Syst. Biol. 63(3):349–367, 2014

DOI:10.1093/sysbio/syu006
Advance Access publication February 8, 2014

Bayesian Estimation of Speciation and Extinction from Incomplete Fossil Occurrence Data

DANIELE SILVESTRO[1,2,3,*], JAN SCHNITZLER[4,5], LEE HSIANG LIOW[6], ALEXANDRE ANTONELLI[3,7], and NICOLAS SALAMIN[1,2]

[1]*Department of Ecology and Evolution, University of Lausanne, 1015 Lausanne, Switzerland;*
[2]*Swiss Institute of Bioinformatics, Quartier Sorge, 1015 Lausanne, Switzerland;*
[3]*Department of Plant and Environmental Sciences, University of Gothenburg, Carl Skottsbergs gata 22B, 413 19 Gothenburg, Sweden;*
[4]*Biodiversity and Climate Research Centre & Senckenberg Research Institute, Senckenberganlage 25, 60325 Frankfurt am Main, Germany;*
[5]*Department of Biological Sciences, Goethe University, Max-von-Laue-Str. 13, 60438 Frankfurt am Main, Germany;*
[6]*Centre for Ecological and Evolutionary Synthesis, Department of Biosciences, University of Oslo, N-0316 Oslo, Norway; and*
[7]*Gothenburg Botanical Garden, Carl Skottsbergs gata 22A, 413 19 Gothenburg, Sweden*
[*]*Correspondence to be sent to Daniele Silvestro, Department of Ecology and Evolution, University of Lausanne, 1015 Lausanne, Switzerland; E-mail: daniele.silvestro@unil.ch.*

Received 23 August 2013; reviews returned 21 November 2013; accepted 4 February 2014
Associate Editor: Luke Harmon

Abstract.—The temporal dynamics of species diversity are shaped by variations in the rates of speciation and extinction, and there is a long history of inferring these rates using first and last appearances of taxa in the fossil record. Understanding diversity dynamics critically depends on unbiased estimates of the unobserved times of speciation and extinction for all lineages, but the inference of these parameters is challenging due to the complex nature of the available data. Here, we present a new probabilistic framework to jointly estimate species-specific times of speciation and extinction and the rates of the underlying birth–death process based on the fossil record. The rates are allowed to vary through time independently of each other, and the probability of preservation and sampling is explicitly incorporated in the model to estimate the true lifespan of each lineage. We implement a Bayesian algorithm to assess the presence of rate shifts by exploring alternative diversification models. Tests on a range of simulated data sets reveal the accuracy and robustness of our approach against violations of the underlying assumptions and various degrees of data incompleteness. Finally, we demonstrate the application of our method with the diversification of the mammal family Rhinocerotidae and reveal a complex history of repeated and independent temporal shifts of both speciation and extinction rates, leading to the expansion and subsequent decline of the group. The estimated parameters of the birth–death process implemented here are directly comparable with those obtained from dated molecular phylogenies. Thus, our model represents a step towards integrating phylogenetic and fossil information to infer macroevolutionary processes.[BDMCMC; biodiversity trends; Birth–death process; incomplete fossil sampling; macroevolution; species rise and fall.]

Global temporal patterns of species diversity are governed by the complex interactions of biotic and abiotic factors, but they are ultimately the result of speciation and extinction. The dynamics of these processes have traditionally been inferred using data from the fossil record (Simpson 1944; Stanley 1979), but advances in molecular phylogenetics have created opportunities for estimating diversification rates from the phylogenetic hypotheses of extant organisms (Harvey et al. 1994; Sanderson and Donoghue 1996; Paradis 2004). To disentangle the two components of the diversification process, speciation and extinction are usually modeled as a stochastic process. The most commonly used model is the birth–death process (Harvey et al. 1994; Nee et al. 1994; Bokma 2003) in which speciation and extinction occur as random events according to constant or varying rates (Rabosky 2006; Maddison et al. 2007; Alfaro et al. 2009; FitzJohn et al. 2009; Morlon et al. 2011; Stadler 2011; Etienne et al. 2012; Stadler and Bokma 2013). These rates and their variation through time and across clades are usually estimated by maximum likelihood methods, although Bayesian approaches have also been implemented (Bokma 2008; Ryberg et al. 2011; Silvestro et al. 2011; Rabosky et al. 2012).

When a birth–death process is estimated solely from extant taxa, the inference of speciation and extinction rates may be severely biased, in particular if they vary through time or among lineages (Nee 2006; Quental and Marshall 2010; Rabosky 2010). Part of these limitations have been overcome by recent conceptual and methodological improvements, providing significant progress towards understanding past evolutionary dynamics (e.g., Antonelli and Sanmartín 2011; Litsios et al. 2012; Drummond et al. 2012). For example, negative net diversification rates, i.e., rates of extinction surpassing speciation, can now be estimated from molecular phylogenies, despite the fact that extinct lineages are generally not included in these phylogenies (Morlon et al. 2011; Stadler et al. 2013). Yet, neglecting fossil data means discarding useful information that can significantly improve the inference of macroevolutionary dynamics, especially when integrated with phylogenetic hypotheses of extant taxa (Didier et al. 2012; Ronquist, Klopfstein, et al. 2012; Fritz et al. 2013; Slater et al. 2013; Hunt 2013). The importance of using the fossil record to reconstruct the dynamics of diversification is particularly evident in the case of taxa that were very diverse in the past, but are represented today by very few species, if any (e.g., Bapst et al. 2012).

There is a long history of inferring speciation (or origination in the case of higher taxa) and extinction rates, generally based on birth–death models, using first and last appearances of taxa observed in the fossil record (Kurtén 1954; Raup 1975, 1991; Sepkoski 1998; Foote 2000,

2003; Ezard et al. 2011). Given that the times of first and last appearances of fossil taxa cannot be assumed to reflect the true times of speciation/origination and extinction (Strauss and Sadler 1989; Marshall 1990), various approaches have been developed to estimate diversification rates from incomplete data (Foote and Raup 1996; Foote 2000, 2003). In recent years, "internal" occurrences sampled within the time between first and last appearances of taxa have become available (Alroy 2010), which may be used to model the processes of fossil preservation and sampling (Nichols and Pollock 1983; Connolly and Miller 2001; Liow et al. 2008). Alternatively, data may be standardized to reduce the confounding effect of temporally heterogeneous sampling before parameters of the birth–death or other models are estimated (Alroy et al. 2008; Cermeno 2012). While the volume of available fossil data available continues to increase, most paleontological studies use genera as a proxy for species in the inference of diversity dynamics (Alroy 2010), although this might be inappropriate (Lloyd et al. 2012). Diversification rates estimated with this approach are therefore not directly comparable with those derived from birth–death models used in phylogenetics.

Understanding the dynamics of species diversification and extinction through time and across clades is of key importance in evolutionary biology, not least because it will aid in identifying what drives these processes. In order to facilitate direct comparisons between extinct and extant clades and to integrate paleontological and molecular data, it is desirable to apply the same birth–death models on both fossil and phylogenetic data. To this purpose, we develop a Bayesian approach to investigate processes of speciation and extinction based on all available fossil occurrences identified at the species level for a given taxon. Fossil occurrences are modeled as the result of two processes: sampling and species diversification. Sampling, in our definition, includes all historical and geological conditions that led to the preservation of an organism in the paleontological record and its subsequent sampling, description, and identification. Species diversification reflects the temporal changes in species richness due to varying speciation and extinction rates through time. The two processes are combined in a hierarchical Bayesian framework that jointly estimates the times of speciation and extinction of each sampled species and derives the parameters of the underlying birth–death process.

Our framework advances existing approaches by combining several key features. First, we utilize all available fossil occurrences, including those of extant taxa and fossil species that are known only from a single occurrence (hereafter referred to as "singletons"). Our model allows for continuous time and avoids the use of predefined, discrete time bins that are dictated by the geological record and do not necessarily correspond to biological processes (Foote and Miller 2007). The ages of the fossil occurrences are randomly drawn from their temporal ranges, and the procedure can be repeated to account for the associated uncertainty. Second, we explicitly model the temporal heterogeneity of sampling for each individual species in the fossil record. For any fossil species, there is a canonical pattern of decrease in sampled occurrences at both the start and the end of its lifetime (the "hat" trajectory; Liow, Skaug, et al. 2010), such that its first and last appearances may severely underestimate its true speciation and extinction times (Foote et al. 2007; Liow and Stenseth 2007; Liow, Quental, et al. 2010). We model this sampling pattern to estimate the speciation (and extinction) times of each species in the data set. In addition, we model the heterogeneity of the preservation rate in a lineage-specific manner, using a simple parameterization originally developed for substitution models in molecular phylogenetics (Yang 1994), and show that this also captures temporal rate variations. Third, in our birth–death model, extinction can exceed speciation, the rates are allowed to vary through time, and their changes are not assumed to be temporally linked (Foote 2001; Stadler 2011; Stadler et al. 2013). Furthermore, we do not assume the phylogenetic relationships among the fossil taxa to be known, thus greatly expanding the number of potential data sets that can be examined. Nevertheless, phylogenetic information could be used, if available, for example to partition the fossil data and assess differences in the diversification rates among subclades of a higher taxon. Within the hierarchical structure of the model, we include the probability that a clade undergoing the estimated birth–death process results in the number of species observed at the present. This strategy uses a formulation developed in a molecular phylogenetics context (Kubo and Iwasa 1995) and therefore introduces a formal link between the analysis of neontological and paleontological data. Finally, all the aspects outlined above are integrated in a joint Bayesian framework to estimate the model parameters while incorporating several sources of uncertainty. The resulting posterior distributions of the parameters are therefore comparable to those obtained by the standard (Bayesian) phylogenetic methods for molecular dating (Thorne et al. 1998; Drummond et al. 2006; Ronquist, Teslenko et al. 2012) and estimation of diversification rates (Ryberg et al. 2011; Silvestro et al. 2011; Stadler et al. 2013).

We evaluate the robustness of the joint estimation of the times of speciation and extinction of individual species and rates of speciation and extinction using simulations that reflect commonly observed diversity dynamics. The data sets analyzed were simulated under a range of potential biases, including violations of the sampling assumptions, variable preservation rates, and incomplete taxon sampling. Furthermore, we demonstrate the application of our approach by evaluating the temporal dynamics of diversification of the mammal family Rhinocerotidae to explore how shifts in speciation and extinction rates have shaped the expansion and decline of this family.

METHODS

Probabilistic Model

Fossil occurrence data is here modeled as the result of two processes: the preservation that allows an organism to fossilize and be sampled, and the speciation and extinction that generate and shape species diversity. The method described below uses a set of dated fossil occurrences identified to species level as input data. The data set can include all species with at least one known fossil occurrence, including extant species. The number of extant species, with or without a fossil record, is also used as input data in the analysis. From these data we estimate: 1) the parameters of the preservation process, 2) the times of speciation and extinction of each species in the data set, and 3) the speciation and extinction rates and their variation through time. Fossil preservation is modeled by a nonhomogeneous Poisson process (NHPP), in which the rate parameter is a function of time, whereas the species diversification is modeled by a birth–death stochastic process. These processes are combined for joint parameter estimation in a hierarchical Bayesian framework. In the following, we describe the different components of the model.

Times of speciation and extinction.—Let us assume that the probability of a sampled fossil occurrence is a function of a NHPP with a rate parameter $Q(t)$ that represents the probability of sampling at time t, and includes multiple processes such as the death and subsequent fossilization of an organism and its modern day sampling and identification. Since the number of fossil occurrences is known to decline close to the start and end of the lifetime of a taxon (Liow, Skaug, et al. 2010), beta distributions have been used to describe the probability of sampling a species through time after rescaling its lifespan to a time range of 0 to 1 (Liow, Quental, et al. 2010). To reproduce this pattern, we model the sampling probability $Q(t)$ using a generalized form of the beta distribution that can take any range of values between two boundaries, often referred to as a PERT distribution (Program Evaluation and Review Technique; Vose 2008). Let us consider a species i that originates at an unknown time of speciation s_i and becomes extinct at an unknown time of extinction e_i. In our notation, the ages of all events are measured as time before the present, so that $s_i > e_i$. The probability of sampling at time t is:

$$Q(t) = (s_i - e_i) q f_{PERT}(t; s_i, e_i) \quad (1)$$

where q is the mean preservation rate, $s_i - e_i$ is the species lifespan, and $f_{PERT}(t; s_i, e_i)$ is the PERT distribution with boundaries s_i and e_i. The preservation rate q indicates the average expected number of fossil records per species in a time unit (e.g., 1 myr) and is at this point assumed to be equal for all species in a given data set (but see below). The sampling probability $Q(t)$ resulting from Equation (1) is $Q(t) = 0$ for $t \notin [s_i, e_i]$ and $Q(t) > 0$ for $t \in [s_i, e_i]$. In other words, this function approximates the "hat" trajectory (Liow, Quental, et al. 2010) for the preservation rate, with no fossil occurrence allowed outside of a species lifespan (defined by s_i and e_i). The probability density function of a PERT distribution with maximum c and shape parameter l is

$$f_{PERT}(t|s_i, e_i, c, l) = \frac{(s_i - e_i)^{-1-l} (s_i - t)^a (-e_i + t)^b}{\Gamma(a+1)\Gamma(b+1)/\Gamma(a+b+2)} \quad (2)$$

