Is BAMM Flawed? Theoretical and Practical Concerns in the Analysis of Multi-Rate Diversification Models

Daniel L. Rabosky[1,*], Jonathan S. Mitchell[1], and Jonathan Chang[2]

[1]*Department of Ecology and Evolutionary Biology and Museum of Zoology, University of Michigan, Ann Arbor, MI 48109, USA and* [2]*Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA 90095, USA*
*Correspondence to be sent to: Department of Ecology and Evolutionary Biology and Museum of Zoology Ruthven Museums Building, Museum of Zoology, University of Michigan, Ann Arbor, MI 48109-1079, USA; Email: drabosky@umich.edu.

Abstract.—Bayesian analysis of macroevolutionary mixtures (BAMM) is a statistical framework that uses reversible jump Markov chain Monte Carlo to infer complex macroevolutionary dynamics of diversification and phenotypic evolution on phylogenetic trees. A recent article by Moore et al. (MEA) reported a number of theoretical and practical concerns with BAMM. Major claims from MEA are that (i) BAMM's likelihood function is incorrect, because it does not account for unobserved rate shifts; (ii) the posterior distribution on the number of rate shifts is overly sensitive to the prior; and (iii) diversification rate estimates from BAMM are unreliable. Here, we show that these and other conclusions from MEA are generally incorrect or unjustified. We first demonstrate that MEA's numerical assessment of the BAMM likelihood is compromised by their use of an invalid likelihood function. We then show that "unobserved rate shifts" appear to be irrelevant for biologically plausible parameterizations of the diversification process. We find that the purportedly extreme prior sensitivity reported by MEA cannot be replicated with standard usage of BAMM v2.5, or with any other version when conventional Bayesian model selection is performed. Finally, we demonstrate that BAMM performs very well at estimating diversification rate variation across the ∼20% of simulated trees in MEA's data set for which it is theoretically possible to infer rate shifts with confidence. Due to ascertainment bias, the remaining 80% of their purportedly variable-rate phylogenies are statistically indistinguishable from those produced by a constant-rate birth–death process and were thus poorly suited for the summary statistics used in their performance assessment. We demonstrate that inferences about diversification rates have been accurate and consistent across all major previous releases of the BAMM software. We recognize an acute need to address the theoretical foundations of rate-shift models for phylogenetic trees, and we expect BAMM and other modeling frameworks to improve in response to mathematical and computational innovations. However, we remain optimistic that that the imperfect tools currently available to comparative biologists have provided and will continue to provide important insights into the diversification of life on Earth. [BAMM; diversification; macroevolution; phylogeny; speciation.]

Rates of lineage diversification—speciation and extinction—vary widely across the Tree of Life, and this variation underlies many large-scale patterns of biological diversity. Variation in evolutionary rates contributes to striking disparities in species richness among groups of organisms, to the waxing and waning of clades in deep time, and to latitudinal and other spatial diversity gradients. Consequently, there has been widespread interest in methods for inferring the dynamics of speciation and extinction from molecular phylogenies of extant taxa. Dated phylogenies of extant species contain an imperfect record of the history of speciation through time, even if the phylogeny itself has been estimated without error (Nee et al. 1994; Mooers and Heard 1997; Barraclough and Vogler 2002). The record is imperfect because a speciation event cannot be observed in a reconstructed phylogenetic tree unless one or more descendants of both progeny lineages have survived to the present. Many statistical tools have been developed to extract information about speciation and extinction rates from molecular phylogenies (Ricklefs 2007; Stadler 2013; Morlon 2014), including the extent to which those rates have varied through time (Pybus and Harvey 2000; Morlon et al. 2010; Silvestro et al. 2011; Stadler 2011) and among lineages (Alfaro et al. 2009; Morlon et al. 2011; Etienne and Haegeman 2012; Lewitus and Morlon 2016).

Bayesian Analysis of Macroevolutionary Mixtures (BAMM) is a computer program for inferring the dynamics of speciation, extinction, and phenotypic evolution on phylogenetic trees (Rabosky 2013, 2014a; Rabosky 2014). The diversification models implemented in BAMM are based on a simple birth–death process, a feature shared with most other methods for studying diversification rates on phylogenies (O'Meara 2012). BAMM uses reversible jump Markov chain Monte Carlo (rjMCMC) to infer complex mixtures of distinct evolutionary rate dynamics across the branches of phylogenies. In the BAMM framework, a "rate shift" is a transition to a new set of evolutionary parameters along a branch in a reconstructed phylogenetic tree, and a "shift configuration" is a particular mapping of evolutionary rate parameters across the phylogeny as a whole (Rabosky 2014b). A set of rate shifts and associated parameter values mapped to a set of branches on a phylogeny, thus defines a particular shift configuration. Using rjMCMC, BAMM simulates a posterior distribution of shift configurations conditioned on an observed phylogenetic data set. The resulting distribution reflects uncertainty in evolutionary rates across the phylogeny and can be parsed to extract a range of summary attributes, including tip-, branch-, and clade-specific rates of speciation and extinction. The method also provides statistical evidence for the number and location of rate shifts across phylogenies.

BAMM belongs to a more general set of diversification methods that relax the assumption of rate homogeneity across the branches of a phylogenetic tree. Some

TABLE 1.   Key claims by Moore et al. (2016) concerning the BAMM method for macroevolutionary dynamics

MEA claim	How tested by MEA	Assessment in present study
1. BAMM likelihood function contains "serious error" as it ignores rate shifts on extinct lineages	Likelihoods computed by BAMM compared to those from MEA's independent BAMM likelihood calculator that includes Monte Carlo simulation of extinction probabilities	Test whether MEA likelihood satisfies axioms of probability. Test whether MEA correctly implemented BAMM likelihood function
2. Likelihoods for rate-shift models are invalid if they fail to account for rate shifts on extinct/unobserved lineages	Biological relevance asserted, but not tested	Test whether unobserved rate shifts bias extinction probabilities for biologically relevant parameterizations of the diversification process. Test whether MEA Monte Carlo extinction estimates differ from constant-rate birth-death (CRBD) process process approximation
3. CPP prior model for diversification rate shifts in BAMM is problematic	Theoretical demonstration that prior distribution of rate shifts across tree is nonuniform under BAMM's CPP prior model	Assess BAMM's performance on empirical and simulated data sets
4. Posterior on number of shifts extremely sensitive to prior	Analyzed empirical and simulated data sets with BAMM v2.5 after activating an experimental (non-default) program setting; results not compared to default program performance. Compared model posterior probabilities but did not use Bayes factors to account for model prior	Reanalyze MEA data sets with BAMM v2.5 default settings; compare to MEA results. Use Bayes factors for model selection, as recommended by BAMM developers
5. Diversification-rate estimates using BAMM are unreliable	Simulated phylogenies with rate shifts; analyzed each with BAMM. Assessed accuracy of branch-specific diversification estimates	Assess ascertainment bias in MEA simulation study. Test whether MEA "rate variable" phylogenies have sufficient statistical power to infer rate heterogeneity. Determine whether MEA results reflect biased inference with BAMM or low power to infer rate variation

methods, including those that allow rates of speciation (λ) and/or extinction (μ) to vary through time (Nee et al. 1994; Rabosky and Lovette 2008; Morlon et al. 2010; Stadler 2011; Etienne et al. 2012) assume that all contemporaneous lineages have precisely the same rates of speciation and extinction. All lineages in these homogeneous models are said to be of the same "type" and have the property that any two lineages i and j have the same evolutionary rates at a given point in time from the root node, for example, $\lambda_i(t) = \lambda_j(t)$, and $\mu_i(t) = \mu_j(t)$. This homogeneous class of models can be contrasted with multi-type branching process models that allow mixtures of different types of lineages across the branches of a phylogeny, such that lineages potentially differ in their rates of speciation and extinction. The calculations in BAMM are based on the multi-type branching process, as are many other methods (Maddison et al. 2007; Alfaro et al. 2009; Morlon et al. 2011; Beaulieu and O'Meara 2016). As discussed below, BAMM utilizes a subclass of multi-type branching processes that we will refer to as "rate-shift" models. These models enable the calculation of the likelihood of a specific configuration of diversification rate regimes on a phylogenetic tree.

A recent article by Moore et al. (2016) assessed the theoretical foundations of the inference model in BAMM and its Bayesian implementation. Moore et al. (hereafter, MEA) state that the likelihood function in BAMM is incorrect and that the prior model for rate shifts is theoretically unsound. MEA explored the performance of BAMM in practice and reported statistical pathologies that render inferences with BAMM unreliable. The authors conclude, through analysis of simulated and empirical data sets, that (i) posterior estimates of the number of rate shifts are extremely sensitive to the assumed prior; and (ii) diversification rate parameters are uncorrelated with the true values in the simulation model. On the basis of these concerns, the authors concluded that BAMM is flawed and should not be used.

Here, we assess the validity of major claims from MEA and conclude that their results are either incorrect or inconclusive. A summary of these claims, how they were tested by MEA, and how we assess their validity in the current study, is provided in Table 1; claim numbers are cross-referenced in major section headings. Our treatment in this article generally follows Table 1: we first address theoretical issues associated with BAMM and MEA likelihoods, then consider BAMM's performance in practice (including earlier versions of the software). We welcome critical analysis of BAMM and believe that progress in our field requires continued attention to the assumptions that underlie this and other inference tools. However, we believe that MEA have provided an inaccurate assessment of BAMM and its limitations. It is important to evaluate these issues fairly because several concerns raised by MEA pertain to all rate-shift models. We conclude that BAMM is an imperfect tool that has a clear path for theoretical improvement, but which nonetheless performs and has performed well in practice. Furthermore, we believe that the critical future

directions for rate-shift models involve issues that are either neglected or incorrectly emphasized by MEA.

Scope of the present article.—Here we focus primarily on the validity of MEA's claims regarding BAMM v2.5, which was used to obtain all results in their article. This version is the most recent major release of BAMM and has been available longer than any other single release of the program as of January 2017. Nonetheless, we have included a comprehensive performance assessment of all major previous releases of BAMM to determine the reliability of the program as used in earlier research. We also describe the theoretical and implementational differences between these major releases of the software in the Supplementary Material that accompanies this article (available on Dryad at http://dx.doi.org/10.5061/dryad.36g21). We note that our conclusions are restricted to diversification models, as MEA did not critique the phenotypic evolutionary models implemented in BAMM.

Reanalysis of MEA input files.—The Dryad submission to accompany MEA's article does not include their complete BAMM output, but does include all input files required to replicate their analyses. We repeated all BAMM analyses exactly as performed by MEA using their input (control) files using BAMM v2.5. As described below, MEA included a nonstandard option ("combineExtinctionAtNodes = random"), not visible to end users, that altered a default value in their analyses; we therefore performed a second set of analyses where we restored this undocumented feature to its default value in order to obtain BAMM results that are consistent with those that would be obtained by a typical user of the program. Throughout our article, we reference MEA's results but acknowledge that these were obtained by repeating their analyses using their analysis files exactly as they were provided on Dryad. Given the stochastic nature of MCMC, we expect to observe minor numerical discrepancies between our results and theirs.

CLAIM 1. ARE THE BAMM AND MEA LIKELIHOODS COMPARABLE?

The fundamental operation in the analysis of rate-shift models involves the calculation of the likelihood of a phylogenetic tree under a given set of evolutionary rate parameters and mapping of transition points (rate shifts) between parameter sets on the tree. MEA claim that the BAMM likelihood is incorrect because it fails to account for rate shifts that may have occurred on unobserved lineages. To assess the consequences of unobserved rate shifts for BAMM, MEA provided an independent implementation of the BAMM likelihood function where the extinction probabilities are estimated under a computationally intensive Monte Carlo approximation that simulates the effects of unobserved rate shifts. They compared extinction probabilities and likelihoods

computed by BAMM to those of their independent implementation with Monte Carlo extinction estimates (MEA Figs. 2 and 3). All analyses from MEA assume that their likelihood function is correct, an assumption that was only tested for the constant-rate (no shift) birth–death process. MEA's results further assume that the only difference between BAMM and their implementation involves the effects of unobserved rate shifts on extinction probabilities. MEA found that their likelihoods differed from those computed by BAMM, which they attributed exclusively to the effects of unobserved rate shifts. However, if MEA did not correctly implement the BAMM likelihood function and/or if their own likelihood function is invalid, then they have not demonstrated an effect of unobserved rate shifts on the likelihoods computed by BAMM, because any discrepancies between their implementation and BAMM could reflect other differences in the algorithms used to compute the likelihood. In this section, we focus solely on whether MEA have demonstrated an effect of unobserved rate shifts on the BAMM likelihood; the next major section (Claim 2) addresses the biological relevance of these unknown quantities.

Likelihoods of Rate-Shift Models

As MEA explain, computing the likelihood of a phylogeny where rate shifts have been placed on the tree can be considered a form of data augmentation; we retain this terminology for comparison with their article but note that many other researchers would simply consider the shifts to be part of the model (e.g., they are not data). The "augmentation," in this case, is the specific mapping of rate regimes across a phylogeny (and their associated parameters) that is used during the calculation of the likelihood. We can contrast the likelihood of a phylogeny under a diversification process with a data-augmented history (as in BAMM) with the corresponding likelihood under the complete process, without data augmentation. A likelihood calculation with BiSSE (binary-state speciation and extinction: Maddison et al. 2007; FitzJohn et al. 2009), for example, does not involve any data augmentation: although the model assumes that there are two classes of diversification rates across the tree, the likelihood is computed without placing the rate shifts (e.g., character state changes) on the branches of the tree. A likelihood calculation with BiSSE involves integrating over all possible histories of diversification rate changes that could yield the observed tree and its associated character state data.

