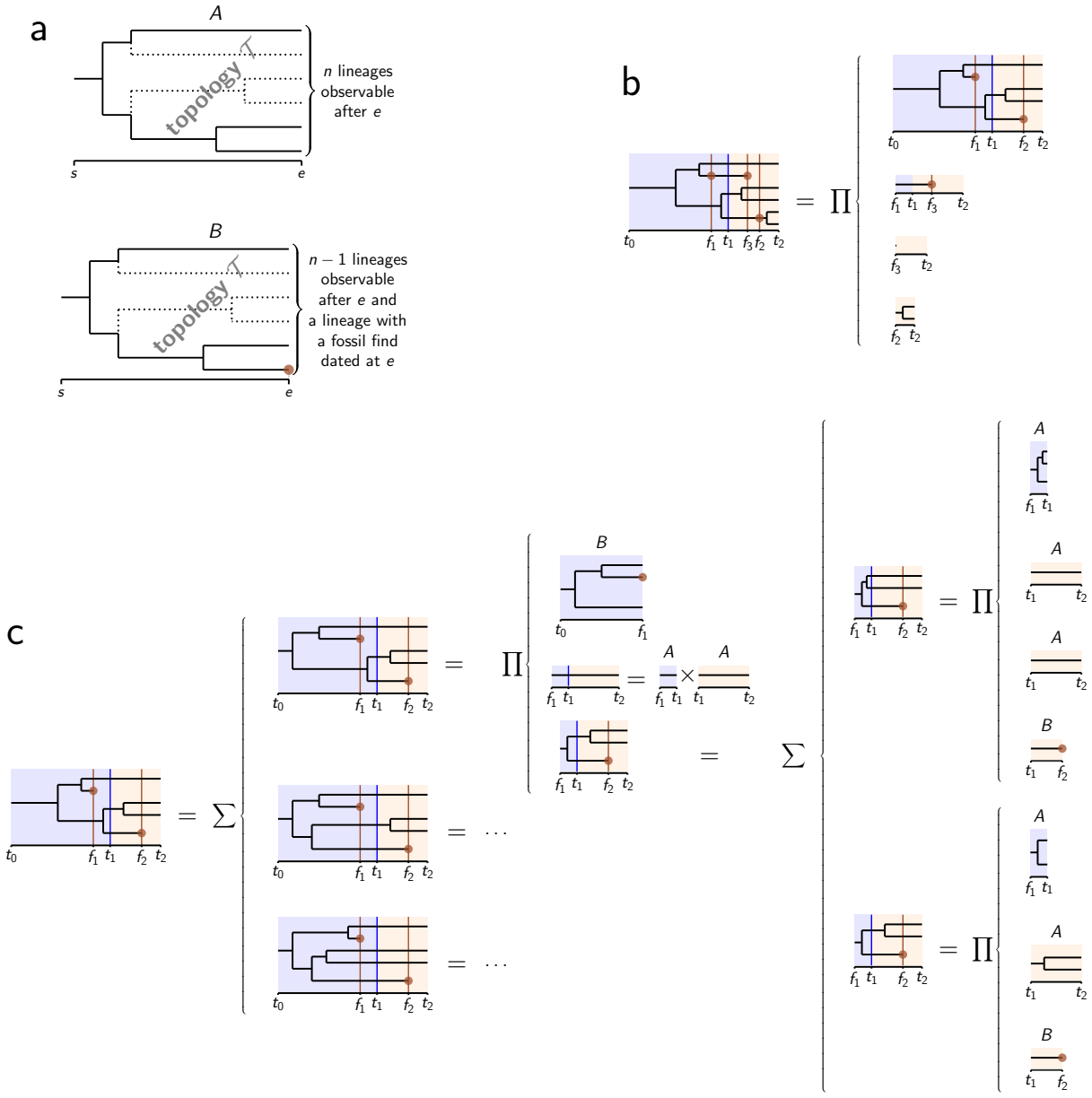

Figure A.1: a) the (diversification) patterns A and B; b) decomposition of the probability density of a tree topology with fossils as a product of probability densities of basic topologies; c) decomposition of the probability density of the first basic topology of (b) as a sum-product of patterns represented with their type (either A or B) on top of them. Bullets "•" represent fossils. Probability densities are computed under a skyline FBD model with the two intervals $[t_0, t_1]$ and $[t_1, t_2]$.

B Maximum likelihood estimates

Fixed (i.e., not estimated) parameters are in gray font.

Model specifications considered on the simulated dataset (Fig. 2)

$\mathcal{M}_0$

λ	μ	ψ	ρ
	350		
0.020079	0.015619	0.020400	1.000000
	0		

$\mathcal{M}_1$

λ	μ	ψ	ρ
	350		
0.027559	0.014278	0.018138	1.000000
	250		
0.062272	0.014278	0.018138	1.000000
	200		
0.013098	0.014278	0.018138	1.000000
	0		

$\mathcal{M}_1^e$

λ	μ	ψ	ρ
	350		
0.045215	0.013830	0.019330	0.038154
	250		
0.104704	0.013830	0.019330	1.000000
	200		
0.014057	0.013830	0.019330	1.000000
	0		

Model specifications considered on the empirical dataset (Fig. 3)

$\mathcal{M}_0$

λ	μ	ψ	ρ
	320		
0.120881	0.126431	0.229720	1.000000
	0		

$\mathcal{M}_1$

λ	μ	ψ	ρ
	320		
0.278303	0.157598	0.109370	1.000000
	298.9		
0.045376	0.126009	0.285345	1.000000
	0		

$\mathcal{M}_1^e$

λ	μ	ψ	ρ
	320		
0.352242	0.083203	0.112049	0.491451
	298.9		
0.036674	0.113780	0.251673	1.000000
	0		

$\mathcal{M}_2$

λ	μ	ψ	ρ
	320		
0.205211	0.146072	0.109980	1.000000
	293.5		
0.021534	0.098555	0.352608	1.000000
	0		

$\mathcal{M}_2^e$

λ	μ	ψ	ρ
	320		
0.230608	0.125493	0.117923	0.804927
	293.5		
0.030445	0.113486	0.267584	1.000000
	0		

$\mathcal{M}_3$

λ	μ	ψ	ρ
	320		
0.168868	0.115746	0.123424	1.000000
	290.5		
0.023005	0.109557	0.330108	1.000000
	0		

$\mathcal{M}_3^e$

λ	μ	ψ	ρ
	320		
0.170746	0.099920	0.112914	0.821527
	290.5		
0.029859	0.136193	0.336135	1.000000
	0		

$\mathcal{M}_4$

λ	μ	ψ	ρ
	320		
0.149066	0.096882	0.207837	1.000000
	283.3		
0.015500	0.231170	0.338781	1.000000
	0		

$\mathcal{M}_4^e$

λ	μ	ψ	ρ
	320		
0.141614	0.081646	0.186913	0.961562
	283.3		
0.014225	0.232888	0.265858	1.000000
	0		

$\mathcal{M}_5$

λ	μ	ψ	ρ
	320		
0.115730	0.095206	0.219508	1.000000
	274.4		
0.177806	0.401765	0.057986	1.000000
	0		

$\mathcal{M}_5^e$

λ	μ	ψ	ρ
	320		
0.115757	0.096317	0.210628	0.479792
	274.4		
0.060875	0.505949	0.160279	1.000000
	0		