where $a = (c - e_i)l/(s_i - e_i)$ and $b = (s_i - c)l/(s_i - e_i)$. Here, we use the simplified form of the PERT, i.e., we set the maximum to $c = (s_i - e_i)/2$ and the shape parameter to $l = 4$ to obtain a symmetric "hat"-shaped function with density

$$f_{PERT}(t|s_i, e_i) = \frac{(s_i - t)^2 (-e_i + t)^2}{(s_i - e_i)^5 B(3,3)} \quad (3)$$

where $B(3,3)$ is the symmetric beta function with shape parameters equal to 3. Based on the NHPP, the likelihood of a species i with K fossil occurrences of ages $t_1 < t_2 < \ldots < t_K$ given $t_0 = s_i$, $t_{K+1} = e_i$, and a preservation rate q, and conditional on at least one sampled fossil occurrence is

$$P_{NHPP}(t_1, \ldots, t_K | s_i, e_i, q) = \frac{\prod_{j=0}^{K} \exp\left(-q(s_i - e_i) \int_{t_j}^{t_{j+1}} f_{PERT}(u|s_i, e_i)\, du\right) \prod_{j=1}^{K} Q(t_j)}{\rho_i} \quad (4)$$

where the first product describes the probability of the waiting times in which no fossils are sampled and the second product indicates the probability of the sampled fossil occurrence. The denominator ρ_i indicates the probability that at least one occurrence is sampled for the species i, i.e., given the NHPP process:

$$\rho_i = 1 - \exp\left(-q(s_i - e_i) \int_{s_i}^{e_i} f_{PERT}(u|s_i, e_i)\, du\right) \quad (5)$$

where $\exp(\cdot)$ represents the probability of zero fossil occurrences in the time $[s_i, e_i]$. Thus the likelihood of a data set X, i.e., a set of N species each including K_i occurrences, is:

$$P(X|\mathbf{s}, \mathbf{e}, q) = \prod_{i=1}^{N} P_{NHPP}(t^i_1, \ldots, t^i_{K_i} | s_i, e_i, q) \quad (6)$$

where $\mathbf{s} = s_1, \ldots, s_N$ and $\mathbf{e} = e_1, \ldots, e_N$ are the vectors of speciation and extinction times, respectively.

Extant species.—A fossil data set can include extant taxa (i.e. $t^i_K = 0$), in which case the reconstructed lifespan $[s_i, 0]$ represents only an unknown fraction of the entire species life. The PERT component of the fossilization rate becomes truncated at time 0 and cannot be defined since the parameter e_i is missing. To overcome this problem we use data augmentation (Tanner and Wing 1987) to generate unobserved times of extinction (the latent data

e_i^{DA}) based on the estimated extinction rate at time 0 i.e. μ_J (see Equation 11). We assume here that the extinction rate will remain constant in the future. The probability of an extinction event in time t given μ_J, according to the birth–death process (Kendall 1948, cf. Equation 9) is described by the density:

$$P(t|\mu_J) = \mu_J \exp(-\mu_J t) \tag{7}$$

from which the latent times of extinction can be simulated. While the augmented data provide a measure of the lifespan of an extant species (s_i, e_i^{DA}), the fossil record remains incomplete because fossils can be observed only in the past. Thus, the likelihood of an extant species with at least one fossil occurrence (Equation 4) is calculated only within the range $[s_i, 0]$:

$$P_{NHPP}(t_1, \ldots, t_K = 0 | s_i, e_i^{DA}, q) = \frac{\prod_{j=0}^{K-1} \exp\left(-q(s_i - e_i^{DA}) \int_{t_j}^{t_{j+1}} f_{PERT}(u|s_i, e_i^{DA}) du\right) \prod_{j=1}^{K-1} Q(t_j)}{1 - \exp\left(-q(s_i - e_i^{DA}) \int_{s_i}^{0} f_{PERT}(z|s_i, e_i^{DA}) dz\right)} \tag{8}$$

The process of data augmentation is performed iteratively during the posterior sampling and the latent data are generated from the current values of speciation times and extinction rates. In our implementation, simulated e_i^{DA} are sampled from the distribution (7) at the quantiles 0.125, 0.375, 0.625, 0.875 and likelihoods obtained under each of the four latent extinction times are averaged (Tanner and Wing 1987). This sampling strategy was chosen, instead of e.g., randomly sampling e_i^{DA} from Equation (7), because it yielded better mixing of the MCMC samples.

Diversification process: Birth–death prior.—The times of speciation and extinction (s_i and e_i) must be estimated for each species (i) in order to use them directly in Equation (6). A Bayesian estimation of the parameters $\mathbf{s}, \mathbf{e}$, and q requires the definition of the respective prior distributions. An appropriate prior for the times of speciation and extinction is the birth–death stochastic process that describes the tempo of species origination and disappearance with two parameters: speciation and extinction rates indicated with λ and μ, respectively.

The prior probability of $\mathbf{s}$ and $\mathbf{e}$ given λ and μ is calculated, based on (Keiding 1975), as

$$P(\mathbf{s}, \mathbf{e}|\lambda, \mu) \propto \lambda^B \mu^D e^{-(\lambda+\mu)S} \tag{9}$$

where B and D are the numbers of speciation and extinction events, and S is the total time lived (Foote and Miller 2007), which is summed over all species:

$$S = \sum_{i=1}^{N} s_i - e_i. \tag{10}$$

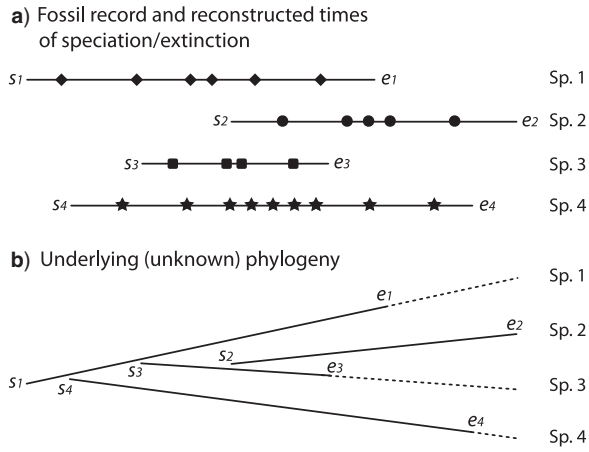

FIGURE 1. An example of fossil data set with four species. a) Fossil occurrences (indicated by symbols) are found as randomly sampled along each lineage according to a PERT-distributed rate. The lifespan of a species is defined by the estimated times of speciation and extinction (s_i, e_i). b) All species are assumed to be connected by an underlying (unknown) phylogeny, which we do not attempt to reconstruct, here represented as a "broken tree" in analogy with Fig. 1 of Nee (2001).

In this formulation, the species are treated as lineages of an underlying complete phylogeny (Fig. 1). The lineages can be connected in many different topologies, but these, as in the reconstructed process originally described by Nee et al. (1994), do not affect the estimation of speciation and extinction rates that is based exclusively on the times of speciation and extinction ($\mathbf{s}, \mathbf{e}$). Hence, the topology of the complete tree is here regarded as unknown.

The assumption of constant speciation and extinction rates (Equation 9) can be relaxed by introducing rate shifts through time (Stadler 2011; Stadler et al. 2013). We indicate the times of rate shifts with the vectors τ^Λ and τ^M of length $I-1$ and $J-1$ respectively to define I time frames with speciation rates $\Lambda = \lambda_1, \ldots, \lambda_I$ and J time frames with extinction rates $M = \mu_1, \ldots, \mu_J$. We set $\tau_I^\Lambda = \tau_J^M = 0$ and $\tau_0^\Lambda = \tau_0^M = s_{\max}$ where $s_{\max}$ is the time of origin of the process (here defined as the speciation time of the oldest taxon included in the analysis) Thus, we can rewrite the log probability of speciation and extinction times (Equation 9) as

$$\log p(\mathbf{s}, \mathbf{e}|\Lambda, M, \tau^\Lambda, \tau^M, s_{\max}) \propto \sum_{i=1}^{I} \left(\log(\lambda_i) B_i - \lambda_i S_{[\tau_{i-1}^\Lambda, \tau_i^\Lambda]}\right) + \sum_{j=1}^{J} \left(\log(\mu_j) D_j - \mu_j S_{[\tau_{j-1}^M, \tau_j^M]}\right) \tag{11}$$

where B_i and D_j are the number of speciation and extinction events in the time frames $[\tau_{i-1}^\Lambda, \tau_i^\Lambda]$ and $[\tau_{j-1}^M, \tau_j^M]$, respectively. Because the parameters of the birth–death process are not known, they must be estimated from the data, along with the number of rate shifts and their temporal placement. Times of shift

are sampled in continuous time and assumed to be independent for speciation and extinction.

Taxon sampling and birth–death hyperprior.—The Bayesian estimation of the speciation and extinction rates from the data requires the definition of a hyperprior on the birth–death parameters. The birth–death process described above can be used to estimate the birth–death parameters $(\Lambda, M, \tau^{\Lambda}, \tau^{M})$ underlying the diversification of a clade under the assumption that the number of speciation and extinction events (B, D) and the total time lived (S) are known. If all species of a clade are sampled, i.e. they appear with at least one occurrence in the fossil record, B and D are observed values and S is obtained by estimating the speciation and extinction times ($\mathbf{s}, \mathbf{e}$; Equations 4, 10). In the case of incomplete taxon sampling, the observed number of speciation and extinction events represents only a fraction of the total, and an estimation of the birth–death rates might be expected to be biased as observed in the case of extant taxa phylogenies (Yang and Rannala 1997; Stadler 2009). However, removing a random set of species from the sample has the effect of reducing both the births and deaths counts (B, D) as well as the total time lived S. This results in a flatter likelihood surface, but leaves the maximum likelihood point estimate of the rates unchanged (Supplementary Fig. S1 available at Dryad under http://dx.doi.org/10.5061/dryad.87d8s).

The effect of nonrandom incomplete sampling on the birth–death rate estimation is more difficult to predict and to correct for. In particular, the distribution of the true species richness through time of a given clade that includes the taxa not appearing in the fossil record, is unknown (cf. Discussion). Nonrandom taxon sampling can be generated by the preservation process itself, which tends to overrepresent long-living species as the preservation rate (q) decreases and short-living taxa are more likely to disappear entirely from the fossil record (Foote 2000). We investigated the potential bias deriving from low preservation rates and consequent incomplete taxon sampling empirically on simulated data sets (see below). Nonrandom taxon sampling is also expected to play an important role in particular if recently evolved species might not have had time to fossilize. However, we can assume that the true current species richness of a clade is the number of observed extant species, indicated by N_{OBS}. We construct an informative hyperprior for the birth–death parameters as the probability that a clade originating at time $s_{\max}$ results in N_{OBS} extant species after diversifying under a birth–death process defined by $\Lambda, M, \tau^{\Lambda}, \tau^{M}$. This is calculated using equations (1a–c) of (Kubo and Iwasa 1995), assuming a single starting lineage at time $s_{\max}$:

$$P_{N_{OBS}} = P(\Lambda, M, \tau^{\Lambda}, \tau^{M} | N_{OBS}, s_{\max}) = \frac{\beta}{\alpha(1+\alpha)} \left(\frac{\alpha}{1+\alpha} \right)^{N_{OBS}}, N_{OBS} > 0 \quad (12)$$

where

$$\alpha = \int_{s_{\max}}^{0} \lambda_u \exp\left[\int_{u}^{0} (\lambda_z - \mu_z) dz \right] du \quad (13)$$

and

$$\beta = \exp\left[\int_{s_{\max}}^{0} (\lambda_u - \mu_u) du \right]. \quad (14)$$

In the case of a clade that has gone completely extinct, i.e., $N_{OBS} = 0$, the hyperprior of Equation (12) becomes (Kubo and Iwasa 1995):

$$P_{N_{OBS}} = P(\Lambda, M, \tau^{\Lambda}, \tau^{M} | N_{OBS} = 0, s_{\max}) = \frac{1+\alpha-\beta}{\alpha(1+\alpha)}, N_{OBS} = 0 \quad (15)$$

The age of origin of the diversification process ($s_{\max}$), i.e., the "root age" of the clade from a phylogenetic perspective, is assumed here to be the time of speciation of the oldest sampled species.

Posterior.—After assigning Gamma (hyper-)priors to the preservation rate q and the time of origin of the process $s_{\max}$ (the latter truncated at the age of the oldest fossil in the data set), we can write the joint posterior of all parameters as:

$$P(\mathbf{s}, \mathbf{e}, \Lambda, M, \tau^{\Lambda}, \tau^{M}, s_{\max}, q | X, N_{OBS}) \propto P(X | \mathbf{s}, \mathbf{e}, q)\, P(\mathbf{s}, \mathbf{e} | \Lambda, M, \tau^{\Lambda}, \tau^{M}) P(q) P(\Lambda, M, \tau^{\Lambda}, \tau^{M} | N_{OBS}, s_{\max}) P(s_{\max}). \quad (16)$$

Heterogeneity of preservation rates across species.—The model of fossil preservation described by Equations (1–8) is based on a preservation rate q that is assumed constant through time and across species. While the preservation rates could ideally be estimated independently for each species in the data set, the number of per-species fossil occurrences might be insufficient for this estimation and the model of fossil preservation would be likely over-parameterized. Alternatively, the across-species variation of the preservation rate can be incorporated without attempting to estimate species-specific rates. This approach has been used in molecular phylogenetics to deal with the heterogeneity of mutation rates among the sites of aligned sequences (Yang 1993). We implemented the approximate method for rate heterogeneity proposed by Yang (1994) to model the variation of the preservation rates across species. That is, we assumed the rates of preservation to vary according to a Gamma distribution with an overall mean equal q. The Gamma distribution was approximated by four discrete categories with equal probabilities and the rescaled median values were applied as multipliers of the parameter q. The likelihood of each species was then calculated using Equation (4) under each of the four rates and averaged. The parameters of the Gamma

distribution were assumed to be equal (Yang 1994), assigned a uniform prior within the range [0, 20], and estimated from the data.