Data augmentation is essential for likelihood calculations under rate-shift models for two reasons. First, the number of possible diversification processes is infinite, rendering the BiSSE-type calculations infeasible; this point is discussed clearly by MEA. Perhaps more importantly, we are not generally interested in the likelihood of a phylogeny after integrating over all possible rate shifts that could produce the observed data. Most researchers are interested in the likelihoods of alternative data-augmented histories

TABLE 2.   Diversification models and software implementations that allow speciation and extinction rates to vary among lineages ("rate shift" models)

Model/reference	Extinction probability E(*t*) includes unobserved rate shifts?	E(*t*) accounts for data augmentation?
MEDUSA (Alfaro et al. 2009)	No	No
Split-BiSSE and all split-class SSE models (FitzJohn et al. 2009, 2012)	No	No
Split-clade models (Morlon et al. 2011)	No	Conditions on survival of all subclades that have undergone rate shifts
DDD with subclade shifts (Etienne and Haegeman 2012)	No	Conditions on survival of the crown clade, given that one of the descendant lineages undergoes a rate shift
BAMM (Rabosky 2014)	No	Implicitly conditions on existence of subtrees leading to rate shifts
Moore et al. (2016)	Simulates expected effects of unobserved rate shifts on E(*t*) under prior assumptions	No

(e.g., configurations of rate shifts). A question such as: how does the likelihood of a given phylogeny with no rate shifts compare to the corresponding likelihood of the same tree with one rate shift is a question about the relative probability of two data-augmented histories. Hence, the data-augmented histories and their associated likelihoods are the objects of interest for most empirical researchers.

Rate-shift models with data augmentation have been widely used to model diversification heterogeneity across phylogenies (Alfaro et al. 2009; Morlon et al. 2011; Etienne and Haegeman 2012; Rabosky 2014). Moreover, state-dependent models with "split" parameter sets (FitzJohn 2010, 2012) also fall into this category, as they assume discrete shifts in parameter sets at specific locations on the tree with no formal stochastic process for transitioning between parameter sets on each partition. In split-BiSSE, for example (FitzJohn 2010, 2012), we assume that there are two or more rate partitions across a phylogeny, each of which has a set of parameters associated with it. Within partitions, we may be integrating over possible diversification histories, but the assumption of distinct parameter partitions and associated shift locations is data augmentation, at least under MEA's terminology. In fact, if we constrain a split-BiSSE model to the special case where there is no state-dependence of diversification rates, we have specified a mapping of diversification parameters and rate-shift locations across a phylogeny, precisely as in MEDUSA and BAMM.

The likelihood of a particular rate-shift configuration is based on a set of differential equations that describe transition probabilities for a stochastic birth–death process. These equations are solved backward in time along individual branches of the tree, from the tips to the root. The differential equations for the likelihood involve two probabilities. The first, $D_i(t)$, is the probability that a lineage in the i-th rate regime at some point in time (i.e., a location on an observed branch of a phylogeny) gives rise to all observed lineages descended from that particular

point on the tree. The second equation, $E_i(t)$, describes the probability that a lineage in the i-th rate regime has gone extinct before the present. Letting λ_i and μ_i denote the corresponding speciation and extinction rates for the regime, we have

$$\frac{dD_i}{dt} = -(\lambda_i + \mu_i + \eta)D_i(t) + 2\lambda_i D_i(t)E_i(t) \qquad (1)$$

and

$$\frac{dE_i}{dt} = \mu_i - (\lambda_i + \mu_i + \eta)E_i(t) + \lambda_i E_i(t)^2 + \eta\Phi(t,\eta) \qquad (2)$$

The parameter η corresponds to the rate at which a lineage shifts to a new rate regime, similar to the Λ term in Rabosky (2014). The $\eta\Phi(t, \eta)$ term describes the probability that a lineage at time t undergoes a rate shift to new rate regime, drawn from the set of all possible shift parameters, and is subsequently unobserved; MEA describe this term as pertaining to extinct rate shifts, but it applies more generally to rate shifts that occur on branches that are unobserved for any reason. For consistency with MEA, we will generally interpret this term as the chance of extinction due to rate shifts on unobserved branches of a phylogeny; MEA's incorporation of this term into the formal mechanics of rate-shift models is an advance that has largely been overlooked by prior work on this topic. Note that BiSSE and related models allow rate shifts (e.g., state changes) on unobserved lineages, but accommodating these is straightforward, as the model specifies a finite number of lineage types: there are only two possible types of lineages in BiSSE, corresponding to the two character states. The state-space in BAMM and related models is infinite, such that there is no solution to $\eta\Phi(t, \eta)$.

With the exception of MEA, all published rate-shift models at the time of this writing have implicitly assumed $\eta = 0$ (Table 2). The likelihood functions in all such models, including BAMM, condition the likelihood on the nonoccurrence of rate shifts on extinct side branches of the observed phylogeny (see MEA
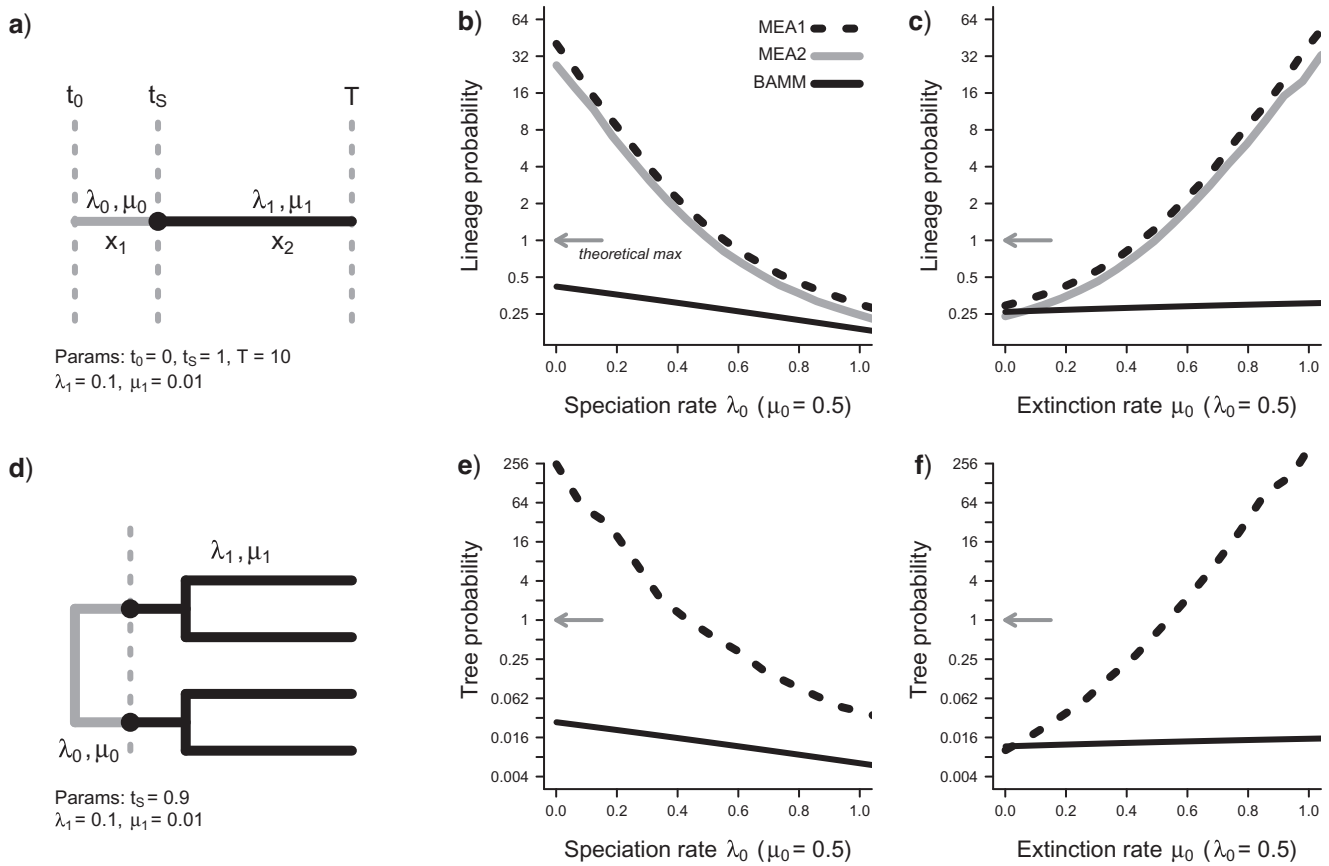
FIGURE 1.   Probabilities of hypothetical lineage histories under MEA and BAMM-type algorithms for computing likelihoods. Top row, a–c: probability of a single lineage history, where a lineage begins at time t with parameters (λ_0, μ_0) and shifts to new parameters (λ_1, μ_1) at time t_s. The probability of the lineage history is shown as a function of the speciation rate λ (b) and extinction rate (c), with all other parameters fixed to the values indicated in the figure. In (b) and (c), MEA1 and MEA2 refer to independent implementations of the MEA probability; MEA1 uses the likelihood calculator distributed with MEA, and MEA2 is an independent implementation of the same equations using a different numerical integration method. Bottom row, d–f: probability of a four-taxon tree with two independent rate shifts to identical parameter values. MEA probability (dashed line) is the likelihood obtained with MEA's Monte Carlo likelihood calculator after conditioning the likelihood on the observed speciation events. In all cases, high extinction probabilities at the beginning of the process lead to invalid and potentially infinite probabilities under MEA's incorrect conditioning scheme for E(t).

Fig. 1). We disagree with MEA's terminology that this conditioning constitutes a theoretical error—all models involve simplifying assumptions—but we agree that it is important to assess whether or not a simplifying assumption has a negative impact on inference.

Likelihood of Rate Shifts under MEA

MEA provide an innovative solution to approximating the value of $\eta \Phi(t, \eta)$ through computationally intensive Monte Carlo simulation of the extinction probability at each point in time given prior assumptions about the underlying distributions of λ and μ. The η parameter is heavily influenced by the data augmentation (e.g., the number of shifts placed on the observed portion of the tree), but the $\Phi(t, \eta)$ term is influenced both by the inferred shift rate and prior assumptions about the underlying distributions of λ and μ.

At the root of the tree, the likelihood of the data is conditioned on the probability of observing a phylogeny,

as is commonly done for phylogenies of extant taxa (Nee et al. 1994). Thus, the likelihood of the tree τ is divided by the probability that the process at the root of the tree (θ_R) leaves two surviving crown lineages. The likelihood of the data-augmented history is

$$L\left(\tau | \theta_R, \theta_1, \theta_2 \ldots \theta_n; P_S = 1\right) = \frac{L\left(\tau | \theta_R, \theta_1, \theta_2 \ldots \theta_n\right)}{\left(1 - E_R\left(t_0\right)\right)^2} \quad (3)$$

where P_s denotes the probability of crown-clade survival, t_0 is the time of the root node in the tree, and $\theta_1, \theta_2, \ldots, \theta_n$ denotes the set of n non-root rate regimes that have been placed on the tree. The $E_R(t_0)$ term is the probability of extinction for a single lineage originating at the root with parameters θ_R.

The likelihood described above suffers from a deep theoretical problem that has largely gone unrecognized by the field, even when we ignore the potential effects of unobserved rate shifts ($\eta = 0$). Specifically, the extinction probability at the root of the tree, $E_R(t_0)$, must account

for the fact that we are computing the likelihood of a tree with data-augmented histories that can make the process more or less likely to survive than would be expected based on the parameters at the root of the tree. In other words, the extinction probability of a lineage must be conditioned on the data augmentation, or the set of rate shifts that are assumed to have occurred.

Under MEA's likelihood, all calculations for E(t), including those used for conditioning at the root of the tree, are performed only with the current values of speciation and extinction (e.g., $\lambda_i(t)$, $\mu_i(t)$). The extinction probability of the entire process is a simple function of the age of the tree and the rates at the root (θ_R) and ignores all information regarding rate shifts that have been placed on the tree. It is straightforward to demonstrate that this conditioning scheme, which is used by several modeling frameworks (Table 2), can lead to nonsensical probabilities and incorrect inference. To be clear, this problem is not consistently resolved in the literature: three methods that implement some form of E(t) conditioning (Morlon et al. 2011; Etienne and Haegeman 2012; Rabosky 2014) entail different assumptions (Table 2), suggesting an acute need for a comprehensive theoretical treatment of this topic.

To demonstrate that the MEA likelihood is not valid, we will first convert the likelihood expression into a formal probability. We will then demonstrate that the method of computation can yield probabilities that are not bounded on the interval (0, 1). A formal proof of this property is provided in the Appendix, but here we consider two empirical scenarios with data augmentation. In the first scenario, we have a single observed lineage segment where some initial set of parameters (λ_0, μ_0) shifts to a new set of parameters (λ_1, μ_1) at some time t_s (Fig. 1a). In the second scenario, a phylogeny of four taxa undergoes independent rate shifts on each of the lineages descended from the crown node (Fig. 1d). This four-taxon tree scenario enables us to compute the probability a phylogenetic tree using the likelihood calculator distributed with MEA's Dryad files, which ensures that the properties we describe here are not strictly a function of our implementation of the MEA likelihood expression.

The unconditioned likelihood of the data-augmented history shown in Figure 1a, with distinct sets of rate parameters on intervals x_1 and x_2, is

$$P(x_0|\lambda_0,\mu_0,x_1)P(x_1|\lambda_1,\mu_1)\eta \qquad (4)$$

where P(x|λ,μ) denotes the probability of the corresponding segment as computed using Equations 1 and 2. This expression is a probability density (see Maddison et al. 2007 for discussion of this point), due to the presence of the η term. By conditioning on the observation of a shift at t_s and on the nonextinction of the process, we convert the expression into a formal probability, or

$$P(x_0|\lambda_0,\mu_0,x_1)P(x_1|\lambda_1,\mu_1)/[1-E_0(t_0)] \qquad (5)$$

where $E_0(t_0)$ is the probability that a single lineage with parameters (λ_0, μ_0) at the start of the process goes extinct before the present. We can perform a similar conditioning for the four-taxon tree, except that we also condition on the existence of two non-root speciation events, the root node, and the survival of the crown clade.

The behavior of the MEA probability as a function of the speciation and extinction rates on the root segment (x_0) is shown in Figure 1b,c. When the chance of extinction of the root process (λ_0, μ_0) is high, the probability of the data can increase arbitrarily to infinity (Appendix). A similar effect is noted for the probability of the four-taxon tree using the MEA likelihood calculator (Fig. 1e,f). The root conditioning in MEA's code for likelihood calculations is described in our Supplementary Material (sections S2.1–S2.2, available on Dryad); we also demonstrate that the MEA likelihood can tend to infinity for the cetacean data set (Supplementary Fig. S1, available on Dryad).

The failure of the MEA likelihood to remain bounded between 0 and 1 is a simple consequence of incomplete data augmentation. The data-augmented process contains a rate shift (or two rate shifts, in the four-taxon tree). This rate shift, under these parameterizations, renders the extinction of the overall process unlikely (in this example, the shift parameters involve a rate regime with a very small chance of extinction). Hence, if the process survives to time t_s, a rate shift will occur that dramatically changes the overall probability of extinction of the process. The MEA equation does not account for this: because the extinction probability is a strict function of the age of the process and the parameters at the root, the chance of extinction can asymptotically approach unity, even if the data-augmented history implies that the true chance of extinction is low. Specifically, the denominator in Equation (5) can approach zero, even when the numerator is unity (Appendix). Hence, the equation is unbalanced, and the unconditioned probability in Equation (4) (which can take any value between 0 and 1) is divided by probability of survival that, in the limit, can approach zero. In contrast, the corresponding BAMM probability is conditioned on the data augmentation and is bounded on the correct interval.

BAMM and Other Likelihoods with E(t) Augmentation

The BAMM likelihood differs from the MEA likelihood, and simply because of the assumption that $\eta = 0$ (Table 2). Specifically, BAMM attempts to obtain a consistent definition of E(t) at the root of the phylogeny that can be used to condition survival of the process by recursively passing down previously computed E(t) values from the tips to the root (Fig. 2). Hence, E(t) at any point in time is the extinction probability of a lineage at time t given the complete downstream (tipward) shift history. At internal nodes that differ in shift histories of their descendants (e.g., node n_{AB} in Fig. 2, but not node n_{CD}), the probabilities are
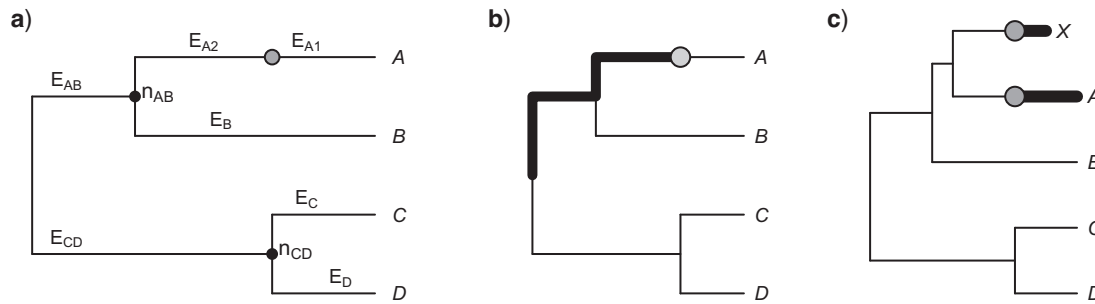
FIGURE 2. Likelihood calculations under BAMM v2.5, illustrated on a hypothetical phylogeny with a single rate shift (on branch A). Due to rate shift, branch segment E_{A1} has rate parameters that are distinct from the rest of the tree. Likelihoods are computed from the tips to the root by passing previously computed extinction probabilities $E(t)$ down the tree. Thus, the $E(t)$ component of the likelihood calculation on branch segment E_{A2} is initialized with the value previously computed on E_{A1}. When a node is reached, the extinction probabilities $E(t)$ are multiplied together if the nodes differ in their downstream shift histories (node n_{AB}). Hence, the initial value for the E_{AB} segment is the product of extinction probabilities on E_{A2} and E_B. This multiplication does not occur at node n_{CD}, because the lineages have identical shift histories. This algorithm for computing the likelihood entails two assumptions (b, c). The likelihood of the tree as a whole is effectively conditioned on the existence of a subtree leading to the mapped rate shifts (bold branches in (b)). Second, any extinct lineages that branched off before the rate shift but after the most recent surviving node (n_{AB}), are also assumed to have undergone a rate shift (e.g., extinct lineage X in (c)). These unusual assumptions were incorporated into BAMM v2.5 pending the development of a theoretically coherent method for computing $E(t)$ with data-augmented histories (Table 2) and avoid the infinite likelihood trap of several other approaches (Appendix; Fig. 1).

multiplied together and used as initial $E(t)$ inputs for the next (rootward) branch segment. The advantage to this algorithm is that the extinction probability $E_R(t)$ at the root of the tree is conditioned on the observed shift history: the two $E(t)$ values that one obtains at the root (one each for right and left descendant branches) after computing the likelihood can immediately be used to condition $D(t)$ at the root on the probability that the process survives (Equation (3)). However, this approach entails several unusual assumptions and is best viewed as an approximation to the true likelihood, which is currently unknown. First, these calculations lead to strong topological conditioning of $E(t)$: the likelihood is essentially conditioned on the existence of a subtree leading to the set of shifts that have been placed on the tree (Fig. 2b). Second, the model assumes that any extinct lineages that branched off of a focal branch prior to the rate shift (e.g., lineage X in Fig. 2c) have also undergone a rate shift at the same time. A comprehensive theoretical treatment of the assumptions that underlie these calculations is provided in the Supplementary Material (sections S2.4–S2.5).

We acknowledge that these calculations are approximate and were motivated by our observation that MEA-type calculations (as in split-BiSSE), with incomplete data augmentation, could incorrectly lead to infinite likelihoods; we demonstrate this point in the Appendix. Several other approaches use data augmentation to compute $E(t)$, which avoids the more significant pathologies associated with the unconditioned likelihood expression. Morlon et al. (2011) compute the likelihood of a phylogeny with subclade shifts, conditioning on the survival of each subclade that has undergone a rate shift. Etienne and Haegeman (2012) compute the likelihood of a branching process where a single lineage shifts to a new diversity-dependent regime at some time t_s. Their solution is to treat the $E(t)$ calculations at the root of the tree as

distinct from the $E(t)$ terms used to solve the differential equations (Equations (1–2)) along individual branches. Once the (unconditioned) likelihood of the data has been computed, the probability that both crown lineages have survived is computed, given that one of the subclade lineages will undergo a rate shift at time t_s if the process as a whole survives to time t_s. It is this extinction probability that is used to condition the probability of the tree. However, the Etienne and Haegeman (2012) approach cannot easily be applied to scenarios where two or more rate shifts have been placed on the tree, and the proper approach to conditioning likelihoods under rate-shift models is currently an unresolved theoretical problem.

In contrast to the BAMM likelihood, the MEA implementation computes $E(t)$ as a strict function of the parameters of the process at time t. In Figure 2, for example, the MEA extinction probability on segment E_{A2} would be independent of the extinction probability used to compute the likelihood of segment E_{A1}. Hence, MEA did not correctly implement the BAMM likelihood function for the comparisons that underlie their Figures 2 and 3: their function differs from the BAMM likelihood in numerous respects, including the incorrect data augmentation scheme referenced above (Appendix; Supplementary Material section S2.4, available on Dryad). The MEA and BAMM likelihoods should thus be expected to differ numerically for many reasons, but MEA attributed these differences exclusively to the effects of unobserved rate shifts.

Summary: Likelihoods of Rate-Shift Models

We agree with MEA that many outstanding theoretical issues remain to be resolved with rate-shift models; we also agree that MEA have correctly identified an issue with most rate-shift models in that they ignore
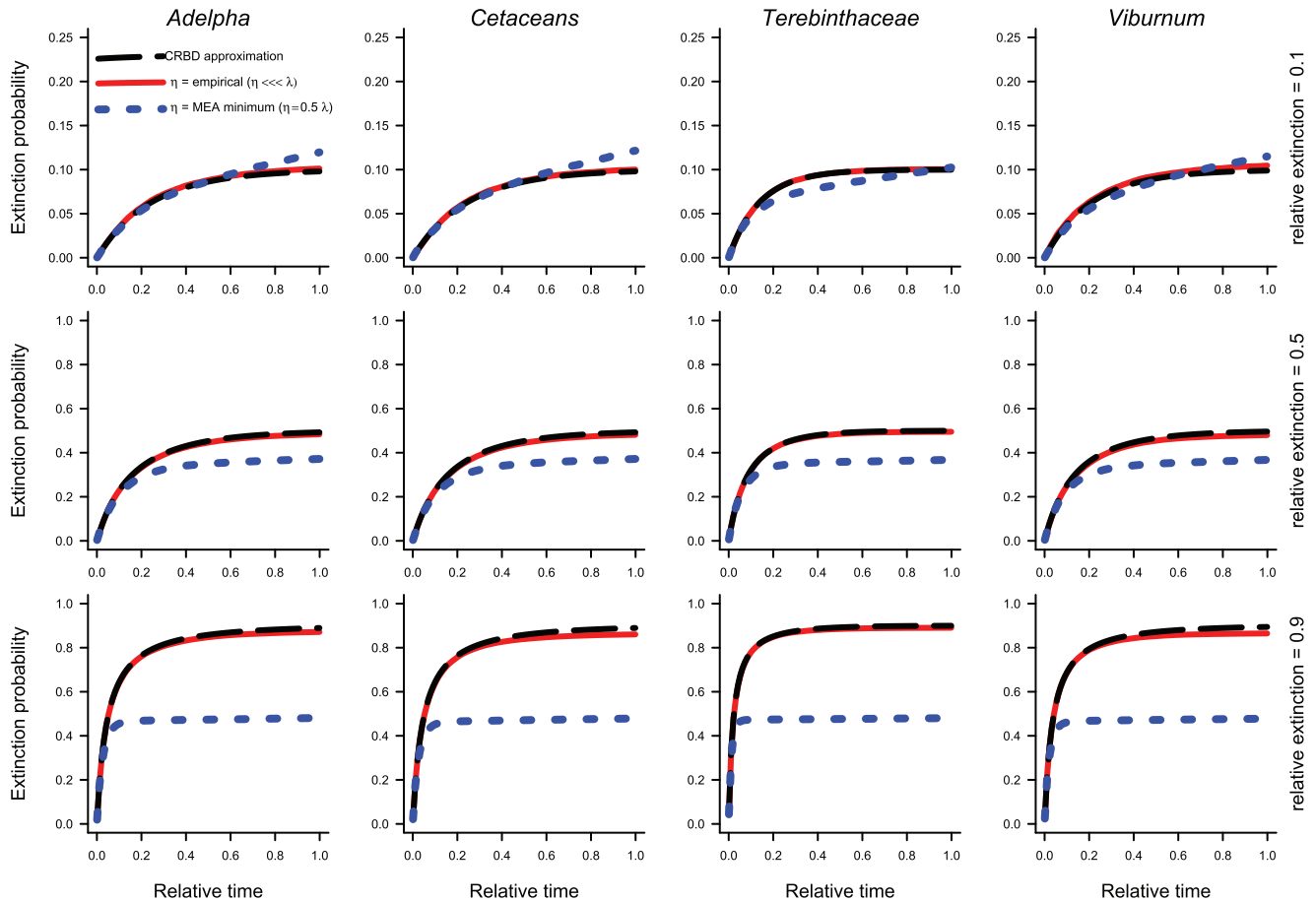
FIGURE 3. The chance of lineage extinction as a function of relative time from the root node for four empirical data sets analyzed by MEA (columns) and under three different relative extinction rates (rows). When rate shifts are rare relative to speciation events, the effect of unobserved rate shifts on the extinction probability is negligible. Red (solid) lines shows cumulative chance of extinction estimated using MEA's Monte Carlo simulator and parameterized with empirical estimates of the rate shift frequency (transition rate) η. Black (dashed) line shows the corresponding extinction probability under a simple constant-rate birth–death process. Blue line (dotted) shows the change in the extinction probability under the minimum nonzero rate-shift frequency illustrated by MEA (Supplementary Fig. S5, available on Dryad). MEA are correct that unobserved rate shifts can influence the extinction probability, but the parameter space over which the effect becomes important involves rate shift frequencies that are approximately 100–1000× greater than for the empirical data sets considered in their paper.

rate shifts on unobserved lineages (Table 2). If concerns about unobserved rate shifts are of practical significance, then the likelihood calculations implemented in many software programs will be similarly affected, including MEDUSA (Alfaro et al. 2009), diversitree (FitzJohn 2012), DDD (Etienne et al. 2012), and RPANDA (Morlon et al. 2016). However, the MEA likelihood function itself contains an important error—an error that is shared with several other software implementations of rate-shift models. Specifically, the MEA likelihood is compromised by incorrect data augmentation (Table 2) that leads to invalid probabilities and potentially infinite likelihoods. The MEA likelihood expression is fundamentally different from the BAMM expression; differences include the nature of conditioning at the root and the handling of extinction probabilities at shift points and internal nodes (Fig. 2). Thus, differences in the likelihood and E(t) terms for these implementations (e.g., Figs. 2 and 3 from MEA) do not demonstrate pathologies with BAMM that can be conclusively

attributed to the effects of unobserved rate shifts on extinct lineages, as claimed by MEA.

CLAIM 2. RATE SHIFTS ON UNOBSERVED LINEAGES: ARE THEY IMPORTANT?

The MEA likelihood is an important theoretical contribution for its clear discussion of rate shifts on unobserved lineages and their potential consequences for inference. However, as noted above, the likelihoods computed by BAMM and using MEA's independent implementation are not comparable and provide no information about the effects of unobserved rate shifts on the likelihood. It is unclear whether the computationally intensive procedure suggested by MEA significantly enhances our ability to accurately model phylogenetic data, especially as the MEA likelihood calculations are many orders of magnitude slower than BAMM's calculations. In this section, we ask a simple question: does the Monte Carlo extinction

simulator used by MEA yield extinction probabilities that are appreciably different from those computed under a constant-rate birth–death (CRBD) process that ignores rate shifts on extinct lineages?

As justification for the importance of considering unobserved rate shifts, MEA provide a Supplementary Figure S5 from MEA (available on Dryad) that illustrates substantial effects of unobserved rate shifts on the extinction probability. However, the MEA parameterizations involve rate-shift frequencies that are much higher than the values obtained for their empirical analyses. MEA illustrate the effects of unobserved rate shifts for processes where rate shifts occur at rates from one-half to eight times the rate at which speciation events occur (e.g., minimum $\eta = 0.5\lambda$). For example, under MEA's illustrative parameterizations, a phylogeny of 100 extant tips would potentially have been generated under a process that included 50 or more rate shifts on the observed (extant) portion of the tree; their most extreme parameterizations would have such trees generated under processes that include many hundreds of rate shifts. We suspect that, when contemplating the effects of rate shifts or key innovations on phylogenetic diversification patterns, most empirical researchers are not conceptualizing a process where rate shifts outnumber the branches on which to place them.