Sampling Algorithms and Implementation

The Bayesian framework described above was implemented using a Markov chain Monte Carlo (MCMC) with an acceptance ratio based on Equation (16) to sample the model parameters from their posterior distribution. In MCMC analyses the number of rate shifts is fixed *a priori*, but their temporal placement $(\tau^{\Lambda}, \tau^{M})$ is estimated from the data. Birth–death models with different number of rate shifts can be compared by their respective marginal likelihoods calculated by thermodynamic integration (TI; Lartillot and Philippe 2006). While the estimation of marginal likelihood via TI provides an accurate way to compare alternative hypotheses of diversification, its application can be time consuming and even inefficient considering the potentially infinite number of models that can be fitted to the data. Thus, an alternative algorithm called birth–death MCMC (BDMCMC; Stephens 2000) was developed to estimate the number of parameters of the model (i.e., the number of rate shifts) and the speciation and extinction rate in a joint analysis. This method represented a great improvement over the MCMC+TI algorithms in terms of computational time (see Appendix), and was therefore used in all the analyses presented below. The vectors of speciation and extinction times (**s** and **e**) were assumed to derive from an underlying model of diversification with unknown number of speciation rates ($k_{\lambda} \geq 1$) and extinction rates ($k_{\mu} \geq 1$). The number of rate shifts ($k_{\lambda} - 1; k_{\mu} - 1$) was thus estimated from the data. More details on the algorithms implemented for this study are provided in the Appendix. Marginal rates within one myr time bins were logged to obtain posterior estimates of speciation and extinction rates through time. Posterior distributions of individual rates (e.g., λ_i) or times of rate-shift (e.g., τ_i^{Λ}) could not be obtained in BDMCMC analyses, because the vectors Λ, M have variable lengths depending on the currently estimated number of rate shifts. Note that posterior distributions of individual rates or time of rate-shift can be obtained instead by running a standard MCMC analysis with fixed number of rates.

The methods described in this study were implemented in an open-source command-line program, PyRate, available at http://sourceforge.net/projects/pyrate/. PyRate is written in Python version 2.7 (www.python.org) and supports multi-thread computation of the likelihoods and parallel MCMC runs. It has been developed and tested on UNIX machines.

Simulations

The Bayesian framework described above was tested using simulated data. Data sets were generated by simulating the speciation and extinction of lineages (**s**, **e**) based on a stochastic birth–death process and assuming that new species are budding off the parent lineage (asymmetric speciation). Simulated data sets were generated based on six patterns of species diversification chosen to yield scenarios of varying species richness commonly observed in empirical data (Fig. 2 and Table 1; Sepkoski 1981):

I. expanding diversity with constant speciation and extinction rates ($\lambda > \mu$);

II. expanding diversity followed by a decline with all taxa going extinct before the present ($\lambda_1 > \mu_1$, $\lambda_2 < \mu_2$);

III. expanding and then declining diversity followed by turnover at equilibrium ($\lambda_1 > \mu_1$, $\lambda_2 < \mu_2$, $\lambda_3 = \mu_3$);

IV. expanding diversity followed by turnover at equilibrium due to a decrease in speciation rate ($\lambda_1 > \mu_1$, $\lambda_2 = \mu_1$);

V. expanding diversity followed by turnover at equilibrium due to a decrease in speciation rate and increase in extinction rate ($\lambda_1 > \mu_1$, $\lambda_2 = \mu_2$); and

VI. constant speciation rate and a mass extinction event ($\lambda_1 > \mu_1$, $\lambda_1 \ll \mu_2$, $\lambda_1 > \mu_1$).

Based on the complete birth–death realizations, fossil occurrences were simulated for each species. Along each lineage i a number of occurrences K_i was derived from a Poisson distribution with rate parameter $q_{POI} = q(s_i - e_i)$ where q is the preservation rate (cf. Equation 1), and s_i, e_i are the true times of speciation and extinction. The preservation rate was assumed $q = 3$ unless stated otherwise. A number of fossils K_i were then randomly drawn from the PERT distribution for all species, resulting in a synthetic data set that mimics the fossil record. The shape parameter of the PERT distribution was assumed to be $l = 4$ (as in Equation 3) unless stated otherwise. The number of occurrences K depended only on the preservation rate q and on the species lifespan, i.e., it was not conditioned on being greater than 0. The lineages without a fossil record ($K = 0$) were disregarded in the analyses because of the condition stated in Equation (4). All extant species were then truncated at time 0 (i.e., the present). The number of extant species based on the complete birth–death realization (N_{OBS}) was kept to construct the hyperprior of Equations (12–15).

Sensitivity analyses.—To assess the robustness of our approach to different violations of the assumptions and quality of the data, a range of 14 settings were used to sample the fossil occurrences from each birth–death realization (Table 2).

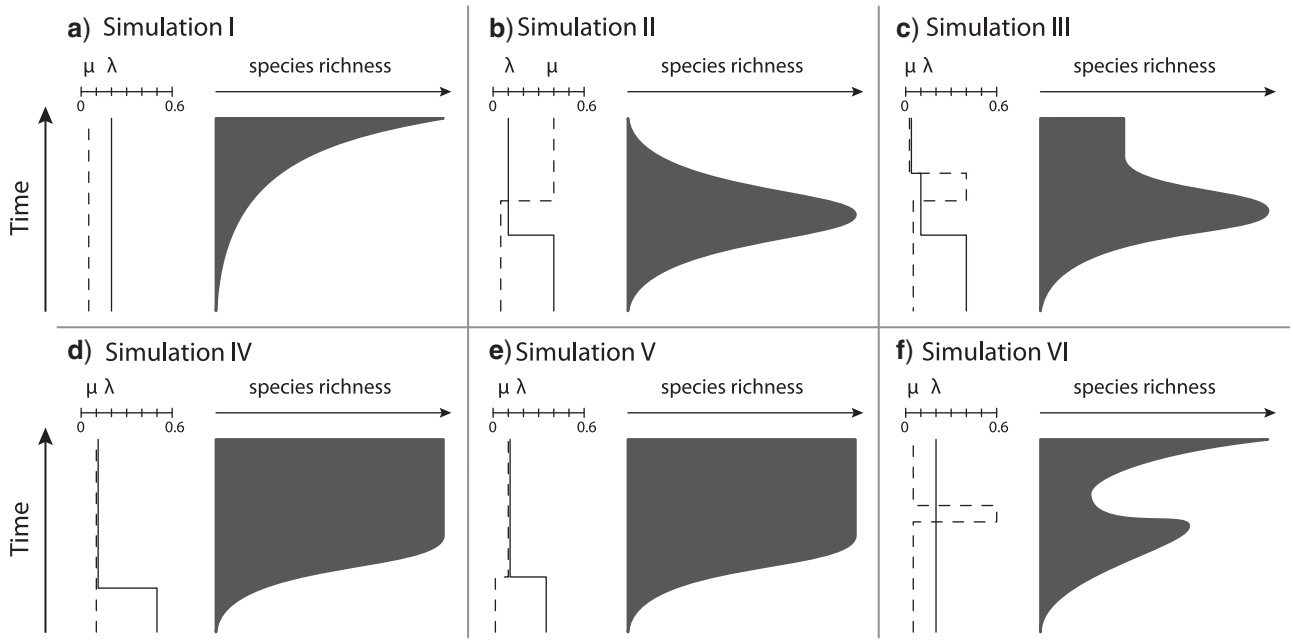

FIGURE 2. Simulated temporal patterns of species diversity. Six birth–death scenarios (a–f) were simulated to reproduce common patterns of increasing and/or declining species richness observed in the fossil record. Different variations of speciation (solid line) and extinction (dotted line) rates through time are shown on the left side of each panel (see text and Table 1 for details).

TABLE 1. Summary of birth–death simulations

Birth-death scenario	root age	Speciation rates (λ)	Extinction rates (μ)	Times of shift		Species min/max	Frac. extant min/max
				τ^{Λ}	τ^{M}		
I $(1_{\lambda}, 1_{\mu})$	45	0.15	0.07	NA	NA	106–270	0.50–0.63
II $(2_{\lambda}, 2_{\mu})$	30	0.4, 0.1	0.05, 0.4	20	15	111–273	0–0.03
III $(3_{\lambda}, 3_{\mu})$	35	0.4, 0.1, 0.01	0.05, 0.3, 0.01	20, 10	15, 10	119–289	0.10–0.18
IV $(2_{\lambda}, 1_{\mu})$	35	0.5, 0.1	0.1	27	NA	100–300	0.18–0.31
V $(2_{\lambda}, 2_{\mu})$	35	0.35, 0.1	0.01, 0.1	25	18	117–268	0.28–0.40
VI $(1_{\lambda}, 3_{\mu})$	35	0.2	0.05, 0.6, 0.05	NA	15, 12	122–294	0.37–0.69

Notes: Numbers between parentheses indicate the number of rates between shifts (e.g., $1_{\lambda}, 3_{\mu}$: 1 speciation rate, 3 extinction rates). The root age and the times of rate shifts are given in Ma. Ten birth–death realizations were generated from each scenario and the resulting number of species (and the fraction of extant) is reported as min and max values across replicates.

I. We varied the shape parameter (l) of the PERT distribution and sampled occurrences under a uniform distribution ($l=0$), a PERT distribution with the same parameter l as assumed in the model ($l=4$, cf. Equation 3), a PERT distribution with greater convexity ($l=10$), and finally a PERT distribution shape parameter randomly varying across species $l \in [0, 10]$.

II. We allowed the preservation rate to change across species by drawing species-specific rates q from a gamma distribution. Three different shape parameters were tested, namely $\Gamma(5), \Gamma(3)$, and $\Gamma(1)$, to simulate various degrees of rate heterogeneity, with variances equal to 0.2, 0.33, 1, respectively. In addition, a varying-through-time fossilization process was simulated by introducing a 60-fold linear decrease in the preservation rate from an initial value $q=3$ at the present to a value of $q=0.05$ at the origin of the clade.

III. The number of fossil records was reduced by decreasing the preservation rate that was set to $q=1, 0.5$, and 0.25. This resulted in data sets with increasingly poor taxon sampling (below 50% in the worst cases) and high fraction of singletons (up to around 30% of the species; Table 2).

IV. Data sets with incomplete taxon sampling were generated by removing from the complete birth–death realizations entire lineages at random to obtain sampling fractions of $\Psi=0.75, 0.50$, and 0.25. It is noted that the effective taxon sampling might further decrease after excluding the species that lacked a fossil record, i.e., with $K=0$ (Table 2).

For each of the 6 birth–death scenarios, 10 independent realizations were generated, resulting in data sets of

TABLE 2. Summary of the settings applied to sample fossil occurrences from each of the birth–death scenario

	Sampling settings				Effective sampling fraction	Occurrences per species	Proportion of singletons
	l	q	α	Ψ			
PERT shape	4	3	NA	1	0.76–0.98	10.77–29.49	0.05–0.14
	0	3	NA	1	0.89–0.99	13.60–30.24	0.01–0.10
	10	3	NA	1	0.70–0.98	9.50–30.35	0.01–0.13
	0–10	3	NA	1	0.74–0.99	10.46–29.75	0.01–0.12
Rate heterogeneity	4	3	1	1	0.63–0.93	9.90–33.59	0.01–0.09
	4	3	3	1	0.75–0.98	8.89–29.14	0.01–0.11
	4	3	5	1	0.79–0.98	9.98–32.81	0.01–0.10
	4	3–0.05	NA	1	0.55–0.86	3.53–12.36	0.05–0.21
Preservation rate	4	1	NA	1	0.68–0.94	3.56–10.27	0.05–0.21
	4	0.5	NA	1	0.55–0.88	1.68–5.05	0.11–0.26
	4	0.25	NA	1	0.43–0.75	0.97–2.65	0.16–0.34
Random taxon sampling	4	3	NA	0.75	0.58–0.75	7.63–20.55	0.01–0.09
	4	3	NA	0.50	0.40–0.50	5.36–15.99	0.00–0.06
	4	3	NA	0.25	0.19–0.25	2.41–10.20	0.00–0.05

Notes: Fossil occurrences were obtained from PERT distributions with different shape parameters (l), a range of preservation rates (q), and under different proportions of random taxon sampling (Ψ). The preservation rate was also varied across species according to a gamma distribution with shape parameter α. Summary statistics for the resulting data sets are calculated over all birth–death scenarios and provided as min and max values. "Effective sampling fraction" indicates the proportion of species that appear in the record with at least one occurrence.