The parameterizations considered by MEA far exceed values of η estimated in MEA's empirical analyses (Supplementary Figs. S9–S11, available on Dryad). For example, across all empirical data sets they consider, the estimated ratio of η to λ is approximately 0.01 (and sometimes much lower), indicating that speciation events are 100 times more likely—on average—than rate-shift events. This result is intuitively appealing: rate shifts are rare, and BAMM in general does not infer many of them, at least under the most commonly used prior expectation of 1.0 shift per tree. If the rate shift frequency η approaches or exceeds the speciation rate λ, we would, in general have low power to infer shifts, because inferential power is a function of the number of taxa in the shift regime. With $\eta > \lambda$, the mean number of taxa per shift regime becomes too small to infer shift regimes with confidence.

We assessed the extent to which the MEA rate-shift parameterizations matched the rate-shift frequencies from their empirical analyses, and we considered whether rate shifts on unobserved/extinct branches significantly affected the overall extinction probability of the clade. For each empirical data set in MEA, we estimated rates of speciation and extinction under a CRBD process under three relative extinction rates: $\mu/\lambda = 0.1$, 0.5, and 0.9. We then estimated the transition rate η from the rate-shift frequency ("event rate") sampled using rjMCMC with BAMM. As pointed out by MEA, estimates of η from the observed portion of a phylogeny may be biased relative to the true value of η; however, we found that, under MEA's parameterizations, estimates of η across the observed portion of the tree are highly correlated with the value that would be estimated if we knew the complete tree with extinct lineages and unobserved shifts (Supplementary Fig. S8, available on Dryad).

We used MEA's Monte Carlo extinction simulator to generate estimates of the extinction probabilities through time for each of the empirical data sets they considered. The Monte Carlo simulator estimates the extinction probability of an independent lineage where rate shifts to new parameters are permitted to occur with rate η. When a rate shift occurs, the simulator samples new parameters from prior distributions (e.g., these results are dependent on the prior assumptions about speciation and extinction rate distributions). We parameterized the priors on speciation and extinction rate distributions to have a mean equal to the inferred speciation rate for the tree, as in MEA's likelihood calculations. We parameterized estimates of η by selecting the 95th percentile of the distribution of estimates for each empirical tree; this is conservative, as using any lower quantile of this distribution (e.g., the median) would decrease the frequency of rate shifts relative to speciation. By deliberately overestimating the empirical shift rate η, we are maximizing the effects of any unobserved rate shifts on parameter estimates.

We then used MEA's simulator to generate the corresponding extinction-through-time curve using $\eta = 0.5\lambda$, the minimum nonzero ratio of transition-to-speciation events considered in MEA Supplementary Figure S5 (available on Dryad). Finally, we computed the corresponding extinction-through-time probabilities for each clade under the assumption of a simple CRBD process. This latter probability is analytical (Kendall 1948; Bailey 1964; Raup 1985) and depends only on the age of the process and the diversification rates, as it assumes no rate shifts ($\eta = 0$).

Across all relative extinction rates, we find only a marginal contribution of unobserved/extinct rate shifts to the total extinction probability of the clade (Fig. 3; Supplementary Figs. S11–S12, available on Dryad). The extinction trajectories using the empirically parameterized Monte Carlo simulator are nearly identical to the analytical expectation assuming $\eta = 0$ (red vs. black lines). Under the smallest nonzero η parameterization considered by MEA ($\eta = 0.5\lambda$), there is a substantial effect with respect to the total extinction probability (Fig. 3; blue dotted lines). Supplementary Figures S11–S12 (available on Dryad) illustrate the lack of effect that unobserved rate shifts have on the extinction probability across all empirical data sets from MEA's study. The equivalence of analytical and empirical extinction curves indicates that rate-shift models parameterized with biologically relevant values of η are unlikely to experience a significant contribution of this term to the chance of extinction. However, as MEA show, rate-shift frequencies that are 100–1000 times greater than empirical estimates have the potential to compromise likelihood calculations unless accounted for through simulation.

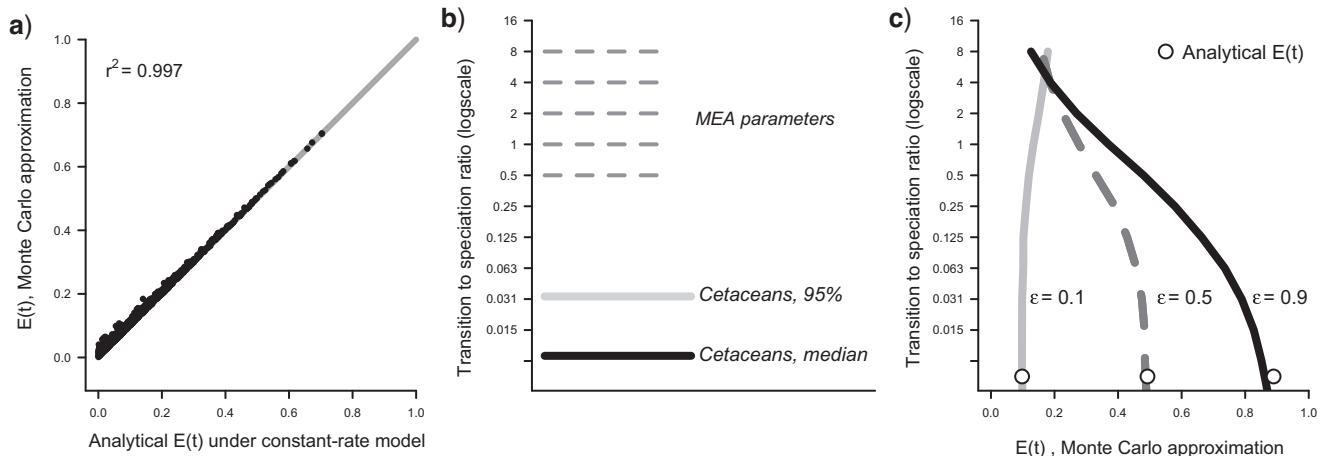We then looked in detail at the extinction probabilities computed by MEA for the cetacean data set, which

FIGURE 4. Rate shifts on unobserved lineages have a marginal effect on likelihood calculations when the transition rate η is low relative to speciation. (a) Estimates of the extinction probability E(t) at the root of the cetacean phylogeny estimated using MEA's Monte Carlo simulator and compared to the corresponding analytical E(t) probabilities from simple CRBD process. Each point represents the extinction probability for a set of root parameters sampled from the posterior simulated with BAMM, exactly as for MEA Figure 2. (b) Ratios of transition rates (η) to speciation rates for the cetacean data set and compared to the five illustrative parameterizations considered by MEA (Supplementary Fig. S5, available on Dryad); note logarithmic scale for rates. (c) The extinction probability at the root of the cetacean phylogeny for a constant-rate diversification process under three relative extinction scenarios (ε), as a function of the transition-to-speciation rate ratio. When transition rates are low relative to speciation, as for the cetaceans, the extinction probabilities are similar to the corresponding constant-rate estimates (open circles), which assume η = 0. The strongest effects are observed when rate shifts are approximately as frequent or more frequent than speciation events; we expect that rate shifts in such scenarios are difficult to detect.

figures prominently in their assessment of BAMM's performance. This test is important, because MEA Figure 2 purports to show proportional error in extinction probabilities computed by BAMM relative to those estimated with their Monte Carlo simulator for this data set. The authors attribute the difference in extinction probabilities to the effects of unobserved rate shifts. We compared the extinction probabilities computed at the root of the cetacean phylogeny exactly as computed by MEA, to the corresponding analytical probabilities under CRBD process with no rate shifts (η = 0). The resulting probabilities are very highly correlated (Fig. 4a; $r^2 > 0.99$) and suggest minimal contribution of unobserved shifts to the overall extinction probability. This result further suggests that the demonstrable numerical discrepancies between likelihoods computed by BAMM and MEA (MEA Figs. 2 and 3) primarily reflect the combined influence of error in their likelihood equation (Appendix) as well as algorithmic differences between the two implementations (Supplementary Material section S2.4, available on Dryad). We also plotted the empirical transition rates for the cetaceans to illustrate the discrepancy in scale between these empirical rates and those used in MEA's illustrative curves (Fig. 4b). Finally, we estimated the root extinction probabilities for the cetacean data set as a function of the transition-to-speciation rate ratio (η/λ), under three relative extinction rates (Fig. 4c). These latter results indicate that the overall extinction probability is largely invariant across empirically relevant parameterizations, but changes dramatically for the rate ratios considered by MEA.

Summary: Unobserved Rate Shifts

MEA claim that the likelihood function in BAMM (and other models; Table 2) is invalid because it ignores the effects of unobserved rate shifts. However, we find a marginal contribution of extinct rate shifts to the total extinction probability under biologically plausible (empirically parameterized) values of the shift rate η. Our results indicate that the primary difference between extinction probabilities and likelihoods in BAMM versus MEA has little to do with the contribution of unobserved rate shifts (Fig. 4a), and much to do with fundamental differences in the way the likelihoods are computed (Fig. 2; Table 2). MEA likelihoods should differ from those computed by BAMM even if MEA likelihoods assume η = 0, due to MEA's incorrect implementation of the BAMM likelihood and theoretically invalid root conditioning. Finally, if unobserved rate shifts are sufficiently frequent as to impact inference with formal rate shift models, then they would also affect inference with all other diversification models that ignore rate shifts on unobserved lineages. In the unlikely possibility that true clade extinction probabilities are strongly influenced by unobserved rate shifts to high-extinction regimes, one cannot simply make the problem go away by using theoretically coherent models (e.g., BiSSE; constant-rate birth-death process) that assume an absence of such shifts. The effects of unobserved shifts would still be present in the data, and the use of simpler models that ignore their effects cannot remove the footprint of unobserved shifts from the data itself.
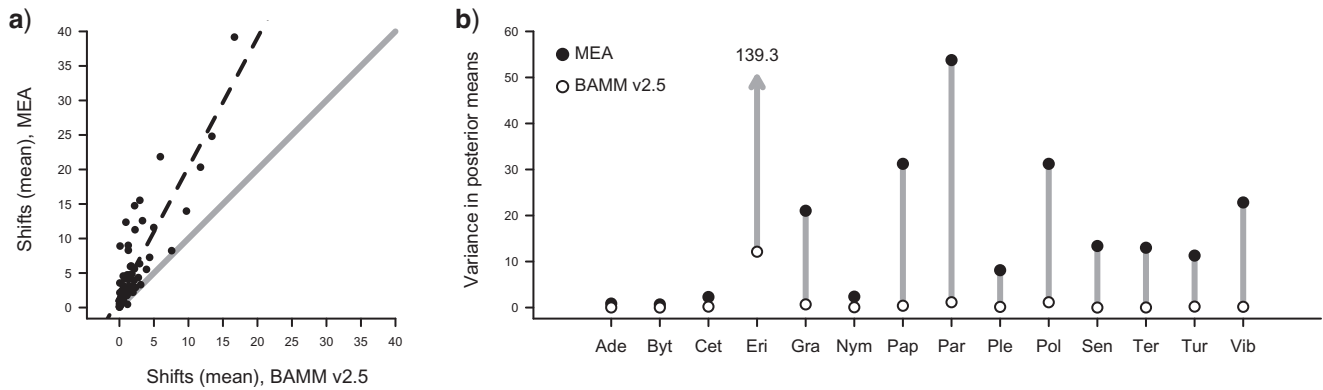
FIGURE 5. The posterior distribution of the number of rate shifts is much more sensitive to the prior when the "random" option activated by MEA is used for analysis, relative to BAMM v2.5 default values. (a) Means of the marginal posterior distributions on the number of shifts across all data sets and priors considered by MEA as analyzed with BAMM defaults (x-axis) and as used by MEA (y-axis). The mean number of shifts is biased upward with MEA's analysis configuration. (b) The variance in mean number of shifts for 14 empirical data sets across 5 prior parameterizations considered by MEA, estimated under BAMM v2.5 default settings (open circles) and as used by MEA (solid circles). Data sets are abbreviated by first three letters in names used by MEA.

CLAIM 3. IS THE COMPOUND POISSON PROCESS PRIOR MODEL IN BAMM INCOHERENT?

We largely agree with MEA that the prior model in BAMM deserves further exploration. MEA demonstrate that the compound Poisson process (CPP) model used in BAMM induces nonuniform probabilities for rate shifts on single branches, and that the model induces non-Poisson behavior in the posterior distribution of rate shifts (MEA Fig. 5). MEA claim that these theoretical concerns lead to statistical pathologies in practice, including prior sensitivity and unreliable diversification estimates. However, as we demonstrate in the following three sections, MEA did not provide valid evidence that BAMM performs poorly. Nonuniform, poorly specified, and/or arbitrary priors are used widely in ecology and evolution, including phylogenetics and divergence time estimation (Pickett and Randle 2005; Alfaro and Holder 2006; Brown et al. 2009; Heled and Drummond 2012; Heath and Moore 2014). We contend that concerns about the shapes of the underlying prior distributions in BAMM, and other methods, are largely irrelevant unless they result in demonstrable pathologies in the shape of the posterior.

CLAIM 4. IS PRIOR SENSITIVITY PROBLEMATIC FOR BAMM?

MEA claim that the posterior distribution on the number of shifts obtained with BAMM shows extreme prior sensitivity, reflecting fundamental pathologies with the underlying CPP model. All Bayesian methods are characterized by some prior sensitivity, as the posterior is a function of both the likelihood and the prior. Even if the posterior on the number of shifts is prior-sensitive, we do not consider this result to be inherently problematic: if the data are sufficiently informative, the likelihood will shift the posterior away

from the prior. Furthermore, prior sensitivity is not necessarily problematic if Bayes factors are used to assess the evidence in favor of rate variation, or if a conservative prior is used for analysis. Rabosky (2014) clearly demonstrated that use of liberal priors influenced the marginal posterior distribution of shifts, with minimal effect on rate estimates, for early versions of the BAMM software (see Figs. 3 and 10 from Rabosky 2014). MEA presented only the raw posterior distributions simulated with BAMM but did not use Bayes factors to assess whether the resulting inferences were dominated by the prior. MEA do not demonstrate that spurious inferences result from the purported prior sensitivity in BAMM; they merely re-document the prior sensitivity previously described by Rabosky (2014) in an early software release. Most published studies to date have used a conservative prior ($\gamma=1$) that, even accepting MEA's results at face value, would not have resulted in the inference of spurious rate shifts.