100 to 300 species. Fossil occurrences were sampled for each of the 60 realizations of the birth–death under the 14 preservation and taxon sampling settings described above (Tables 1 and 2).

Statistical Evaluation

All simulated data sets (6 birth–death scenarios × 14 preservation and sampling settings × 10 replicates) were analyzed by BDMCMC with runs of 5,000,000 generations (see Appendix for more details). Efficient chain mixing and effective sample sizes were assessed by examining the log files in Tracer (Rambaut and Drummond 2007). The sampling frequencies of different numbers of rate shifts were estimated from the posterior sample to assess the power of the method to find the correct model of diversification. Marginal speciation and extinction rates within 1 myr time bins were summarized from the posterior samples as mean values and 95% highest posterior densities (HPD). These were displayed in rates-through-time plots and compared against the rates used to simulate the data. The results of the 10 replicates were pooled in a single plot showing the range of mean rate values and HPD upper and lower boundaries across replicates. Mean and 95% HPD of the preservation rate (q) were calculated for the 14 sampling settings (Table 2) averaging over all birth–death scenarios and over all replicates. A similar approach was adopted to assess the accuracy of the estimated age of origin of the data set (s_{max}) in which case the difference between the true values of s_{max} and the estimated values $\hat{s}_{max}$ was calculated and averaged over data sets and replicates.

The BDMCMC analyses were repeated on all data sets using the Gamma model for heterogeneity of the preservation rate. Rate-through-time plots (RTT) of the marginal rates were generated and the posterior sample of Gamma's shape parameter (α) was interpreted as a measure of the rate heterogeneity and compared against the true value used to simulate the data. While proper model testing to choose between constant-rate and Gamma models was not attempted, a much simpler (and computationally cheaper) approach was taken instead. Since increasing values of the shape parameter indicate decreasing rate heterogeneity (rates are constant with $\alpha = \infty$), we considered the Gamma model preferable over the constant rate model whenever α was found significantly smaller than the upper bound defined by the hyperprior range [0,20]. With a shape parameter $\alpha = 20$ the gamma distribution has a variance of 0.05.

A Fossil Data Set of the Family Rhinocerotidae

We used our approach to evaluate the diversification dynamics of the mammal family Rhinocerotidae. This family was selected for its extensive fossil record and long diversification history starting in the early Eocene (Tougard et al. 2001; Steiner and Ryder 2011). Rhinocerotidae also serve as an exemplary group for which phylogeny-based methods are expected to have limited analytical power due to the very little present diversity, i.e., five extant species. Fossil occurrences were assembled from the NOW database (Fortelius 2013), the Paleobiology Database (http://paleodb.org) (both downloaded on January 17th, 2013), and from Geraads (2010). All undetermined species and genera were removed from the data. The ages of most occurrences were provided with a temporal range (of size 0–6.28 myr) from which a random age was uniformly drawn for each occurrence. The data set (deposited at Dryad, http://dx.doi.org/10.5061/dryad.87d8s) comprised 164 species including the five extant species, with a total of

TABLE 3. Sampling frequencies of the number of rates (k_λ and k_μ) estimated by BDMCMC for the different birth–death scenarios sampled under preservation rate $q=3$ and averaged over ten replicates (standard deviation across replicates is given in parentheses if greater than 0.01)

Birth-death scenario	Parameter	Sampling frequency (number of rates k)				
		1	2	3	4	5
I ($1_\lambda, 1_\mu$)	λ	**0.80** (0.13)	0.17 (0.11)	0.02 (0.02)	0	0
	μ	**0.95** (0.03)	0.05 (0.02)	0	0	0
II ($2_\lambda, 2_\mu$)	λ	0.01 (0.01)	**0.66** (0.19)	0.28 (0.15)	0.05 (0.04)	0
	μ	0	**0.83** (0.12)	0.17 (0.12)	0.01 (0.01)	0
III ($3_\lambda, 3_\mu$)	λ	0.02 (0.01)	0.49 (0.11)	**0.39** (0.07)	0.09 (0.05)	0.01 (0.01)
	μ	0.02 (0.01)	0.04 (0.03)	**0.77** (0.21)	0.15 (0.13)	0.03 (0.05)
IV ($2_\lambda, 1_\mu$)	λ	0	**0.79** (0.13)	0.20 (0.12)	0.02 (0.01)	0
	μ	**0.97** (0.01)	0.03 (0.01)	0	0	0
V ($2_\lambda, 2_\mu$)	λ	0.02 (0.02)	**0.78** (0.08)	0.18 (0.06)	0.02 (0.01)	0
	μ	0.02 (0.02)	**0.92** (0.05)	0.05 (0.03)	0	0
VI ($1_\lambda, 3_\mu$)	λ	**0.47** (0.19)	0.45 (0.16)	0.08 (0.03)	0.01	0
	μ	0	0	**0.99** (0.02)	0.01 (0.02)	0

Note: Numbers in bold highlight the settings used to generate the data sets.

2,463 occurrences spanning 45 myr, with 35 species (21%) being represented by a single sampled fossil occurrence.

The data set was first analyzed by 10,000,000 BDMCMC generations both assuming constant preservation rate and under the Gamma model to assess whether heterogeneity of the preservation rate could be detected. The selection of the Gamma model followed the procedure described above, and ten additional BDMCMC analyses were performed after randomizing the age of fossil occurrences within their corresponding range intervals. The posterior samples were finally compared and combined to estimate the rates through time. This approach explicitly incorporates the uncertainty associated with the age of the fossil records, which generally relies on the estimated boundaries of stratigraphic units (Gradstein et al. 2012).

For comparison, phylogeny-based estimations of the diversification rates were obtained by applying the birth–death-shift model as implemented in TreePar (Stadler 2011). Maximum-likelihood analyses were carried out on the clade of extant rhinos extracted from a recently published molecular phylogeny of Perissodactyla dated with relaxed molecular clock and fossil calibrations (Steiner and Ryder 2011). Because one of the five extant rhino species was missing in the tree, the analyses were corrected for incomplete taxon sampling (Stadler 2011). The fit of birth–death models with different number of rate shifts were compared using the Akaike Information Criterion corrected by sample size (AICc; Akaike 1973), Akaike weights were used as measures of their respective relative likelihoods (Burnham and Anderson 2002).

RESULTS

Model Selection

Our analyses on the simulated data sets using BDMCMC show that the correct birth–death models consistently received strong support (i.e., high sampling frequency; Table 3 and Supplementary Table S2 available at Dryad under http://dx.doi.org/10.5061/dryad.87d8s), indicating that the BDMCMC is robust against over- or under-estimation of the number of rate shifts. Nevertheless a considerable uncertainty was detected in some cases around the number of shifts in speciation and extinction rates, k_λ and k_μ, respectively. For example, in data set III (Fig. 2c), models with one or two shifts in the speciation rate received similar support based on the sampling frequencies of k_λ (Table 3). For data sets sampled under strong violations of the model assumptions (e.g., under constant preservation rate along the species lifespan, sampling scheme 2; Table 2) or under very low preservation rates and taxon sampling, the accuracy of the model selection tended to decrease (Table 3 and Supplementary Table S1 available at Dryad under http://dx.doi.org/10.5061/dryad.87d8s). This led to a failure to detect small rate changes, highlighted by the second shift in the speciation rate of the birth–death scenario III. In addition, the speciation rate in the birth–death scenario VI was often estimated to undergo a modest decrease after the mass extinction event (Table 1, Fig. 3 and Supplementary Figs. S2–29 available at Dryad under http://dx.doi.org/10.5061/dryad.87d8s). However, the highest sampling frequency was found in favor of a false rate shift (thus representing "false positives") in only 3.5% of the analyses, and always with low support (0.40–0.55).

Parameter Estimation

Speciation and extinction rates.—The estimation of the birth–death parameters was accurate throughout the simulations (Fig. 3), and the correct trends of rate changes and their timing were always recovered (e.g., rate increases/decreases, turnover phases, negative/positive diversification). Cases of spikes of rate changes, such as the sudden increase in extinction rates causing a mass extinction (Fig. 3c,f), were correctly identified. While the correct trend of decline

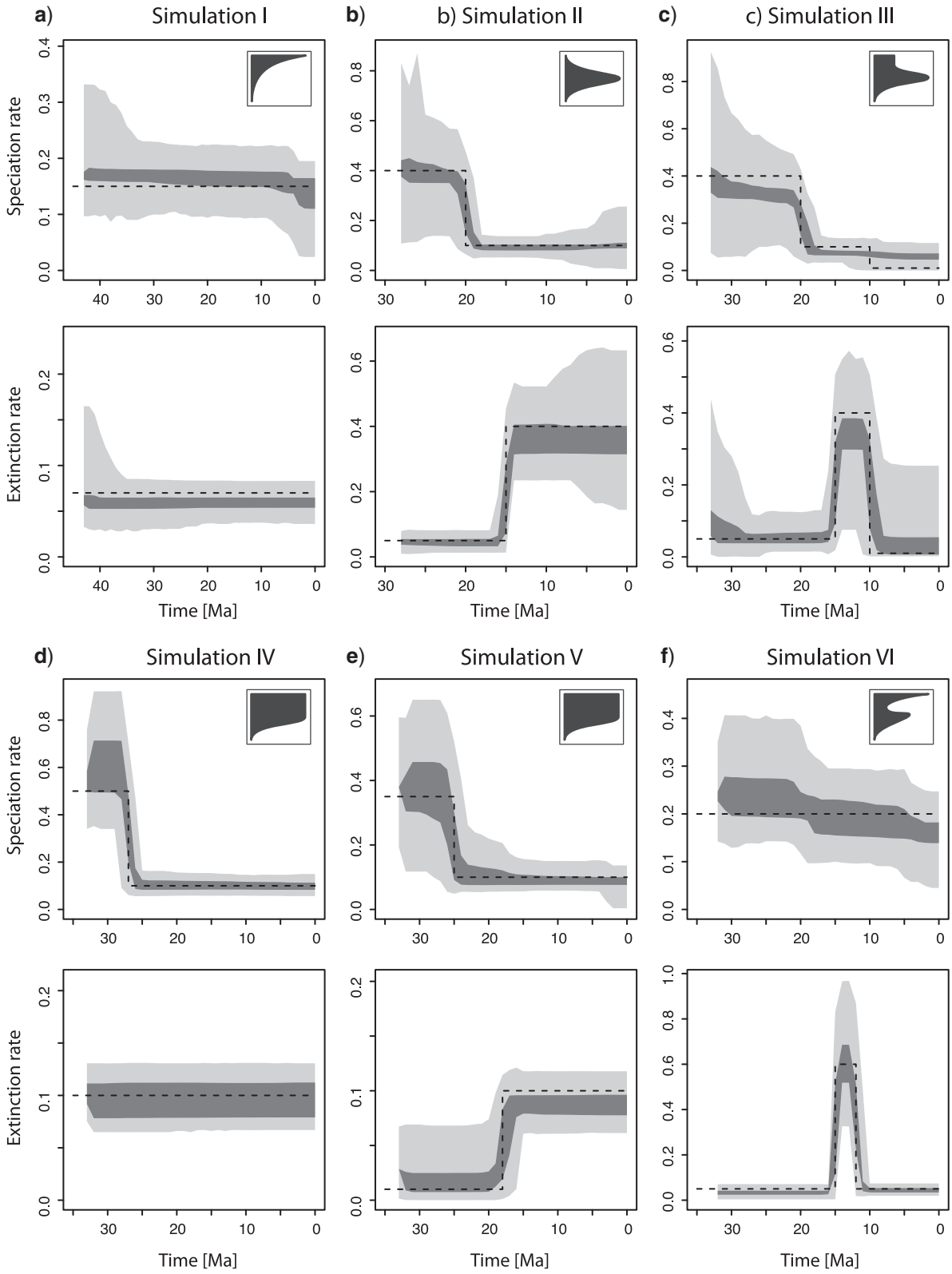

FIGURE 3. Rate-through-time plots. Marginal rates of speciation and extinction summarized over ten replicates for each simulated data set obtained through BDMCMC sampling. Dark gray areas show the range of mean values obtained from the different replicates and the gray shaded areas display the range of the associated 95% credibility intervals, black dashed lines indicate the parameters used in the simulations. The inserts in the speciation rate plots represent simplified diagrams of the simulated species richness dynamics as shown in Figure 1.