However, results in MEA cannot be replicated with standard usage of BAMM v2.5. As detailed in Supplementary Material section S2.3 (available on Dryad), MEA added an undocumented option to their BAMM control files that induces unpredictable behavior in the likelihood calculations (Supplementary Figs. S2 and S3, available on Dryad); this setting is not visible to end users and MEA do not justify why this option was used in their article. Unfortunately, the setting used by MEA (referred to here as "random") does not replicate behaviors from previous releases of the program and has not been tested by us. Indeed, the likelihood calculations in BAMM v2.3 (released 8 March 2015) and v2.4 (13 June 2015) are nearly identical to the MEA likelihood (but with $\eta=0$) and used a similarly incorrect data augmentation scheme (see Supplementary Material sections S2.4–S2.5, available on Dryad); it is BAMM v2.5 that introduced the subtree conditioning shown in Figure 2. In BAMM v2.3 and v2.4, E(t) is always computed using the current parameter
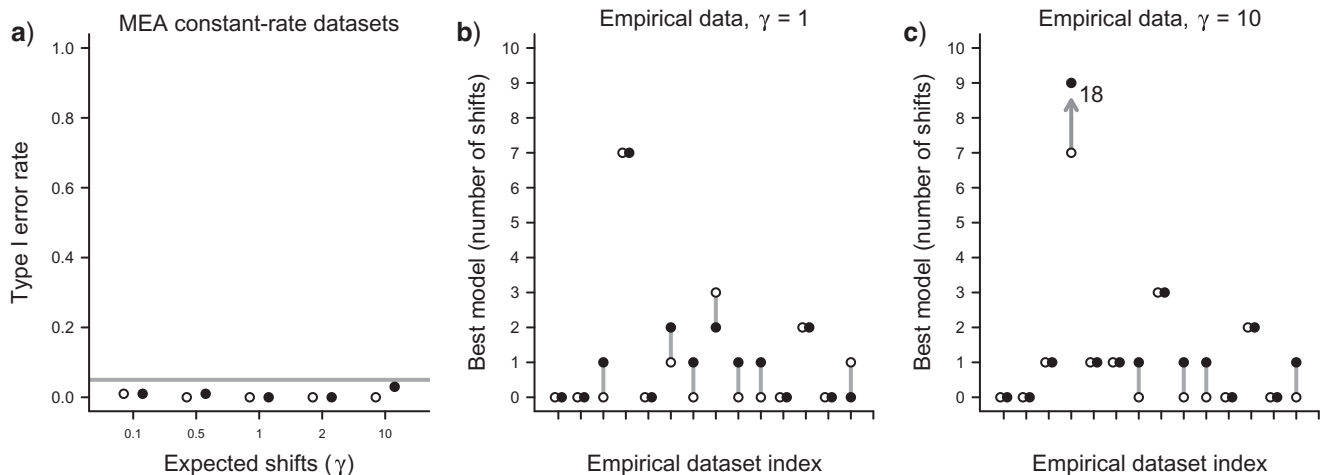
FIGURE 6.   Model selection with Bayes factors is generally robust to choice of model prior, regardless of how BAMM v2.5 is used. a) Type I error rates for the MEA constant-rate phylogenies are low regardless of whether BAMM is used with v2.5 defaults (open circles) or MEA settings (filled circles). Type I error rate is the proportion of simulated trees where Bayes factor comparisons led to rejection of the true (zero shift) model. (b) Best model for 14 empirical data sets considered by MEA with $\gamma = 1$ (expected number of shifts); (c) Best model for empirical data sets with $\gamma = 10$. Regardless of the posterior on the number of shifts, Bayes factors generally favor models with zero or one shift for this set of phylogenies. A single data set was found to have large numbers of shifts (Ericaceae; 7/18 for BAMM v2.5/MEA settings), but this was also the largest data set and includes 4500 species (450 sampled in the phylogeny).

values at time t, as in MEA. It is unclear why the "random" setting was used in MEA, as this choice deviates from the default setting for BAMM v2.5 and, to our knowledge, no other publications have used this setting. The only BAMM versions that would have enabled MEA to draw conclusions about the effects of unobserved rate shifts on the likelihood, given the implementation they used in their own their likelihood calculator, are BAMM v2.3 and v2.4; all other versions involve fundamentally different data augmentation with respect to E(t), as illustrated in Figure 2. Given this unusual usage of BAMM v2.5, we assessed whether the recommended (and commonly used) version of the program demonstrates the same prior sensitivity reported by MEA.

We repeated all BAMM analyses using input files from MEA, both with the default and MEA ("random") settings. All BAMM runs were performed across prior expectations of $\gamma = 0.1$, 0.5, 1, 2, and 10 expected shifts across the phylogeny, as in MEA. We then computed several statistics to describe BAMM's sensitivity to the prior on the number of rate shifts (γ). For the 14 empirical data sets considered by MEA (Supplementary Figs S19–S32 from their article), we simply plotted the mean of the marginal posterior distribution on the number of shifts for BAMM v2.5 with default settings versus BAMM with the MEA "random" option. We then computed, for each data set, the variance in the posterior shift means across the five prior scenarios. If, for a given data set, the posterior mean is relatively insensitive to the prior, we expect the variance to be small.

BAMM with the MEA "random" option shows much greater prior sensitivity than standard BAMM v2.5 (Fig. 5). The mean of the marginal posterior distribution on the number of shifts is consistently higher with the

MEA setting. Likewise, the variance in posterior means (across $\gamma = 0.1$, 0.5, 1, 2, and 10) is much lower with standard usage of BAMM v2.5 (Fig. 5b).

Despite the differences observed between "random" and standard BAMM v2.5 (Fig. 5), we do not consider prior sensitivity inherently problematic, unless it consistently leads to incorrect inference. We have previously advocated the use of Bayes factors for model selection with BAMM (Rabosky 2014b; Mitchell and Rabosky 2016), which permits robust model selection that is less sensitive to the prior. MEA found that, with the "random" setting, the posterior is highly sensitive to the prior for constant-rate trees that lack rate shifts, suggesting that BAMM may be associated with a high Type I error rate if nondefault priors are used ($\gamma \gg 1$). We performed stepwise model selection for constant-rate and empirical phylogenies using Bayes factors. Starting with the least complex model (e.g., fewest shifts), we tested whether a model with one additional shift fit the data significantly better. If so, we rejected the lower complexity model in favor of the model with the additional shift. The process is repeated until an incremental increase in model complexity does not yield a significant improvement in model fit. We considered Bayes factor evidence of 20 in favor of one model over another to be "significant", as in Rabosky et al. (2014b) and May et al. (2016).

When Bayes factors are used for model selection, there is virtually no tendency toward model overfitting regardless of whether standard BAMM or the MEA "random" option is used for analysis (Fig. 6a). Type I error rates for all prior parameterizations were less than 0.05, indicating that—after appropriately controlling for the prior on the number of shifts—BAMM does not tend to overfit the data. Similar results were observed for the

empirical data sets. In general, the "random" setting leads to support for greater model complexity relative to standard BAMM (Fig. 6b), but the effect is weak to moderate.

Summary: Prior Sensitivity in BAMM

With standard usage of BAMM v2.5, we cannot replicate MEA's finding of extreme prior sensitivity with respect to the number of rate shifts. Moreover, when BAMM is used with nonstandard settings ("random") from MEA, proper model selection with Bayes factors does not lead to spurious inference of rate shifts. Most importantly, even if the posterior on the number of shifts is prior-sensitive, we see no reason why this should preclude usage of the program, provided that researchers choose a conservative prior on the number of shifts. Most published results obtained with BAMM have used a prior of $\gamma = 1$; even under MEA's analytical protocol, results obtained with this prior are conservative. This value has always been the default setting in BAMM and specifies that a model with zero shifts is twice as likely (under the prior) as the most probable model that includes rate shifts.

CLAIM 5. DOES BAMM PROVIDE RELIABLE DIVERSIFICATION RATE ESTIMATES?

MEA report that BAMM performs poorly at estimating diversification rates across phylogenies simulated with rate shifts. They used a Poisson process model of rate variation to simulate trees, which is very similar to the true process modeled by BAMM. Extinction rates estimated with BAMM and other methods are generally poor (Davis et al. 2013, Rabosky 2015; but see Beaulieu and O'Meara 2015), but MEA's results for speciation are surprising given the reasonable performance reported by Rabosky (2014). In particular, Rabosky (2014) found good performance of BAMM even when the inference model did not match the generating model.

Do MEA's Analyses Have Sufficient Power to Justify their Conclusions?

MEA's variable-rate trees are characterized by small total size and small tip-to-shift ratios (Fig. 7), suggesting that the trees may contain low information content with respect to rate heterogeneity. MEA's variable rate trees are restricted to sizes of 50–150 tips, and the median size of a shift regime is just a single tip. Fully 58% of MEA's trees did not include a shift regime with more than five tips (Fig. 7), and we believe most researchers would not expect any software program to reliably detect rate regimes that include five or fewer tips.

The prevalence of these small shift regimes (Fig. 7b) is due to a substantial ascertainment bias in MEA's simulation design that selected for trees with small numbers of rate shifts and/or rate shifts of small
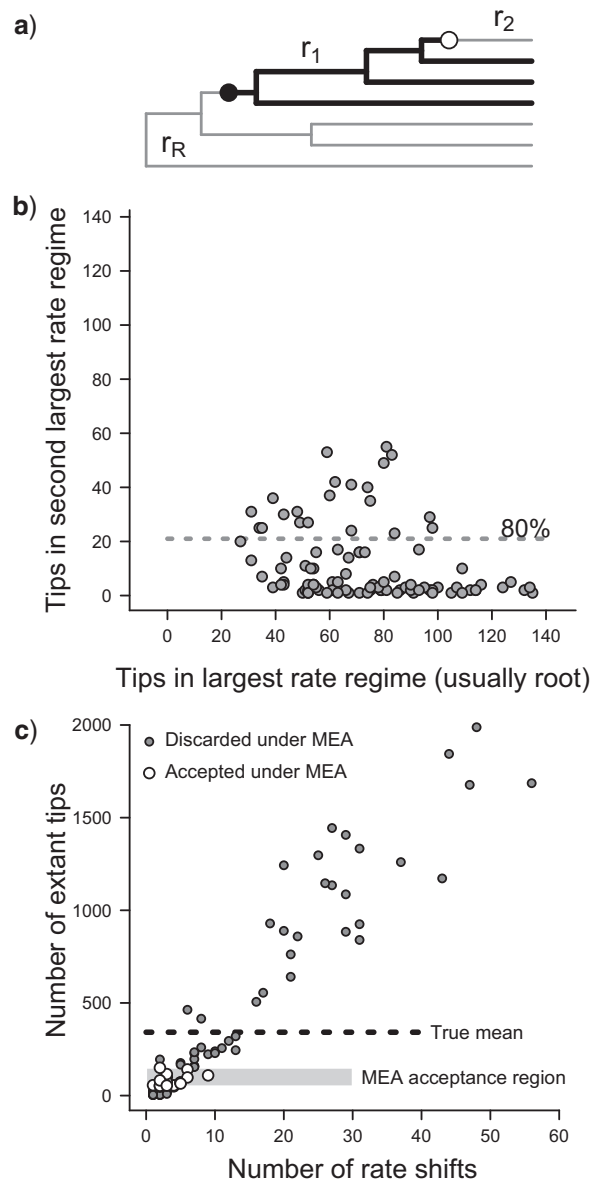

FIGURE 7. a) Phylogeny illustrating three rate regimes: r_1 (bold lines), r_2, and the root regime (r_R). Regimes r_R and r_1 have three extant tips, and r_2 includes a single tip. b) Numbers of tips in the largest and second-largest rate regimes for each tree in MEA's variable-rate dataset. In general, trees are dominated by a single rate regime (usually the root regime); the second-largest regime included fewer than 20 tips in 80% of trees (<5 tips in 58% of trees). BAMM and other methods would be unlikely to detect rate variation in phylogenies where most rate shifts lead to very small subclades, as in this data set. c) MEA's variable rate simulations are comprised by ascertainment bias, as they discarded all simulations that did not meet highly restrictive size criteria (gray polygon). A comparable data set of 100 phylogenies simulated using their code and parameter values without imposing this arbitrary criterion demonstrates that tree sizes are on average much larger and contain far more shifts, with higher information content, than the trees in their data set (true mean: 342 tips; MEA data set mean: 89 tips). Panel illustrates the relationship between tree size (extant tips) and the number of shifts in each tree; all trees identified with filled circles would have been rejected under MEA's acceptance criteria. This sampling bias generates phylogenies that are dominated by excessively small rate regimes (panel b) relative to the Poisson generating process described in their paper.

effect. Using MEA's simulation code with identical input parameters, we generated an equivalent data set of 100 phylogenies but retained all trees with 2000 or fewer extant tips. It is clear that trees selected by MEA represent a small fraction of the simulation outcomes produced under their parameterization: the mean tree size in the data set they analyzed contained merely 89 extant tips, but the mean we obtained from simulations under their parameters (without restrictions) was 342 (Fig. 7c). The consequences of this ascertainment bias are clear when we plot the number of tips as a function of the number of shifts present in each simulated tree (Fig. 7c): MEA selected for a set of trees that are unusual, relative to the Poisson process expectation, in that (i) they contain a small number of tips overall, (ii) they contain fewer shifts, and (iii) rate shifts tend to be of small effect, such that they lead to clades with small numbers of extant tips (Fig. 7b). Rate shifts of large effect, no matter how common under the MEA parameterization, tend to generate trees that exceed the 150 taxon upper bound that they imposed on their simulations and would have generally been excluded from their analyses (e.g., Fig. 7c: filled circles). Section 5.1 in the Supplementary Material (available on Dryad) that accompanies this article describes these and other ascertainment biases in detail.

If BAMM requires reasonable numbers of tips in a rate regime in order to detect a shift (e.g., $n > 10$), the method would have been unable to infer rate heterogeneity for most of their simulated trees. We tested the information content associated with each rate shift in the MEA data set to determine whether it would have been possible, in principle, to infer rate shifts from these data. For each rate regime in the MEA variable-rates tree set, we computed the exact likelihood of the rate regime under the true parameters in MEA's generating model. We compared this likelihood to the corresponding likelihood of the same rate regime under the parameter values we would estimate for the whole tree under a constant-rate birth-death process. Thus, for each of the 100 variable-rates phylogenies in MEA, we estimated λ and μ under a CRBD process. Let $\tau_{S,i}$ denote the subtree from tree τ containing all branch segments and nodes to which the i-th rate regime has been mapped, as illustrated in Figure 7a. The information content $\Delta LogL_i$ associated with this rate regime is

$$\Delta LogL_i = \log L(\tau_{S,i}|\theta_{TRUE}) - \log L(\tau_{S,i}|\theta_{CRBD})$$

where $L(\tau_{S,i}|\theta_{TRUE})$ is the likelihood of the corresponding subtree under the true parameters in the generating model and θ_{CRBD} denotes the corresponding speciation and extinction rates obtained for the entire tree under a two-parameter, CRBD process. This statistic tells us how much information a given shift regime has relative to the information we would obtain from simply computing the likelihood of the shift regime under the whole-tree estimates for speciation and extinction (e.g., assuming no rate variation across

the tree). The likelihood is a summary of the information in the data. Consequently, if a given rate regime has any information with which it can be inferred, $\Delta LogL$ must be greater than zero. If $\Delta LogL < 0$, the rate regime is more probable under a simple whole-tree estimate of diversification rates. Such negative values are possible because we are not optimizing parameters for each shift regime; for comparison, we perform such optimization in the Supplementary Material and show that it yields virtually identical results but with all $\Delta LogL > 0$.

It is not sufficient for $\Delta LogL$ to be greater than zero, because the parameters and location of the shift (θ_{TRUE}) are fixed exactly to their true values in the calculation above (e.g., $\Delta LogL$ is a difference in likelihoods between two nested models with different numbers of parameters). Hence, $\Delta LogL$ must be somewhat greater than zero in order to conclude that rate variation is present, a fact that becomes obvious when we consider this analysis in an Akaike information criterion (AIC) framework. The expression for AIC in terms of the difference in likelihoods ($\Delta LogL$) and the number of parameters k required to fit a shift can be rearranged to give $\Delta LogL = S/2 + k$, where S is the ΔAIC score that we would require to accept the more complex model (e.g., a model with an additional shift). If we interpret ΔAIC evidence greater than or equal to zero as supporting the more complex model, the minimum possible $\Delta LogL$ score that will enable us to infer the rate shift is 3.0 with $k = 3$; $k = 3$ corresponds to the number of parameters associated with each shift in MEA's generating process (parameters: shift location, λ, μ). Any stricter AIC threshold, including the generally accepted value of 2.0 as well as all AICc-type modifications, can only increase the requisite $\Delta LogL$ criterion above this minimal estimate.

We computed $\Delta LogL$ for all 435 rate shifts present in the 100 trees included in the MEA variable rates data set. We computed likelihoods following MEA's approach of accounting for unobserved rate shifts, using their Monte Carlo simulator with the exact parameters that were used in the generating model (λ prior mean = 0.15; μ prior mean = 0.05; $\eta = 0.006$). For comparison, we also computed the likelihood of each rate shift under the analytical birth–death process ($\eta = 0$; no Monte Carlo simulation of extinction probabilities). If $\Delta LogL$ values with the fast analytical approximation are approximately identical to those computed while accounting for unobserved rate shifts, there is little reason in practice to adopt the more computationally intensive inference scheme.

We find that there is virtually no information content associated with the vast majority of rate shifts in the MEA variable rates data set. Of 435 rate shifts across 100 phylogenies, 411 have $\Delta LogL < 3$ (Fig. 8a). We find it difficult to imagine that BAMM—or any other method—would be able to infer rate shifts in any scenario where $\Delta LogL < 3$. Indeed, given that BAMM is sampling from prior distributions on all parameters, we predict that $\Delta LogL$ would need to be considerably higher in order to detect a rate shift. Only 14 shifts,
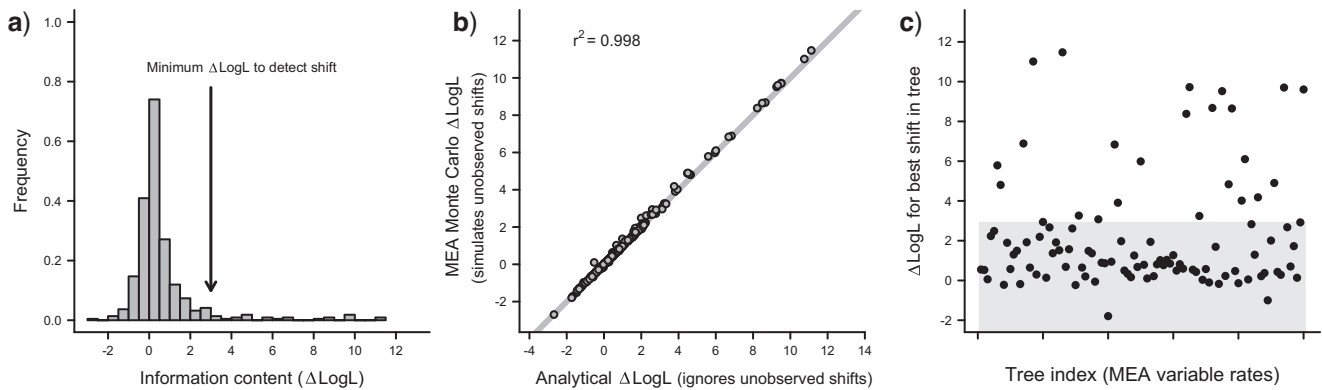
FIGURE 8.   Information content for all rate shifts in the MEA variable-rates data set. ΔLogL denotes the difference in likelihoods for each rate shift under the true parameter values relative to the likelihood for the rate shift assuming that the entire tree has a single invariant rate of speciation and extinction. In an AIC framework, a rate shift is unlikely to be detectable unless ΔLogL is somewhat greater than 3. a) Frequency distribution of ΔLogL for all 435 rate shifts in the MEA data set, where all likelihoods were computed with MEA's Monte Carlo simulator to account for unobserved rate shifts. The vast majority (95%) of all rate shifts in the MEA data set contain virtually no information and probably cannot be inferred by any method. b) ΔLogL values computed using MEA's Monte Carlo simulator are virtually identical to ΔLogL values computed analytically assuming $\eta = 0$; integrating over prior distributions on unobserved rate shifts thus contributes little to our ability to model the data. c) ΔLogL values for the best (most information-rich; most "inferrable") shift in each of the 100 trees in the MEA data set. The vast majority of trees have effectively no information with which to infer rate heterogeneity (gray zone). Given the contribution of the prior and the stochasticity inherent in MCMC, it is likely that ΔLogL = 3 is a conservative estimate of information required to infer a shift in the BAMM framework. For the Monte Carlo likelihood estimates, transition rates and speciation/extinction priors were set to the exact values used by MEA in the generating model.

across the entire MEA set of 100 trees, have ΔLogL >5. Importantly, we find that likelihoods computed with MEA's computationally intensive Monte Carlo simulator are nearly identical to those computed under the fast analytical approximation (Fig. 8b), consistent with results presented in a previous section (Figs. 3 and 4). We performed a second analysis where we optimized rate parameters for every shift regime, thus substituting the maximum likelihood estimates of θ for θ_{TRUE}. Optimizing parameters in this fashion is a simple check that enables us to ensure that there were no errors in the recording of parameter values in MEA's data files and also accounts for ascertainment biases and other factors that could shift the ML parameter estimates for each rate regime away from their true values. This procedure had minimal effect on the distribution of ΔLogL, and only 17 shifts had ΔLogL >5 with this approach (Supplementary Material Fig. S14).

These results demonstrate that MEA's "variable rates" phylogenies contain, with few exceptions, the same amount of information as trees produced by a CRBD process. If BAMM performs well, we should expect that (i) the program should generally not recover evidence for rate variation for most phylogenies in this data set; (ii) that BAMM estimates of rates for low-power trees should converge on the tree-wide average rate; and (iii) that BAMM should be able to recover branch-specific variation in diversification rates for the few trees where the information content is sufficient for the method to infer rate heterogeneity. The simple power calculations performed here suggest that MEA's simulated data sets should not be used to study within-tree patterns of diversification rate variation: across the majority of their trees, there is minimal information that can be used to infer rate heterogeneity.

Assessment of BAMM Rate Reliability

We repeated all BAMM analyses of MEA's variable rates trees but we constrained BAMM to be identical to the generating model (e.g., within-regime rates assumed to be constant through time). This constraint also serves to minimize the impact of well-known ascertainment biases that compromise temporal analyses of diversification when trees are selected nonrandomly with respect to total size (Phillimore and Price 2008); note again that MEA's tree set contains substantial size-related ascertainment bias (Fig. 7c; Supplementary Material section S5.1, available on Dryad). We summarized BAMM estimates of evolutionary rates by branch and by rate regime. In the former analysis, as in MEA, we tested whether branch-specific estimates of diversification rates were correlated with the true values. We also tested whether mean BAMM rates for rate regimes were correlated with the true values across the full data set. We present all analyses with respect to the theoretical information content ΔLogL associated with each rate shift. Corresponding results for BAMM with time-varying rate regimes are very similar and are provided in the Supplementary Figures S15–S18 (available on Dryad).

Figure 9 (top row) illustrates the correlation between true speciation rates and the corresponding branch-specific estimates from BAMM, as a function of maximum ΔLogL (information content) associated with the tree. Each data point in Figure 9 is the correlation for a single tree in MEA's data set; the max ΔLogL for each tree is the largest ΔLogL value observed across all shifts in a given tree (e.g., max ΔLogL = 2 implies that all shifts in the tree have ΔLogL ⩽ 2). It is critical to note that if
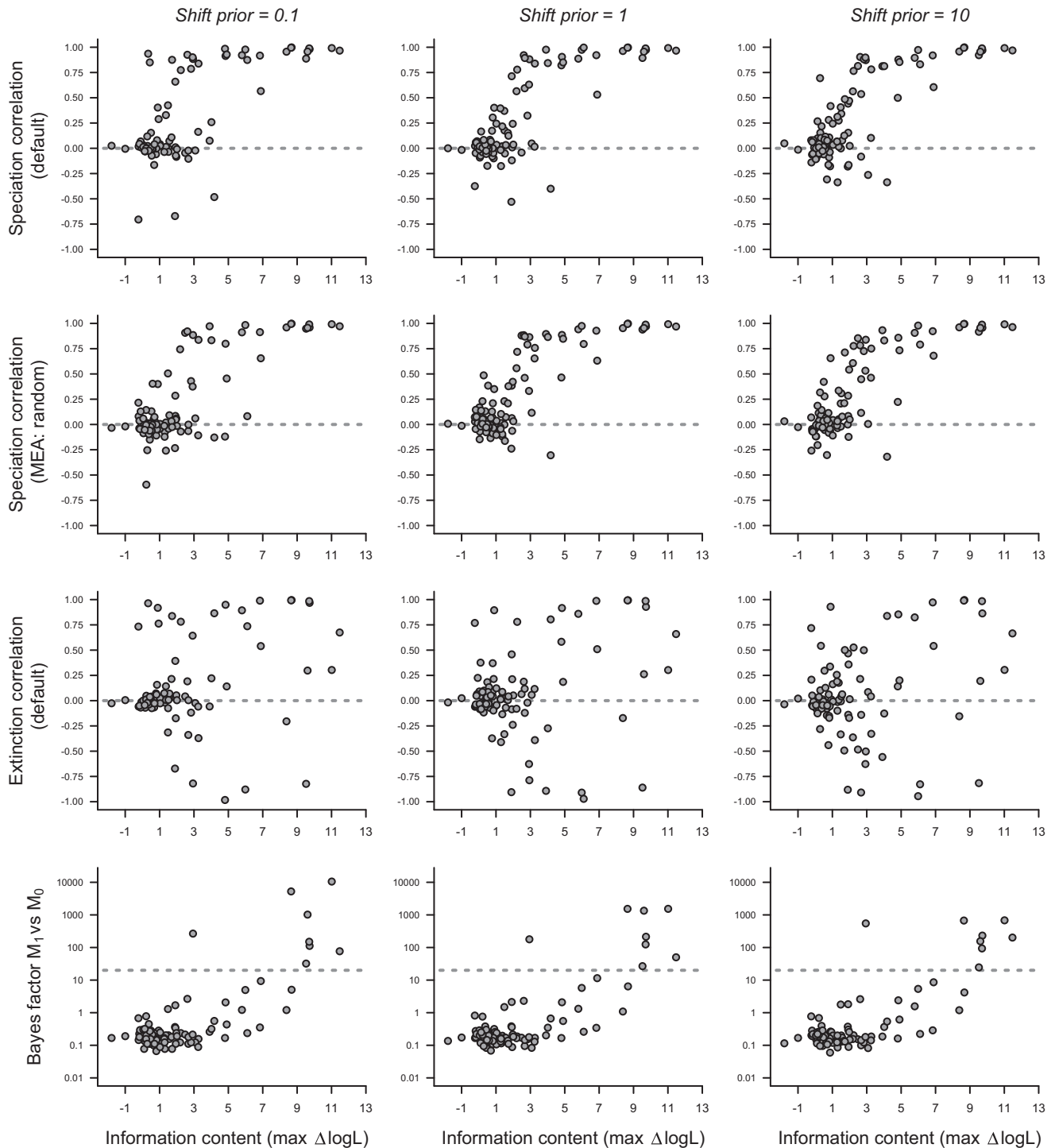
FIGURE 9. Summary attributes of BAMM analysis of all trees in MEA "variable rates" data set analyzed under three priors (columns), plotted by the information content in each tree. Information content for each tree is the maximum Δ LogL value across all shifts occurring in the tree. Each tree is represented by a single point in each plot. Row 1 (from top): Pearson correlations between branch-specific BAMM estimates of speciation rates and the true rates. Speciation rates were estimated as the mean of the marginal posterior distribution for the branch, as in MEA. BAMM-estimated rates are highly correlated with the true rates for most trees with sufficient information to infer rate shifts (max ΔLogL >3). Row 2: correlations between branch-specific BAMM estimates of speciation rates and true rates, using the "random" setting for combineExtinctionAtNodes (as in MEA). Results are virtually identical to those obtained with BAMM defaults (top row). Row 3: correlations between branch-specific extinction rate estimates and true values; extinction estimates are generally poor across this data set, potentially reflecting the low information content of small rate regimes for this parameter. Row 4: Bayes factor evidence for rate heterogeneity as a function of the maximum ΔLogL for each phylogeny. Bayes factors were computed as the evidence favoring a one-shift model (M_1) relative to a zero-shift model (M_0); dotted line indicates "significant" Bayes factor evidence for rate heterogeneity (BF = 20). For trees with maximum ΔLogL <3 (77 of 100 trees), there is generally no evidence for rate heterogeneity, regardless of the prior. Inferences are highly concordant across 100-fold differences in the mean of the prior on the expected number of rate shifts (columns).
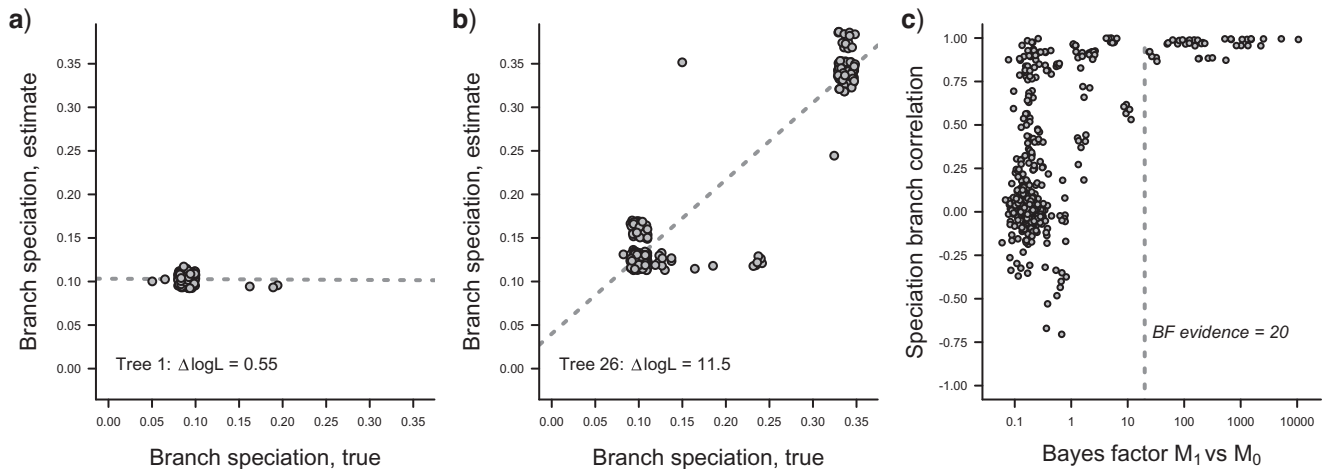
FIGURE 10.   Low statistical power confounds MEA's assessment of BAMM speciation rate reliability. a) Regression analysis of branch-specific speciation rate estimates as a function of the true rates for the first tree in MEA's variable-rates data set, which is characterized by low information content (ΔLogL max $= 0.55$; only 1 or 2 tips per rate shift). Slope of fitted regression line is zero (dotted line). However, BAMM accurately estimates the speciation rate shared by the vast majority of branches: across 160 of 166 branches in the root rate regime, we observe $\lambda_{\mathrm{TRUE}} = 0.09$ versus mean $\lambda_{\mathrm{BAMM}} = 0.10$. Points have been jittered to reduce overplotting. For phylogenies that lack detectable rate variation, the slopes from such regression analyses are expected to equal zero, even when the overall tree-wide rate is estimated with high accuracy b) Comparable analysis for a tree with high information content (tree 26). Slope of fitted regression line is 0.88 and $r^2 = 0.934$. c) Relationship between within-tree correlation in branch-specific speciation rates as a function of the Bayes factor evidence for rate heterogeneity in the tree. Bayes factors were computed as the evidence favoring a one-shift model (M_1) relative to a zero-shift model (M_0); dotted line indicates "significant" Bayes factor evidence for rate heterogeneity (BF = 20).