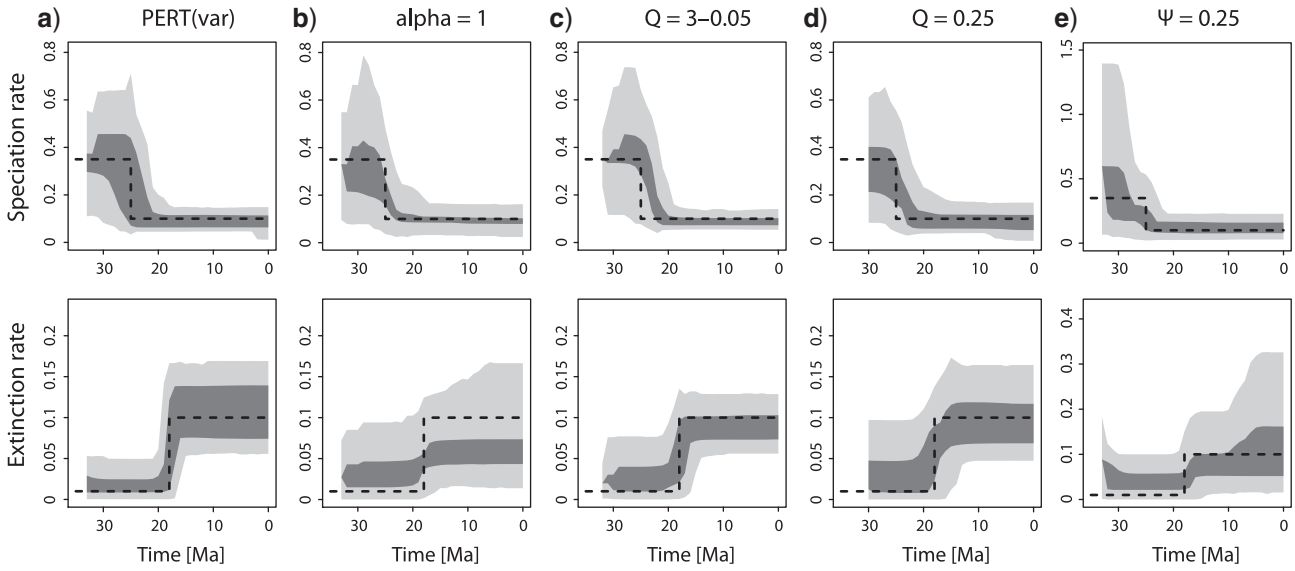

FIGURE 4. Rate-through-time plots. Marginal rates of speciation and extinction summarized over ten replicates for birth–death scenario V (cf. Fig. 2) under different sampling settings: a) variable shape parameter of the PERT distributed preservation rate (ranging from 0 to 10); b) strong heterogeneity of the preservation rate on a species-specific basis or c) through time; d) severely decreased preservation rate (yielding an effective sampling fraction of 43–75%); e) random incomplete taxon sampling (75% of the species removed). More details on the sampling settings used for fossil occurrences are provided in the Table 2.

of the speciation rate was reconstructed for the birth–death scenario III (Fig. 3c), identifying the two successive rate shifts was difficult (as shown by the sampling frequencies of the BDMCMC model selection; Table 3), and the magnitude of the rate variation tended to be slightly underestimated. The mean posterior rates were found to be generally within 10–20% variation across replicates and the size of the 95% HPDs was generally of the same order of magnitude as the rates and similar for both speciation and extinction (Fig. 3). It should be noted that the gray-shaded areas in the RTT plots displays the maximum extent of the credibility intervals over ten replicates, thus overestimating the size of the actual HPDs for a single analysis. Larger credibility intervals were generally observed when clades had smaller sizes, e.g., close to their origin (scenarios I and III) or close to the present (scenario II; Figs. 2 and 3a,b,c).

The posterior speciation and extinction rates are consistent throughout the different sampling settings, although model violations and decreased quality of the data had an impact on the accuracy of the estimates. The assumption of uniform preservation rates throughout a species' lifespan, i.e., violating the assumption of a PERT(4) "hat" distribution, led to an underestimation of the birth–death rates (Supplementary Figs. S3 and S17 available at Dryad under http://dx.doi.org/10.5061/dryad.87d8s), while variable shape parameters across species (ranging from uniform to very convex distributions, PERT(0–10)) did not affect substantially the accuracy of the posterior rates (Fig. 4a Supplementary Figs. S5 and S19 available at Dryad under http://dx.doi.org/10.5061/dryad.87d8s). Strong variations of the preservation rates across taxa or through time resulted in increased uncertainty around the rate estimates, hence yielding wider credibility intervals (Fig. 4b,c). The marginal rate estimated under the constant preservation rate model and under the Gamma model were found to be very similar throughout all simulations.

Analyses on data sets with low-average preservation rate or with incomplete taxon sampling distributed randomly across species showed that the marginal rates were not affected by any consistent bias (Fig. 4d,e and Supplementary Figs. S10–15 available at Dryad under http://dx.doi.org/10.5061/dryad.87d8s), even with data sets with as many as 30% of singletons and with taxon sampling below 25% (Table 2). In particular, we did not find evidence of over-/under-estimation of the rates or strong deviations of their dynamics. With increasing incompleteness of the data sets, however, the posterior rates showed wider HPDs and the magnitude of the rate changes often appeared smoothed out (Fig. 4d,e). While the mass extinction event simulated in the birth–death scenario VI was correctly detected under all sampling settings, a spurious decline of the speciation rate was observed under some conditions (Fig. 4 and Supplementary Figs. S10, S15 available at Dryad under http://dx.doi.org/10.5061/dryad.87d8s). This is however limited in its magnitude and did not receive significant support in the BDMCMC model sampling (Table 3 and Supplementary Table S2 available at Dryad under http://dx.doi.org/10.5061/dryad.87d8s).

Preservation rate.—Estimates of the mean preservation rate q were found to be very accurate throughout most of the simulations with credibility intervals centered on

TABLE 4. Posterior estimates of the preservation rates and accuracy of the root age estimate (expressed as difference from the true value, in myr)

	Simulation settings	Constant q model		Gamma model		
		$\hat{q}$	s_{max} (error)	$\hat{q}$	$\hat{\alpha}$	s_{max} (error)
PERT shape	l = 4	2.69–3.14	−2.39–1.42	2.62–3.14	13.87–20	−2.60–1.48
	l = 0	1.93–2.24	−10.39– −1.19	1.89–2.22	14.17–20	−10.83– −1.28
	l = 10	3.56–4.12	−0.33–3.27	3.45–4.17	12.68–20	−0.45–3.29
	l = 0–10	2.11–3.90	−5.24–3.03	2.06–3.94	13.17–20	−5.46–3.16
Rate heterogeneity	α = 1	2.69–4.68	−5.66–4.65	1.44–6.74	0.56–1.77	−9.00–3.69
	α = 3	2.60–3.63	−3.15–2.16	2.45–4.16	1.16–3.62	−4.02–2.05
	α = 5	2.70–3.46	−3.15–1.72	2.53–3.70	1.68–6.07	−3.33–1.81
	q = 3–0.05	0.65–1.69	−2.35–5.40	0.61–2.20	0.78–16.03	−5.88–5.56
Preservation rate	q = 1	0.82–1.11	−5.21–2.48	0.79–1.13	8.22–20	−6.27–2.62
	q = 0.5	0.40–0.61	−8.62–2.71	0.38–0.61	7.78–20	−10.66–2.82
	q = 0.25	0.19–0.35	−14.20–4.23	0.18–0.36	6.06–20	−16.70–4.72
Random taxon sampling	Ψ = 0.75	2.64–3.16	−3.19–5.95	2.56–3.16	12.48–20	−3.70–4.65
	Ψ = 0.50	2.63–3.19	−3.39–6.03	2.53–3.23	11.21–20	−3.74–6.41
	Ψ = 0.25	2.46–3.30	−3.62–8.29	2.26–3.36	6.86–20	−4.01–8.42

Notes: Results are provided from the constant rate and Gamma model. In the latter case estimates of the shape parameter (α) of the gamma distribution are also provided. Parameter values are provided as 95% HPD ranges calculated over all birth death scenarios and respective replicates. When the upper limit of this range equals 20 for the α parameter, the posterior distribution is truncated at the upper bound of the prior, indicating constant rates (cf. text).

the correct values and limited degrees of uncertainty, e.g., around ±10% of their true value (Table 4). The preservation rate was underestimated (around 30%) in the case of uniform fossil sampling (i.e., PERT(0)) and overestimated (by around 20%) under more convex distribution of the preservation rate (i.e., PERT(10)), whereas variable shapes of the PERT distribution yielded correct rate estimates. The posterior values of q obtained under the constant rate model and the Gamma model were similar even in the cases of data sets simulated under rate heterogeneity. However, the Gamma model was able to capture the correct degree of rate variation across species through the estimation of the shape parameter α. Indeed, in all data sets simulated under constant rate, the 95% HPD of α included its upper prior boundary (i.e., 20) indicating negligible variation. Instead, α was found to be significantly lower in the other cases, with the HPDs centered on their correct values (Table 4). Preservation rates increasing through time were also detected and resulted in values of $\alpha < 20$, although this led to a higher level of uncertainty around the shape parameter.

Root age.—The age of the birth–death process, measured as the oldest time of a speciation event in the clade (s_{max}), was found to be reconstructed with a high degree of accuracy in the large majority of simulations (Table 4). The correct root age was always found within the 95% HPD, and errors ranged between ± 2 and 5 myr when the preservation rate and taxon sampling were comparatively high i.e., $q \geq 1$, $\Psi \geq 0.75$. The only consistent bias was observed in the case of uniform fossil sampling (i.e., PERT(0)), in which case the root age was estimated between 1 and 10 myr too old. With lower preservation rates, more uncertainty was detected toward older root ages whereas an opposite trend was observed under very low random taxon sampling.

Diversification of Rhinocerotidae

The BDMCMC analyses of the Rhinocerotidae data indicated that speciation and extinction rates are likely to have varied throughout their evolutionary history. A hypothesis of constant rate was rejected for speciation (sampling frequencies of $k_\lambda = 1$ smaller than 0.01) and received low support for extinction (sampling frequencies of $k_\mu = 1$ smaller than 0.50; Table 5). Our analyses strongly support at least one shift in speciation rate at around 23 Ma, with potentially one additional rate shift earlier in the Oligocene. More uncertainty was found in the case of the extinction rates with over 50% of the BDMCMC samples scattered between one and three rate shifts. The marginal rates through time (Fig. 5a) displayed wide credibility intervals during the early stages of diversification of the family (at 40 Ma $\lambda =$ 0.344, 95% HPD: 0.135–0.671, $\mu = 0.177$, 95% HPD: 0.056–0.424). Between 35 and 25 Ma the extinction rate shows a nearly 60% decrease (from 0.177 to 0.102) mainly due to a reduction of the upper bound of the credibility interval (from 0.412 to 0.163). Similarly, the size of the credibility interval for speciation rate significantly decreased at around 35 Ma and speciation was found constant around $\lambda = 0.285$ (95% HPD: 0.139–0.423) until 23 Ma. A phase of positive net diversification is observed at the transition between the Oligocene and Miocene (29–21 Ma) with extinction significantly lower than speciation (at 26 Ma $\lambda - \mu = 0.184$; 95% HPD: 0.035–0.339; Fig. 5b). Around 20 Ma the speciation rate shows a 2.6-fold decrease

TABLE 5. Diversification of Rhinocerotidae: Estimated number of rates (k_λ and k_μ) obtained by BDMCMC (sampling frequencies) from the fossil record and by maximum likelihood (Akaike weights) from phylogenetic data

Data set	Parameter	Sampling frequency (number of rates k)					
		1	2	3	4	5	6
fossil (164 spp.)	λ	0.011	0.690	0.255	0.039	0.005	0
	μ	0.446	0.257	0.204	0.076	0.016	0.002
		Akaike weights (number of rates k)					
phylogeny (4 spp.)	λ & μ	≈ 1	≈ 0	0	0	0	0

(from 0.285 to 0.109) and afterwards remains constant until the present. Throughout most of the Miocene, the Rhinocerotidae diversification is characterized by a phase of stable species turnover, with both speciation and extinction rates being comparatively low. Starting from the late Miocene (around 10 Ma), a gradual increase of the marginal extinction rates is inferred, culminating at the Pleistocene, with $\mu = 0.177$ (95% HPD: 0.062–0.440). This pattern paired by constant speciation rates results in negative mean diversification rates over the past 11 myr, i.e., in the Pleistocene and Holocene $\lambda - \mu = -0.067$ (95% HPD: -0.344– 0.077).

Our analyses provide evidence for a very strong heterogeneity of the preservation rate across species with an estimated shape parameter $\alpha = 0.576$ (95% HPD: 0.490–0.667) and the mean likelihood being over 800 log units higher for the Gamma model compared to the constant rate model. The average preservation rate was found to be equal to $q = 5.701$ fossil occurrences per species per myr (95% HPD: 5.197–6.206), but effectively varying across species, based on the estimated Gamma model, from ca. 0.25 to 15 estimated fossils per species per myr. The root age of the Rhinocerotidae clade was inferred at 47.33 Ma (95% HPD: 45.24–50.39 Ma; Fig. 5c).

Phylogeny-based analyses on the dated molecular tree of extant taxa strongly favored a constant rate birth–death against models with one or more rate shifts based on AICc scores and Akaike weights (Table 5). Under the constant rate model, the estimated speciation rate was 0.026 and the extinction rate was found to be approximately 0 (Fig. 5b).

DISCUSSION

We proposed a novel Bayesian framework to model the sampling process and the dynamics of species diversification and extinction. Our method uses all available temporal occurrences assigned to a species, including singletons and extant species with at least one fossil occurrence, which are usually excluded from fossil analyses, to infer the parameters of a birth–death process while jointly estimating the times of speciation and extinction. The joint parameter estimation, in a hierarchical Bayesian context, provides a natural way of incorporating the temporal uncertainties of speciation and extinction events imposed by the incompleteness of the fossil record, and reduces the risk of error propagation (Lartillot and Poujol 2011). Furthermore the use of mechanistic, process-based priors such as the birth–death (hyper-)priors applied here (Equations 9 and 12; Keiding 1975; Kubo and Iwasa 1995) renders their definition less subjective and easier to interpret.