BAMM infers no rate shifts, branch-specific correlations and regression slopes will be zero (Fig. 10a), regardless of the accuracy of the BAMM estimates. Across all trees and γ priors, the mean Pearson correlation between branch-specific BAMM speciation estimates and true rates was 0.255. However, this value is largely driven by lack of power associated with low ΔLogL trees. When we drop the 77% of all MEA trees where the most-inferable shift had ΔLogL <3 and consider only the remaining 23 informative trees, we find that the mean correlation rises to 0.75. If we restrict this even further to the 14 trees with maximum ΔLogL >5, we find a mean correlation of 0.92. These results are virtually identical when the "random" option is used, as in MEA (2nd row from top). With "random" the mean correlations across all trees with ΔLogL >3 and ΔLogL >5 are 0.75 and 0.93, respectively. Interestingly, speciation rate correlations are both consistent and robust across all prior parameterizations on γ (columns in Fig. 9). For the 14 trees with a maximum ΔLogL >5, the branch-by-branch correlation in speciation rates was effectively independent of the prior: for the "random" option used by MEA, the five prior scenarios ($\gamma = 0.1$, 0.5, 1.0, 2, 10) had mean correlations of 0.88, 0.93, 0.93, 0.93, and 0.93, respectively. Corresponding values with BAMM v2.5 defaults were 0.93, 0.91, 0.93, 0.93, and 0.93. For low-power trees with ΔLogL <3, within-tree speciation correlations are close to zero (mean correlation across all priors and analysis options = 0.111) as expected when BAMM fails to infer rate variation (Fig. 10a). Extinction rates are poorly estimated with trees of this size (Fig. 9, row 3).

The Bayes factor evidence for rate heterogeneity in each of MEA's trees is a function of the information content (Fig. 9, row 4). Regardless of the prior on the number of shifts, only a single tree with max ΔLogL <8 was observed to show significant rate variation using Bayes factors. For the nine trees with max ΔLogL >8, seven had significant Bayes factor evidence for rate variation and this result was consistent across all priors. These results are identical regardless of whether BAMM v2.5 defaults or the MEA "random" settings are used.

Figure 10 illustrates the consequences of MEA's low-power design for the regression analyses presented in their paper. For most of MEA's variable rates phylogenies, there is almost no information with which to infer rate shifts, leading to regression slopes that are approximately zero even when the overall tree-wide rates are estimated accurately (Fig. 10a). However, for trees with sufficient information to infer shifts, BAMM is able to infer branch-specific rates with considerable accuracy (Fig. 10b,c). Indeed, for all phylogenies with significant Bayes factor evidence for rate heterogeneity, the correlation between true and BAMM-estimated branch rates is very high (Fig. 10c).

For each tree in the MEA variable rates data set, we computed the mean proportional accuracy in branch-specific speciation estimates as well as the corresponding regression slopes for branch rates (as in Fig. 10b). We found mean rate accuracy—the ratio of inferred-to-true speciation rates—was close to unity, especially for the eight phylogenies where BAMM detected strong evidence for rate heterogeneity (Fig. 11a). MEA conclude that the average slope for branch-specific estimates of
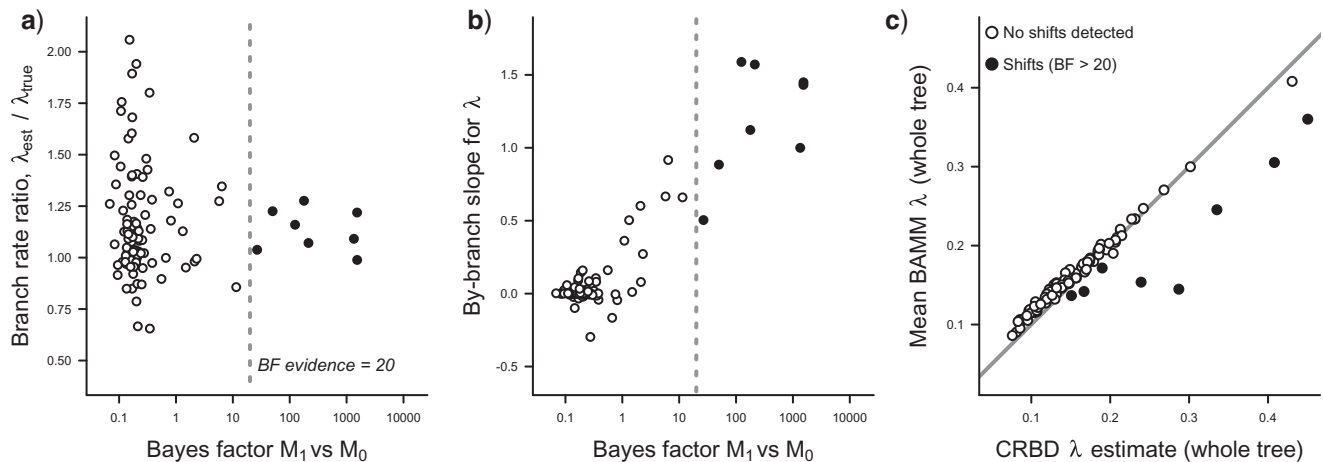
FIGURE 11. Branch-specific estimates of speciation rates obtained with BAMM are more accurate when trees contain evidence for rate variation a) Proportional accuracy of branch-specific speciation rate estimates as a function of the Bayes factor evidence for rate heterogeneity in the tree; each point is the tree-wide mean value. Filled circles denote trees with strong evidence (BF >20) for rate shifts. b) Within-tree regression slopes for branch-specific speciation rates; mean slope for trees where significant heterogeneity was detected (filled circles) is 1.19. c) For trees lacking a strong signal of rate heterogeneity, BAMM estimates of speciation rates are highly correlated with the overall tree-wide rate that would be estimated under a CRBD process. Bayes factors in (a) and (b) were computed as the evidence favoring a one-shift model (M_1) relative to a zero-shift model (M_0); dotted line indicates "significant" Bayes factor evidence for rate heterogeneity (BF = 20). Analyses shown here performed with a prior of $\gamma = 1$ and default BAMM v2.5 settings, but virtually identical results are obtained with other priors and with the "random" setting activated by MEA (Supplementary Figs. S19–S23, available on Dryad).

rates (Fig. 10b) is zero and that estimated rates are uncorrelated with their true values. This result is strictly the result of assessing branch-specific rate accuracy across their set of low-power trees, using regression statistics that would perform poorly if BAMM failed to identify rate shifts (Fig. 10). The mean branch-specific slope was 0.05 for the 92 trees where BAMM did not find strong evidence for rate variation (BF <20), but BAMM performed well at estimating branch-specific rates across the remaining 8 trees (Fig. 11b), where the mean branch slope was 1.19. For the 92 trees where BAMM failed to detect strong evidence for rate variation, the BAMM-estimated mean rates across the entire tree are very highly correlated with values estimated under a simple CRBD process (Fig. 11c, open circles; regression slope = 0.89, $r^2 = 0.987$). A clear picture thus emerges across the MEA "variable rates" data set: when BAMM has little power to infer rate variation, speciation rate estimates are similar to those that would be obtained from a CRBD process (Fig. 11c, open circles). If BAMM has power to detect rate variation, branch-specific variation in rates is inferred with reasonable accuracy (Figs. 9–11).

A similar picture of speciation rate accuracy emerges from considering mean estimates of rates for each shift regime relative to their true values (Fig. 12). Across all rate regimes, the overall correlation between true and BAMM-estimated rates is positive but relatively poor (Fig. 12a). However, when we include only root regimes and shift regimes with ΔLogL >3, we find that BAMM estimates are reasonably correlated with the true values. Across all 4374 branches from the 23 phylogenies with maximum ΔLogL >3, we find a strong correlation between true and BAMM-estimated rates (Fig. 12c). Finally, although branch-specific variation in extinction

rates is poorly estimated (Fig. 9), the whole-tree estimates recovered by BAMM are at least as accurate as one would obtain from a simple CRBD process (Supplementary Figs. S23–S24, available on Dryad).

Summary: Diversification Rate Estimates with BAMM

The overwhelming majority of variable rates phylogenies from MEA are approximately identical to trees generated from a constant-rate birth-death process. There is minimal information available with which BAMM, or any other method, could infer rate heterogeneity (Fig. 8), due to the small number of tips in each rate regime (Fig. 7) and the proportionally small changes in rates for rate shifts. MEA's results are thus a simple consequence of low statistical power: when BAMM does not detect rate variation, the summary statistics used in MEA's assessment are expected to imply poor performance, even when the tree-wide rate is estimated with high accuracy (Fig. 10). Despite the limitations of MEA's simulation design, we find that BAMM does quite well at inferring diversification rates for those regimes that are—in principle—inferable (Figs. 9–12). For sufficiently informative data sets, we find virtually identical results across a broad range of prior parameterizations (Fig. 9, columns) as well as for different analysis modes (e.g., BAMM defaults vs. "random"; time-varying vs. time-constant regimes: Fig. 9; Supplementary Figs. S15 - S24 available on Dryad).

How Reliable are Previous Versions of bamm?