One advantage of our Bayesian method over previously developed likelihood or nonprobabilistic (fossil-based) approaches is the parameterization of the model. New parameters, i.e., rate shifts, are introduced in the model only if they significantly improve the likelihood of the data. Moreover, the BDMCMC approach has the ability to explore a virtually infinite set of models that would not be feasible to test using marginal likelihoods. This effectively corresponds to model averaging and makes the estimates more robust, particularly if several models have a similar fit (Wasserman 2000), because the posterior marginal rates through time incorporate the uncertainty of both the time of rate shifts and their number.

Birth–death Process

Our analyses showed that the dynamics of speciation and extinction rates, including sudden rate changes and mass extinction, are correctly estimated under a wide range of conditions, such as low levels of preservation (down to 1–3 fossil occurrences per species on average), severely incomplete taxon sampling (up to 80% missing), and high proportion of singletons (exceeding 30% of the taxa in some cases). Speciation and extinction rates can appear to vary in a continuous fashion (e.g., Fig. 4c,d) despite the discrete nature of the birth–death model with shifts reflecting incompleteness of the data and model and parameter uncertainty (Stadler et al. 2013). It should be noted, however, that when the incompleteness of the data reduced the power of the algorithm to find the correct model of diversification, the analyses tended to be conservative and yielded near-uniform rates with wide credibility intervals, rather than spurious rate shifts.

Our approach allows flexibility in model definition and parameterization. For instance, we do not constrain

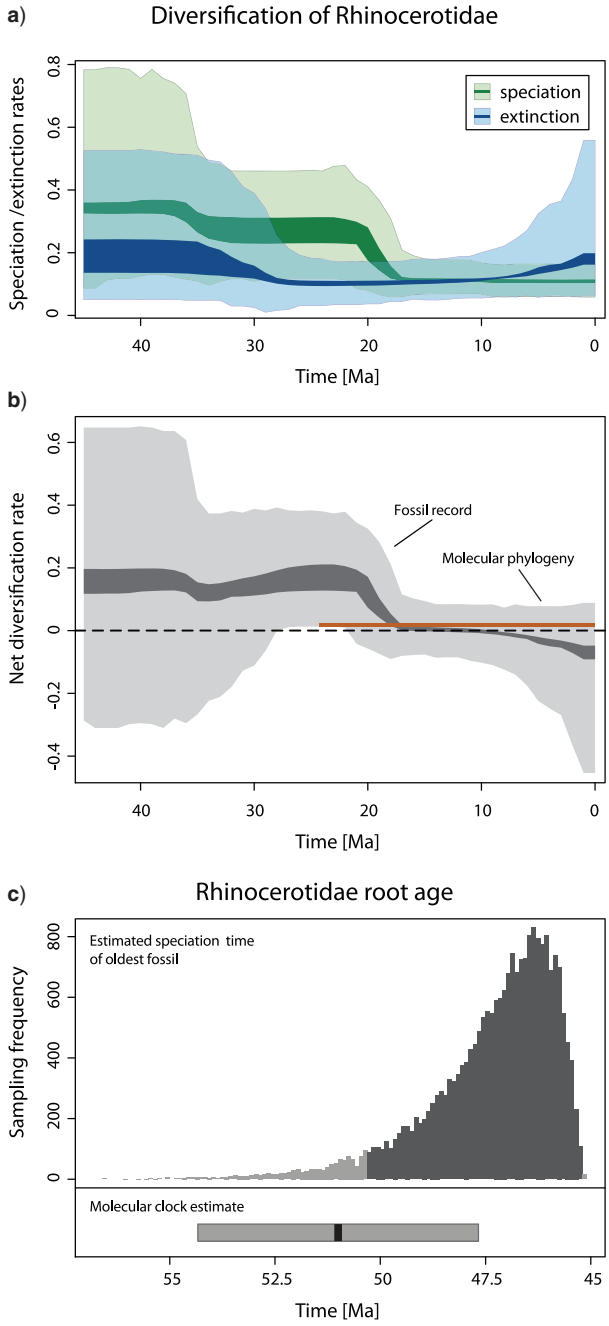

FIGURE 5. Diversification of Rhinocerotidae. a) Marginal rates of speciation (green) and extinction (blue) through time for the Rhinocerotidae. Light shaded areas show the 95% HPD around the mean rate estimates. b) Net diversification rates through time based on fossil data (black line) using BDMCMC, and from a molecular phylogenetic tree of the extant taxa (orange line) using TreePar. c) Posterior estimates of the root age (s_{max} for the fossil data) and stem age of the clade based on Steiner and Ryder (2011).

the net diversification rate to be positive, unlike in several phylogeny-based methods (e.g., Rabosky 2006; Alfaro et al. 2009; Didier et al. 2012) and we do not assume that the temporal dynamics of speciation and extinction are linked to each other. This flexibility results in a high power to distinguish among different diversification scenarios, as highlighted in the simulated data sets IV and V. These were based on slightly different extinction dynamics that resulted in virtually identical patterns of species accumulation through time (Fig. 2d,e). Nevertheless, the BDMCMC provided strong positive support for the correct model in both cases (Table 3). Given that speciation and extinction and their temporal dynamics may be driven by different biotic and abiotic factors (Ezard et al. 2011; Vamosi J.C. and Vamosi S.M. 2011; Bapst et al. 2012; Peters et al. 2013; Quental and Marshall 2013; Silvestro et al. 2013), it is important to unlink these processes.

Arguably the strongest bias in birth–death rate estimates is caused by incomplete data (in particular in the case of molecular phylogenies), unless appropriately corrected (Yang and Rannala 1997; Stadler 2009; Morlon et al. 2011; Cusimano et al. 2012). The reconstructed evolutionary process is biased by the incompleteness of the data because missing lineages alter the distribution of nodes in a dated tree (Ricklefs 2007). The incompleteness of the data in the fossil record appears to have a less problematic effect on the estimation of speciation and extinction rates because in contrast to molecular phylogenies, removing a random set of taxa does not affect the observed occurrences of other lineages. Both the likelihood surface plots (Supplementary Fig. S1 available at Dryad under http://dx.doi.org/10.5061/dryad.87d8s) and our simulations confirm the absence of consistent biases in the marginal rates beyond an increase of parameter uncertainty as the taxon sampling decreases, which can be attributed to the smaller size of the data set (Liow, Quental, et al. 2010). Furthermore, the use of the standing diversity of a clade (N_{OBS}) in constructing the birth–death hyperprior provides additional information by implicitly correcting for biases due to a possible under-sampling of extant species that did not leave any fossil records yet. Finally, as the formulation of (Kubo and Iwasa 1995) was originally derived for application to phylogenies of extant taxa, our model provides a direct link between fossil and phylogeny-based macroevolutionary analyses.

The time of origin of a clade (s_{max}) was estimated with comparatively high accuracy throughout the whole set of simulations, even when the age of the oldest simulated fossil occurrence was several million years younger than the true age of the data set. Estimating the correct age of origin of a clade is crucial for an accurate dating of extant taxa phylogenies and combining analyses of fossils and molecular data may constitute the best solution for calibrating a tree (Ronquist, Klopfstein, et al. 2012; Wilkinson et al. 2011; Steiper and Seiffert 2012; Heled and Drummond 2012; Nowak et al. 2013). The posterior distribution of the time of origin of a clade obtained by our method (e.g., Fig. 5c) might therefore provide a valid alternative to often more arbitrary (log-normal or exponential) distributions commonly applied as calibration constraints in molecular dating.

Fossils, the preservation process, and the geological record

Heterogeneity in the temporal sampling of individual fossil species could seriously compromise the estimation of times of speciation and extinction. Specifically, sampled fossil occurrences of individual species are often "hat-shaped" in empirical data sets (Foote et al. 2007; Liow and Stenseth 2007; Liow, Skaug, et al. 2010). To account for this, the fossil occurrences were modeled as the results of a stochastic process (the NHPP) with a generalized symmetric beta distribution (the PERT distribution) describing the varying probability of preservation of a species during its lifespan. We used a fixed shape parameter ($l=4$) because it mimics other distributions previously used to described such trajectories (Liow, Quental, et al. 2010; Pigot et al. 2012) and because of mathematical convenience (Equations 2 and 3). The accuracy of the rate estimates is affected by strong deviations from this distribution, but only when these are consistent throughout the entire data set (for instance in the case of uniform sampling; l=0, Supplementary Fig. S3 available at Dryad under http://dx.doi.org/10.5061/dryad.87d8s). This bias becomes almost negligible under the more realistic scenario of variable shapes of the preservation process (Fig. 4a and Supplementary Fig. S5 available at Dryad under http://dx.doi.org/10.5061/dryad.87d8s). An estimation of the shape and skewness of the preservation rates from empirical data might be difficult due to the limited number of available fossil occurrences per species. However, a relatively dense and continuous fossil record (e.g., in marine plankton) allows the estimation of these trajectories (Liow, Skaug, et al. 2010), and other, potentially asymmetric, curves can be easily incorporated in our model to reproduce different preservation trajectories.

Our probabilistic method can explicitly account for different fossilization rates among taxa using the Gamma model, which represents an equivalent of Yang's discrete model for molecular rate heterogeneity (Yang 1994). While the Gamma model does not infer species-specific preservation rates, it accounts efficiently for rate variation across taxa as demonstrated by the accurate estimation of the shape parameter of the gamma distribution. This is in contrast to many existing studies in which taxa within any given analysis are implicitly or explicitly assumed to have the same preservation potential (Foote 2000, 2003; Liow et al. 2008; Alroy et al. 2008), even when preservation rates are allowed to change over time. The lineage-specific preservation rate introduced in our Gamma model accounts to some extent for temporal changes of the preservation rates, as indicated by the comparatively little bias observed from data sets with strongly decreasing preservation rates (Fig. 4c). Indeed, temporal variation of the preservation rate partly translates into across-lineage variation, in that e.g., older lineages have different average rates from more recent ones. Nevertheless, the explicit parameterization of different mean preservation rates through time might further improve the accuracy of the analyses (Foote 2001; Wagner et al. 2013).

The geological record associated with each fossil is usually referenced to rock or sedimentary units associated with estimated age bins with varying degrees of uncertainty (Marshall 1990; Gradstein et al. 2012). Fossil finds are usually grouped into either time bins based on geological units (Foote 2003; Alroy 2010) or equal duration time bins (Liow et al. 2008) for diversification analyses (but see Ezard et al. 2011). In modeling continuous time rather than interval-to-interval changes in rates, we are able to reflect the continuous nature of the biological processes of speciation and extinction. Furthermore, we can incorporate in our analyses the uncertainty around the temporal placement of the fossil occurrences by randomizing the ages within their min and max ranges (i.e., the respective discrete bins) and running the analyses multiple times.

Rhinocerotidae Diversification

The application of our Bayesian approach was demonstrated on a data set of Rhinocerotidae, revealing that their diversification was shaped by several changes in speciation and extinction rates with a phase of species accumulation (45–23 Ma), followed by species turnover (23–10 Ma) and eventually net negative diversification since the end of the Miocene (Fig. 5). The greater width of the credibility intervals during the early stage of diversification shows a similarity with the patterns obtained by simulations under a nonuniform preservation rate. This suggests that the rate of preservation in Rhinocerotidae may be increasing toward the present. Thus, the rate heterogeneity detected by the Gamma model is likely to capture a strong component of temporal rate variation, since modern species tend to show a much richer fossil record than older ones (as also observed in simulated data; Table 4). The negative rates of diversification that are driven by an increase in extinction rate toward the present have caused a 10-fold decrease of species richness of the family over the past 10 myr.

Diversification rate analyses of the Rhinocerotidae family highlight the shortcomings of limiting the data to extant species. A molecular phylogeny of the five extant taxa can provide at most five data points (the branch lengths) to inform a birth–death model about the speciation and extinction dynamics of the family in over 50 myr. Thus, as expected, phylogeny-based analyses could not detect any rate variation. On the contrary, our analysis of the over 160 species sampled in the fossil record depicted a complex history of expansion and decline of the family resulting from repeated and temporally unlinked shifts of speciation and extinction rates. On this basis, even if extant taxa phylogenies of larger clades can be used to infer complex diversification processes (e.g., Morlon et al. 2011; Stadler 2011), the analyses of the respective fossil record, when available, should provide an important additional source of information to reconstruct the dynamics of speciation and extinction rates.

Conclusion and Outlook

We have introduced with this study a novel approach to infer diversification dynamics based on the fossil record. Our method can be applied to a wide range of data sets at various taxonomic ranks and temporal scales, particularly considering the increasing amount of data that are available in public databases (e.g., the Paleobiology Database, NOW). The approach also offers the opportunity to investigate the diversification dynamics of formerly species-rich clades that comprise today a small extent of their past diversity. This may allow evolutionary biologists to overcome the bias towards studying predominantly large clades, while often neglecting the less diverse ones (Ricklefs 2007).

Finally, rates of speciation and extinction are estimated from fossil data using the same underlying birth–death stochastic process usually applied in molecular phylogenetics. This makes the estimated rates comparable among data sets and represents an important step towards integrating phylogenies and fossils. For instance, the posterior distributions of the speciation times can be potentially applied as prior distributions for calibration in molecular dating. Moreover, the birth–death process estimated from the analysis of the fossil record should constitute a new basis for a more informative prior to the branching times in molecular clock analyses. In future developments, molecular and fossil data should be analyzed together (Ronquist, Klopfstein, et al. 2012; Slater et al. 2013; Fritz et al. 2013) to jointly estimate speciation and extinction times of fossil species, branching times of the phylogeny, and a common underlying birth–death process.

SUPPLEMENTARY MATERIAL

Data available from the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.87d8s

FUNDING

D.S. received funding from the German Academic Exchange Service (DAAD), the University of Lausanne, and the Wenner-Gren Foundation (Sweden). J.S. was supported by the funding program LOEWE—"Landes-Offensive zur Entwicklung Wissenschaftlich-ökonomischer Exzellenz" of Hesse's Ministry of Higher Education, Research, and the Arts. A.A. is supported by grants from the Swedish Research Council (B0569601) and the European Research Council under the European Unions Seventh Framework Programme (FP/2007-2013, ERC Grant Agreement n. 331024). N.S. is supported by the grants PDFMP3_134931 and 3100A0_138282 from the Swiss National Science Foundation.

ACKNOWLEDGMENTS

Tanja Stadler, Michael Foote, Luke Harmon, and two anonymous reviewers provided valuable comments on earlier versions of the manuscript. We thank Cynthia C. Steiner for providing the phylogeny of Rhinocerotidae, Bob O'Hara, Linda Dib, Anna Kostikova, and Nicolas Lartillot for discussion. All simulations and analyses were run at the high-performance computing center Vital-IT of the Swiss Institute of Bioinformatics (Lausanne, Switzerland), and the High-Performance Computing Cluster operated by the BiK-F Data and Modelling Centre (Frankfurt, Germany). We thank Claus Weiland for technical support.

APPENDIX—MCMC IMPLEMENTATION

MCMC and model testing using marginal likelihoods.—In the MCMC the Hasting ratio was maintained constant by symmetric updates of the model parameters within uniform windows centered on their current values, while reflection at the boundary ensures that the parameter values lie within the allowed ranges (e.g., $\lambda, \mu \geq 0$). To improve the chain mixing and reduce the length of the burn-in phase, a Metropolis-coupled MCMC algorithm (Geyer 1991; Altekar et al. 2004) was implemented by pairing the "cold" chain with incrementally heated chains that can move more easily through the parameter space.

The fit of different birth–death models can be measured by the respective marginal likelihood using the thermodynamic integration (TI) originally described by Lartillot and Philippe (2006). The approach uses MCMC to sample a progression of distributions ranging from the posterior to the prior, obtained by raising the likelihood ratio to the power of a parameter $\beta \in [0,1]$. A path of 10 β values is generated from a beta distribution with shape $B(0.3,1)$ to place more values close to 0, where the acceptance probability changes more rapidly (Xie et al. 2011). The marginal likelihoods L_{TI} of a model is obtained by integrating the likelihood along the path of power values $\beta_1, \ldots, \beta_Z$:

$$\log L_{TI} = \int_0^1 \log L d\beta \sim \sum_{z=2}^{Z} (\beta_z - \beta_{z-1}) \frac{(\log L_z + \log L_{z-1})}{2} \tag{17}$$

where $\log L_z$ is the mean log-likelihood obtained under a power value β_z (Baele et al. 2012). Because the model testing is focused on the number of rate shifts, only the birth–death term of Equation (16) is marginalized during the process to improve the performance of the MCMC sampling and reduce the computational cost (cf. Lepage et al. 2007). To evaluate the relative support of a model M_1 against an alternative M_0, log Bayes factors can be calculated as $BF = 2(\log L_{TI}(M_1) - \log L_{TI}(M_0))$ according to Kass and Raftery (1995).

Joint model testing and parameter estimation via BDMCMC.—An alternative algorithm was developed (and used for all the analyses presented in this study) to jointly estimate the number of parameters of the model and their values. We implemented the birth–death

MCMC (BDMCMC; Stephens 2000) to estimate the number of rate shifts in speciation and extinction rates. The vectors of speciation and extinction times (**s** and **e**) were assumed to derive from an underlying model of diversification with a number of speciation rates ($k_\lambda \geq 1$) and extinction rates ($k_\mu \geq 1$). The number of rate shifts ($k_\lambda - 1; k_\mu - 1$) was considered as an unknown variable and estimated from the data.

The diversification model is treated as a point process. Each *component* of the model, indicated as (π, ϕ), is a point in the parameter space with mixture proportion $\pi \in [0,1]$ and parameter ϕ (i.e., speciation or extinction rates). The mixture proportion of one component is given by the length of a time frame, e.g., between rate shifts, expressed as a fraction of the total duration of the data set ($s_{\max}$)

$$\pi_i = \frac{(\tau_{i-1}^{\phi} - \tau_i^{\phi})}{s_{\max}} \tag{18}$$

with the associated diversification parameter ϕ_i being either λ_i or μ_i. Thus given a number k of rates, the diversification model is composed by the components $(\pi_1, \phi_1), \ldots, (\pi_k, \phi_k)$ with the constraint that $\pi_1 + \ldots + \pi_k = 1$. Vectors of speciation and extinction rates (Λ, and M; cf. Equation 11) and the respective time frames defined by τ^{Λ} and τ^{M} are considered as two independent families of components.

With the BDMCMC algorithm the parameter space is explored while moving across models and sampling values of k_ϕ from the joint posterior distribution of k_ϕ, π, ϕ. This approach allows for "jumps" between models with different dimensions by two types of moves: *births* and *deaths* that add or remove one component to the model, respectively. The birth event changes the model $y = \{(\pi_1, \phi_1), \ldots (\pi_k, \phi_k)\}$ by adding at a random position one component $y \cup (\pi_r, \phi_r)$ and rescaling all mixture proportions from π_i to $\pi_i(1 - \pi_r)$. Similarly, a death event causes the deletion of one component (π_j, ϕ_j) so that in the resulting model, indicated by $y \setminus (\pi_j, \phi_j)$, all remaining components have rescaled proportions from π_i to $\pi_i/(1 - \pi_j)$. Thus, changing the dimension of the model (k) does not affect the constraint $\pi_1 + \ldots + \pi_k = 1$. The BDMCMC approach assumes that births and deaths occur as independent Poisson processes following a continuous time Markov birth–death process. Unlike analogous algorithms, e.g., the reversible-jump MCMC (Green 1995), the BDMCMC moves across models are not determined by an acceptance probability, but by varying rates of the birth–death process. This simplifies the implementation by avoiding the need for calculation of the Jacobian (Stephens 2000). The birth rate is set to a constant $\beta(y) = \lambda_b$ whereas death rates vary for each component of the model (π_j, ϕ_j) and are calculated as:

$$\delta_j(y) = \lambda_b \frac{L\left(y \setminus (\pi_j, \phi_j)\right)}{L(y)} \frac{p(k-1)}{kp(k)} \quad (j = 1, \ldots k) \tag{19}$$

where $L(\cdot)$ is the likelihood of the data, in this case (**s**, **e**), for a given birth–death model (Equation 11), and $p(k)$ is the prior probability of a model with k components. The continuous time birth–death process is used to jump between models works as follows:

1. Initialize the model with $k_\lambda = 1$ and $k_\mu = 1$ (i.e., no rate shifts), and parameter values sampled from a prior distribution.
2. Calculate death rates for each component of the model (the vectors Λ and M) based on the (constant) birth rate $\beta(y)$ and using Equation (19).
3. Calculate the total death rate from the rates assigned to each component $\delta(y) = \sum_{j=1}^{k} \delta_j(y)$.
4. Simulate the time for the next jump based on the continuous time Markov birth–death process, i.e., from an exponential distribution with mean $1/(\beta(y) + \delta(y))$. The jump will be either a birth or a death with probabilities
$$p(\text{birth}) = \frac{\beta(y)}{\beta(y) + \delta(y)}, \quad p(\text{death}) = \frac{\delta(y)}{\beta(y) + \delta(y)}. \tag{20}$$
When a birth is selected, a new component (π_r, ϕ_r) is generated by randomly sampling π from a symmetric Dirichlet distribution with $k+1$ dimensions (i.e., the prior applied to the length of the time frames) and ϕ from the gamma distribution set as prior to the rates. The new component is added to the vectors Λ or M with equal probability. When a death occurs, one component $(\pi_j, \phi_j) \in y$ is selected to be deleted with probability $\delta_j(y)/\delta(y)$ for $j = 1, \ldots k$. After a birth or death of one component, all (remaining) mixture proportions are rescaled as described above.
5. Return to step 2 until a fixed time t_0 is reached.

The BDMCMC algorithm was incorporated in the Bayesian framework described in the previous paragraphs so that the numbers of rate shifts (k_λ and k_μ), that define the components of the model, are sampled as an unknown variables together with all the other parameters (e.g., $\mathbf{s}, \mathbf{e}, \Lambda, M$). In our analyses, we set the frequency of model updates to 0.04 (i.e., on average every 25th MCMC generation) and the length of the continuous time birth–death process to $t_0 = 1$, but higher frequencies and a longer duration of t_0 were tested to improve the mixing of the parameters k_λ and k_μ. The birth rate at which new components are added was set to $\lambda_b = 1$. The prior on the number of components p_k was set to a Poisson distribution with shape parameter equal to the birth rate (λ_b), so that the calculation of the death rates (Equation 19) conveniently reduces to a likelihood ratio (Stephens 2000).

It is noted that a BDMCMC analysis requires about 10% of the CPU time that would be necessary to perform model testing through thermodynamic integration and subsequent MCMC for parameter estimation. This highlights the efficiency of jointly exploring the

parameter space and its dimensions (Green 1995; Stephens 2000).

References

Akaike H. 1973. Information theory and an extension of the maximum likelihood principle. Akademiai Kiado, Budapest.

Alfaro M.E., Santini F., Brock C., Alamillo H., Dornburg A., Rabosky D.L., Carnevale G., Harmon L.J. 2009. Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. Proc. Natl Acad. Sci. USA 106:13410–13414.

Alroy J. 2010. The shifting balance of diversity among major marine animal groups. Science 329:1191–1194.

Alroy J., Aberhan M., Bottjer D.J., Foote M., Fürsich F.T., Harries P.J., Hendy A.J.W., Holland S.M., Ivany L.C., Kiessling W., Kosnik M.A., Marshall C.R., McGowan A.J., Miller A.I., Olszewski T.D., Patzkowsky M.E., Peters S.E., Villier L., Wagner P.J., Bonuso N., Borkow P.S., Brenneis B., Clapham M.E., Fall L.M., Ferguson C.A., Hanson V.L., Krug A.Z., Layou K.M., Leckey E.H., Nürnberg S., Powers C.M., Sessa J.A., Simpson C., Tomašových A., Visaggi C.C. 2008. Phanerozoic trends in the global diversity of marine invertebrates. Science 321:97–100.

Altekar G., Dwarkadas S., Huelsenbeck J., Ronquist F. 2004. Parallel Metropolis coupled Markov chain Monte Carlo for Bayesian phylogenetic inference. Bioinformatics 20:407–415.

Antonelli A., Sanmartín I. 2011. Mass extinction, gradual cooling, or rapid radiation? Reconstructing the spatiotemporal evolution of the ancient angiosperm genus *Hedyosmum* (Chloranthaceae) using empirical and simulated approaches. Syst. Biol. 60:596–615.

Baele G., Lemey P., Bedford T., Rambaut A., Suchard M., Alekseyenko A. 2012. Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. Mol. Biol. Evol. 29:2157–2167.

Bapst D.W., Bullock P.C., Melchin M.J., Sheets H.D., Mitchell C.E. 2012. Graptoloid diversity and disparity became decoupled during the Ordovician mass extinction. Proc. Natl Acad. Sci. USA 109:3428–3433.

Bokma F. 2003. Testing for equal rates of cladogenesis in diverse taxa. Evolution 57:2469–2474.

Bokma F. 2008. Bayesian estimation of speciation and extinction probabilities from (in)complete phylogenies. Evolution 62:2441–2445.

Burnham K., Anderson D. 2002. Model selection and multimodel inference: a practical information-theoretic approach. 2nd ed. New York: Springer.

Cermeno P. 2012. Marine planktonic microbes survived climatic instabilities in the past. Proc. R. Soc. Lond. B 279:474–479.

Connolly S.R., Miller A.I. 2001. Joint estimation of sampling and turnover rates from fossil databases: Capture-mark-recapture methods revisited. Paleobiology 27:751–767.

Cusimano N., Stadler T., Renner S. 2012. A new method for handling missing species in diversification analysis applicable to randomly or nonrandomly sampled phylogenies. Syst. Biol. 61:785–792.

Didier G., Royer-Carenzi M., Laurin M. 2012. The reconstructed evolutionary process with the fossil record. J. Theor. Biol. 315:26–37.

Drummond A.J., Ho S., Phillips M., Rambaut A. 2006. Relaxed phylogenetics and dating with confidence. Plos Biol. 4:e88.

Drummond, C.S., Eastwood R.J., Miotto S.T., Hughes C.E. 2012. Multiple continental radiations and correlates of diversification in *Lupinus* (Leguminosae): Testing for key innovation with incomplete taxon sampling. Syst. Biol. 61:443–460.

Etienne R.S., Haegeman B., Stadler T., Aze T., Pearson P.N., Purvis A., Phillimore A.B. 2012. Diversity-dependence brings molecular phylogenies closer to agreement with the fossil record. Proc. R. Soc. Lond. B 279:1300–1309.

Ezard T.H.G., Aze T., Pearson P.N., Purvis A. 2011. Interplay between changing climate and species' ecology drives macroevolutionary dynamics. Science 332:349–351.

FitzJohn R.G., Maddison W.P., Otto S.P. 2009. Estimating trait-dependent speciation and extinction rates from incompletely resolved phylogenies. Syst. Biol. 58:595–611.

Foote M. 2000. Origination and extinction components of taxonomic diversity: General problems. Paleobiology 26:74–102.

Foote M. 2001. Inferring temporal patterns of preservation, origination, and extinction from taxonomic survivorship analysis. Paleobiology 27:602–630.

Foote M. 2003. Origination and extinction through the Phanerozoic: A new approach. J. Geol. 111:125–148.

Foote M., Crampton J.S., Beu A.G., Marshall B.A., Cooper R.A., Maxwell P.A., Matcham I. 2007. Rise and fall of species occupancy in Cenozoic fossil mollusks. Science 318:1131–1134.

Foote M., Miller A.I. 2007. Principles of Paleontology. New York: W.H. Freeman and Co.

Foote M., Raup D.M. 1996. Fossil preservation and the stratigraphic ranges of taxa. Paleobiology 22:121–140.

Fortelius M. 2013. New and Old Worlds Database of Fossil Mammals (NOW). University of Helsinki. http://www.helsinki.fi/science/now/.

Fritz S., Schnitzler J., Eronen J.T., Hof C., Bhning-Gaese K., Graham C. 2013. Diversity in time and space: wanted dead and alive. Trends Ecol. Evol. 28:509–516.

Geraads D. 2010. Rhinocerotidae. In: Werdelin L, Sanders W.J. editors. Cenozoic Mammals of Africa. University of California Press p. 669–683.

Geyer C.J. 1991. Computing Science and Statistics. Proceedings of the 23rd Symposium on the Interface. chap. Markov chain Monte Carlo maximum likelihood. Keramidas editor, Fairfax Station: Interface Foundation.

Gradstein F.M., Ogg J.G., Schmitz M., Ogg G., eds. 2012. The geologic time scale. Elsevier.

Green P.J. 1995. Reversible jump Markov chain Monte Carlo and Bayesian model determination. Biometrika 82:711–732.

Harvey P.H., May R.M., Nee S. 1994. Phylogenies without fossils. Evolution 48:523–529.

Heled J., Drummond A.J. 2012. Calibrated tree priors for relaxed phylogenetics and divergence time estimation. Syst. Biol. 61:138–149.

Hunt G. 2013. Testing the link between phenotypic evolution and speciation: an integrated palaeontological and phylogenetic analysis. Methods Ecol. Evol. 4:714–723.

Kass R.E., Raftery A.E. 1995. Bayes factors. J. Amer. Stat. Assoc. 90:773–795.

Keiding N. 1975. Maximum likelihood estimation in the birth-death process. Ann. Stat. 3:363–372.

Kendall D.G. 1948. On the generalized birth-and-death process. Ann. Math. Stat. 19:1–15.

Kubo T., Iwasa Y. 1995. Inferring the rates of branching and extinction from molecular phylogenies. Evolution 49:694–704.

Kurtén B. 1954. Population dynamics: a new method in paleontology. J. Paleontol. 28:286–292.

Lartillot N., Philippe H. 2006. Computing Bayes factors using thermodynamic integration. Syst. Biol. 55:195–207.

Lartillot N., Poujol R. 2011. A phylogenetic model for investigating correlated evolution of substitution rates and continuous phenotypic characters. Mol. Biol. Evol. 28:729–744.

Lepage T., Bryant D., Philippe H., Lartillot N. 2007. A general comparison of relaxed molecular clock models. Mol. Biol. Evol. 24:2669–2680.

Liow L., Fortelius M., Bingham E., Lintulaakso K., Mannila H., Flynn L., Stenseth N.C. 2008. Higher origination and extinction rates in larger mammals. Proc. Natl. Acad. Sci. USA 105:6097–6102.

Liow L., Quental T., Marshall C. 2010. When can decreasing diversification rates be detected with molecular phylogenies and the fossil record? Syst. Biol. 59:646–59.

Liow L., Skaug H., Ergon T., Schweder T. 2010. Global occurrence trajectories of microfossils: environmental volatility and the rise and fall of individual species. Paleobiology 36:224–252.

Liow L.H., Stenseth N.C. 2007. The rise and fall of species: implications for macroevolutionary and macroecological studies. Proc. R. Soc. Lond. B 274:2745–2752.

Litsios G., Sims C.A., Wüest R.O., Pearman P.B., Zimmermann N.E., N. Salamin. 2012. Mutualism with sea anemones triggered the adaptive radiation of clownfishes. BMC Evol. Biol. 12:212.

Lloyd G.T., Young J.R., Smith A.B. 2012. Taxonomic structure of the fossil record is shaped by sampling bias. Syst. Biol. 61:80–89.
Maddison W., Midford P., Otto S. 2007. Estimating a binary character's effect on speciation and extinction. Syst. Biol. 56:701–710.
Marshall C.R. 1990. Confidence-intervals on stratigraphic ranges. Paleobiology 16:1–10.
Morlon H., Parsons T.L., Plotkin J.B. 2011. Reconciling molecular phylogenies with the fossil record. Proc. Natl. Acad. Sci. USA 108:16327–16332.
Nee S. 2001. Inferring speciation rates from phylogenies. Evolution 55:661–668.
Nee S. 2006. Birth-death models in macroevolution. Annu. Rev. Ecol. Evol. Syst. 37:1–17.
Nee S., May R.M., Harvey P. 1994. The reconstructed evolutionary process. Phil. Trans. R Soc. B 344:305–311.
Nichols J.D., Pollock K.H.. 1983. Estimating taxonomic diversity, extinction rates, and speciation rates from fossil data using capture-recapture models. Paleobiology 9:150–163.
Nowak M., Smith A., Simpson C., Zwickl D. 2013. A simple method for estimating informative node age priors for the fossil calibration of molecular divergence time analyses. PLoS ONE 8:e66245.
Paradis E. 2004. Can extinction rates be estimated without fossils? J. Theor. Biol. 229:19–30.
Peters S.E., Kelly D.C., Fraass A.J. 2013. Oceanographic controls on the diversity and extinction of planktonic foraminifera. Nature 493:398–401.
Pigot A.L., Owens I.P.F., Orme C.D.L. 2012. Speciation and extinction drive the appearance of directional range size evolution in phylogenies and the fossil record. PLoS Biol. 10:e1001260.
Quental T., Marshall C.R. 2010. Diversity dynamics: Molecular phylogenies need the fossil record. Trends Ecol. Evol. 25: 434–441.
Quental T.B., Marshall C.R. 2013. How the red queen drives terrestrial mammals to extinction. Science 341:290–292.
Rabosky D. 2006. Likelihood methods for detecting temporal shifts in diversification rates. Evolution 60:1152–1164.
Rabosky D.L. 2010. Extinction rates should not be estimated from molecular phylogenies. Evolution 64:1816–1824.
Rabosky D.L., Slater G.J., Alfaro M.E., Mace G.M. 2012. Clade age and species richness are decoupled across the eukaryotic tree of life. PLoS Biol. 10:e1001381.
Rambaut A., Drummond A. 2007. Tracer: Available from http://beast.bio.ed.ac.uk/tracer.
Raup D.M. 1975. Taxonomic survivorship curves and Van Valen's law. Paleobiology 1:82–96.
Raup D.M. 1991. A kill curve for Phanerozoic marine species. Paleobiology 17:37–48.
Ricklefs R.E. 2007. Estimating diversification rates from phylogenetic information. Trends Ecol. Evol. 22:601–610.
Ronquist F., Klopfstein S., Vilhelmsen L., Schulmeister S., Murray D.L., Rasnitsyn A.P. 2012. A total-evidence approach to dating with fossils, applied to the early radiation of the Hymenoptera. Syst. Biol. 61:973–999.
Ronquist F., Teslenko M., Van Der Mark P., Ayres D.L., Darling A., Hhna S., Larget B., Liu L., Suchard M.A., Huelsenbeck J. 2012. Mrbayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. 61:539–542.
Ryberg M., Nilsson R.H., Matheny P.B. 2011. DivBayes and SubT: exploring species diversification using bayesian statistics. Bioinformatics 27:2439–2440.
Sanderson M.J., Donoghue M.J. 1996. Reconstructing shifts in diversification rates on phylogenetic trees. Trends Ecol. Evol. 11:15–20.
Sepkoski J.J. 1981. A factor analytic description of the Phanerozoic marine fossil record. Paleobiology 7:36–53.
Sepkoski J.J. 1998. Rates of speciation in the fossil record. Phil. Trans. R Soc. B 353:315–326.
Silvestro D., Schnitzler J., Zizka G. 2011. A Bayesian framework to estimate diversification rates and their variation through time and space. BMC Evol. Biol. 11:311.
Silvestro D., Schulte K., Zizka G. 2013. Disentangling the effects of key innovations on the diversification of Bromelioideae (Bromeliaceae). Evolution Advance Access Published September 11, 2013; DOI:10.1111/evo.12236.
Simpson G.G. 1944. Tempo and mode in Evolution. New York: Columbia University Press.
Slater G., Harmon L.J., Freckleton R. 2013. Unifying fossils and phylogenies for comparative analyses of diversification and trait evolution. Methods Ecol. Evol. 4:699–702.
Stadler T. 2009. On incomplete sampling under birth-death models and connections to the sampling-based coalescent. J. Theor. Biol. 261:58–66.
Stadler T. 2011. Mammalian phylogeny reveals recent diversification rate shifts. Proc. Natl Acad. Sci. USA 108:6187–6192.
Stadler T., Bokma F. 2013. Estimating speciation and extinction rates for phylogenies of higher taxa. Syst. Biol. 62:220–230.
Stadler T., Kuhnert D., Bonhoeffer S., Drummond A. 2013. Birth-death skyline plot reveals temporal changes of epidemic spread in hiv and hepatitis c virus (hcv). Proc. Natl Acad. Sci. USA 110:228–233.
Stanley S.M. 1979. Macroevolution: Pattern and Process. Baltimore and London: The John Hopkins University Press.
Steiner C.C., Ryder O.A. 2011. Molecular phylogeny and evolution of the Perissodactyla. Zool. J. Linn. Soc. 163:1289–1303.
Steiper M.E., Seiffert E.R. 2012. Evidence for a convergent slowdown in primate molecular rates and its implications for the timing of early primate evolution. Proc. Natl Acad. Sci. USA 109:6006–6011.
Stephens M. 2000. Bayesian analysis of mixture models with an unknown number of components—an alternative to reversible jump methods. Ann. Stat. 28:40–74.
Strauss D., Sadler P.M. 1989. Classical confidence-intervals and Bayesian probability estimates for ends of local taxon ranges. Math. Geol. 21:411–421.
Tanner M., Wing H. 1987. The calculation of posterior distributions by data augmentation. J. Amer. Stat. Assoc. 82:528–540.
Thorne J.L., Kishino H., Painter I.S. 1998. Estimating the rate of evolution of the rate of molecular evolution. Mol. Biol. Evol. 15:1647–1657.
Tougard C., Delefosse T., Hnni C., Montgelard C. 2001. Phylogenetic relationships of the five extant Rhinoceros species (Rhinocerotidae, Perissodactyla) based on mitochondrial cytochrome b and 12S rRNA genes. Mol. Phylogenet Evol. 19:34–44.
Vamosi J.C., Vamosi S.M. 2011. Factors influencing diversification in angiosperms: at the crossroads of intrinsic and extrinsic traits. Am. J. Bot. 98:460–471.
Vose D. 2008. Risk Analysis: A quantitative guide. 3rd ed. Wiley.
Wagner P.J., Marcot J., Slater G. 2013. Modelling distributions of fossil sampling rates over time, space and taxa. Assessment and implications for macroevolutionary studies. Methods Ecol. Evol. 4:703–713.
Wasserman L. 2000. Bayesian model selection and model averaging. J. Math. Psychol. 44:92–107.
Wilkinson R.D., Steiper M.E., Soligo C., Martin R.D., Yang Z., Tavaré S. 2011. Dating primate divergences through an integrated analysis of palaeontological and molecular data. Syst. Biol. 60:16–31.
Xie W., Lewis P.O., Fan Y., Kuo L., Chen M.-H. 2011. Improving marginal likelihood estimation for Bayesian phylogenetic model selection. Syst. Biol. 60:150–160.
Yang Z. 1993. Maximum-likelihood estimation of phylogeny from DNA sequences when substitution rates differ over sites. Mol. Biol. Evol. 10:1396–1401.
Yang Z. 1994. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: Approximate methods. J. Mol. Evol. 39:306–314.
Yang Z., Rannala B. 1997. Bayesian phylogenetic inference using dna sequences: A Markov chain Monte Carlo method. Mol. Biol. Evol. 14:717–724.