Several aspects of the BAMM calculations differ between BAMM v2.5 and earlier releases of the program
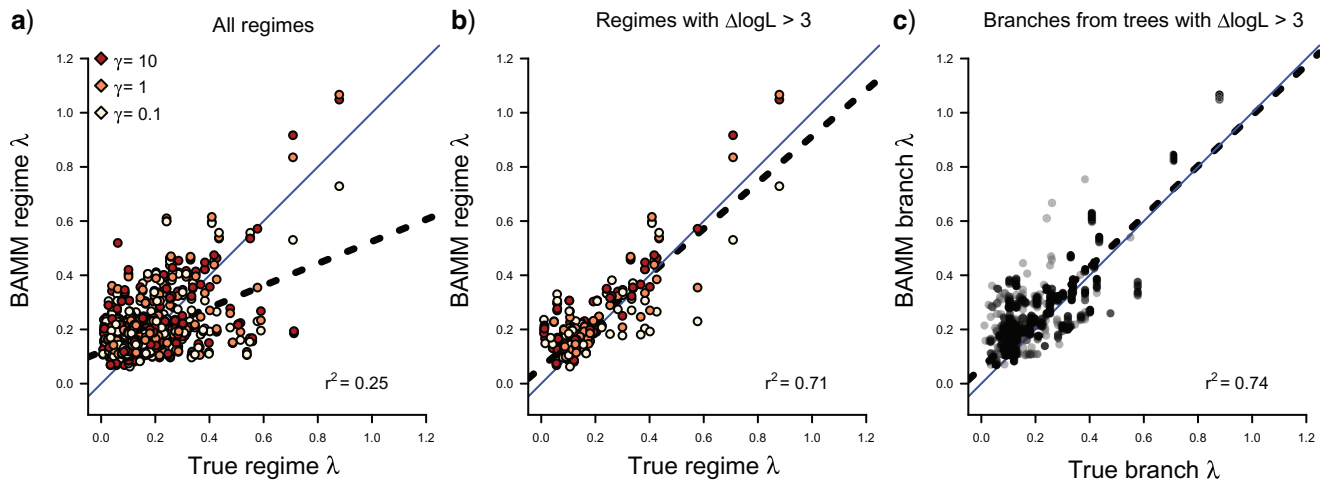
FIGURE 12. a) Relationship between true speciation rates and BAMM-estimated rates across all 535 rate regimes in the MEA variable-rates data set, estimated under three different prior parameterizations (colors; $\gamma = 0.1$, 1, and 10). Dashed line is fitted regression line. Blue line denotes theoretical 1:1 expectation. b) Same as (a), but where all rate regimes with low information content ($\Delta LogL < 3$) are excluded. Slope of fitted regression line is 0.85. c) Plot of all branch-specific estimates of speciation against the true values for all trees where at least one shift had non-trivial information content ($\Delta LogL > 3$) from analyses with $\gamma = 1$ (23 trees; 4374 branches total). Slope of fitted regression line is 0.931; results for other prior parameterizations are virtually identical to those shown in (c). Results obtained using MEA "random" setting; comparable results for BAMM v2.5 default settings shown in Supplementary Figure S19 (available on Dryad).
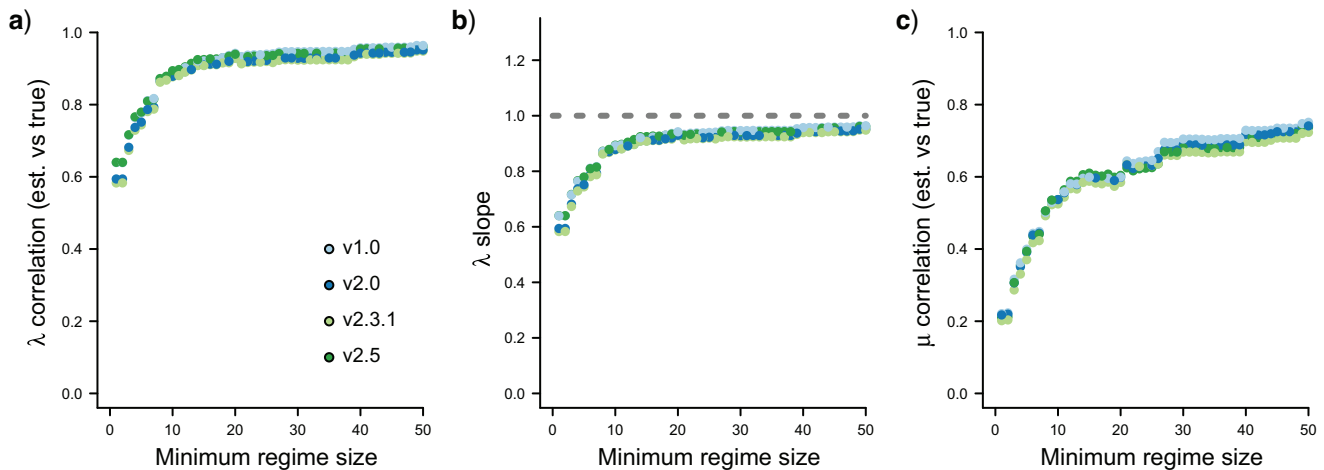

FIGURE 13. BAMM rate estimates are reliable and consistent across all major releases of the software. a) Pearson correlation between true and estimated speciation rates across rate regimes. Results are plotted as a function of minimum regime size; colors denote BAMM version. For example, $x = 10$ corresponds to the correlation computed for the set of regimes with 10 or more tips. The four major releases of BAMM yield virtually identical results. b) Slopes of the linear relationship between true and BAMM-estimated speciation rates across all rate regimes c) Correlations between true and estimated extinction rates across rate regimes. Analyses are based on 200 phylogenies simulated under MEA's general procedure, but minimizing ascertainment biases (Fig. 7c). The primary difference is that our simulations generated greater variation in tree size relative to MEA: although median tree sizes are similar (median size = 129.5 tips vs. 84 tips in MEA), the means are very different (446 tips here vs. 89 tips in MEA). Despite differences in the underlying algorithms used to compute the likelihood, we observe no appreciable differences in performance across these four BAMM versions (see also Supplementary Figs. S25–S30, available on Dryad). Results shown here are based on the most commonly used prior on the number of shifts ($\gamma = 1$). A complete description of these simulations and associated summaries are found in Supplementary Material section S6, available on Dryad.

(Supplementary Material section S2.5, available on Dryad). The initial release of BAMM (Rabosky 2014) was accompanied by extensive performance assessment, but there is yet no published analysis of the extent to which inferences vary across previously released versions of the software. We tested the reliability of inferences about evolutionary rates and the number of rate shifts across four BAMM versions (v1.0, v2.0, v2.3.1,

and v2.5) that capture the most significant changes to the software. Given that MEA's simulated rate-variable phylogenies contain minimal information with which to infer among-lineage rate variation, we simulated a new set of phylogenies under a forward-time Poisson process with rate shifts. We did not impose strong constraints on tree size, to avoid ascertainment biases characteristic of MEA's variable-rates data set (Supplementary Material

section S5.1, available on Dryad). We simulated 200 phylogenies under each of two shift frequencies ($\eta = 0.01$; $\eta = 0.001$) and analyzed each data set with three model priors and four BAMM versions, for 4800 total BAMM analyses. A complete description of these simulations and associated analyses is provided in the Supplementary Material section S6 (available on Dryad).

For the most widely used prior parameterization ($\gamma = 1$), we found virtually no differences in performance across all four BAMM versions (Fig. 13; Supplementary Figs. S25–S30, available on Dryad), despite significant differences in the underlying likelihood calculations. Across all simulation and analysis conditions, speciation rates estimated with BAMM are highly correlated with the true rates (Fig. 13; Supplementary Figs. S25–S28 available on Dryad). Extinction rate estimates are reasonably correlated with the true values under realistic simulation conditions, but as expected, these estimates are much less accurate than those for speciation (Fig. 13c) and extinction estimates are less accurate when rate shifts are common (Supplementary Figs. S25–S28, available on Dryad). We find no evidence that the inferred number of shifts differs appreciably across BAMM versions, at least for the most commonly-used prior parameterizations (Supplementary Figs. S29–S30, available on Dryad). In all metrics we have considered, BAMM v2.5 yields rate estimates that are more accurate than previous versions of BAMM, including a version (v2.3.1) where the likelihood was computed using the same algorithm used by MEA (albeit with $\eta = 0$). These results demonstrate that, in general, results obtained with earlier versions of BAMM are expected to be similar across all major releases of the software, despite numerous differences between these versions. Moreover, this assessment conclusively rejects MEA's claim that BAMM estimates of diversification rate parameters are unreliable. Such a conclusion is only possible when a select set of performance metrics is applied to phylogenies for which it is not theoretically possible to infer rate variation.

CONCLUSIONS

Moore et al. (2016) concluded that BAMM cannot correctly estimate diversification rates, and that the posterior distribution of rate shifts shows extreme prior sensitivity. They attribute these statistical pathologies to an incorrect likelihood function and to a problematic CPP prior model on the number of rate shift events. Our reassessment of the theory and results that underlie their article demonstrate that these conclusions are not justified. First, MEA claim that BAMM's likelihood function is incorrect on theoretical grounds, as it ignores the effects of unobserved rate shifts. We have shown that their assessment was compromised by their use of an invalid likelihood function that, by allowing probabilities to exceed unity, violates the key axiom of probability theory. In addition, numerical discrepancies between BAMM and MEA

likelihoods (MEA Figs. 2 and 3) are a function of multiple algorithmic differences that are not related to unobserved rate shifts (Supplementary Material section S2.4, available on Dryad). Second, we have demonstrated that unobserved rate shifts are unlikely to have relevance in biologically plausible regions of parameter space. Third, we found no evidence that the prior model in BAMM leads to inference problems, regardless of BAMM version or usage mode. BAMM is generally conservative, especially under the most widely used prior parameterizations and/or when proper Bayesian model selection is performed. Fourth, BAMM performs well across all regions of parameter space we have explored thus far. MEA's conclusion that BAMM rate estimates are unreliable is based on a set of low-power analyses for which it would not have been possible for BAMM, or any other method, to infer diversification rate variation. Finally, our assessment of a much richer set of simulated data sets reveals that BAMM infers diversification rates—particularly speciation rates—with high accuracy. We demonstrate that BAMM's performance has been remarkably consistent across all major releases of the software.

We appreciate MEA's clear theoretical discussion of rate shifts on extinct/unobserved lineages, but we question the relevance and practicality of accommodating these fundamentally unknown quantities into empirical inferences. The computationally intensive approach used by MEA is influenced by prior assumptions regarding the rate distributions for unobserved rate shifts, a quantity that—by definition—is not observed. The MEA approach is thus integrating over unknown unknowns and at great computational cost. Moreover, if the marginal effects of unobserved rate shifts on extinction probabilities (Figs. 3 and 4a) lead to biased inference, then surely the arbitrary assumptions about the distributions of rates from which unobserved shift regimes are drawn must also matter.

Leaving these theoretical concerns aside, we find no evidence that accommodating rate shifts on extinct lineages actually contributes to our ability to model phylogenetic data with rate-shift models (Figs. 3 and 4). Under the exact parameters used in MEA's variable-rate simulations, we find that the Monte Carlo likelihoods for rate-shift regimes are virtually identical to those obtained from the orders-of-magnitude faster analytical approximation that ignores unobserved shifts (Fig. 8b). By focusing on largely irrelevant quantities that cannot be accommodated by any modeling framework, we are concerned that MEA's forceful prohibitions serve as a distraction from many important methodological issues in macroevolutionary inference. These include the development of methods for the analysis of combined paleontological and molecular phylogenetic data sets (Didier et al. 2012), for accommodating protracted or ephemeral speciation processes (Etienne and Rosindell 2012; Rosenblum et al. 2012; Dynesius and Jansson 2014), and for teasing apart the relative importance of diversity-dependent and diversity-independent factors

that regulate clade dynamics (Etienne and Haegeman 2012; Rabosky 2013; Manceau et al. 2015; Silvestro et al. 2015).

Macroevolutionary modeling is a dynamic discipline with a rapidly growing data landscape involving phylogenies, fossils, phenotypes, and ecologies. We believe that there is tremendous scope for improvements to existing methods and also that the field offers great opportunities for developing new methods with which to address novel conceptual challenges. Innovating at the frontiers of science has frequently involved building upon valid insights from imperfect tools. We remain optimistic that the imperfect and approximate inference tools currently at our disposal have revealed and will continue to reveal fundamental insights into the processes that generate and maintain biological diversity.

APPENDIX

Formal Analysis of Incomplete Augmentation in MEA

Here, we demonstrate that the data augmentation strategy used by MEA can yield probabilities that exceed unity. Consider the single lineage history illustrated in Fig. 1a, which has a probability that we denote by $P(x_0, x_1|n_T > 0)$. This lineage history begins at time t_0 with rates (λ_0, μ_0), undergoes a rate shift at time t_s, and survives to the present with parameters (λ_1, μ_1). No other lineages are observed, although the process may have produced daughter lineages that failed to survive to time T. Given that the initial probability of the data is unity, let $\lambda_1 = 0$ and $\mu_1 = 0$. Thus, the probability of the segment $(t_s, T) = P(x_1|\lambda_1 = 0, \mu_1 = 0) = 1$, as no events can occur that can change the probability of the data. Letting $\eta = 0$, we will condition the probability expression in Equation (5) on zero unobserved rate shifts to obtain

$$P(x_0|\lambda_0, \mu_0, x_1)[1 - E_0(t_0)]^{-1}\phi^{-1} \qquad (A.1)$$

where ϕ is the probability that there are no unobserved (extinct) rate shifts in the process that gives rise to the lineage history in Figure 2a.

To show that this expression is invalid, we will demonstrate that, as the time interval $\Delta t = t_s - t_0$ becomes very small, the probability $P(x_0|\lambda_0, \mu_0, x_1)$ can become arbitrarily close to 1, given that the probability $P(x_1|\lambda_1, \mu_1)$ is necessarily 1. However, the probability of extinction $E(t)$ is independent of the duration of Δt and depends only on the parameters at the root of the tree and the total age of the process (T). Hence, the survival probability $(1 - E_0(t_0))$ can approach 0, leading to probabilities for the complete process that exceed 1.

The analytical probability of the data at the rootwards end of the segment x_0 is taken from the CRBD process with arbitrary starting conditions and is given by the following expression

$$P(x_0) = \frac{e^{(\mu_0 - \lambda_0)(t_s - t_0)}P(x_1)(\lambda_0 - \mu_0)^2}{\left[\lambda_0 - \lambda_0 E_0(t_s) + e^{(\mu_0 - \lambda_0)(t_s - t_0)}(\lambda_0 E_0(t_s) - \mu_0)\right]^2}$$
$$(A.2)$$

which is equivalent to Rabosky (2014) under the substitutions $\Delta t = t_s - t_0$ and $D_0 = P(x_1)$, and which is also identical to FitzJohn et al. (2009) under the substitution $f = 1 - E_0$. Letting $P(x_1) = 1$ and taking the limit as t_s approaches t_0, we have

$$\lim_{t_s \to t_0} P(x_0) = \frac{(\lambda_0 - \mu_0)^2}{\left[\lambda_0 - \lambda_0 E_0(t_0) + \lambda_0 E_0(t_0) - \mu_0\right]^2} = 1 \quad (A.3)$$

Thus, as the time interval $t_s - t_0$ shrinks to zero, we obtain $P(x_0) = 1$, which is intuitively obvious: the initial probability $P(x_1)$ is 1, and the time interval becomes sufficiently short that it is unlikely that any events will occur to change this probability appreciably.

Note that we can make the term ϕ arbitrarily close to 1 by simply assuming that either the generating process has a very low η, or alternatively, that the prior mean on the extinction probability is very low. If rate shifts occur and the new extinction rate is near zero, the shifts will generally survive to the present to be observed, regardless of η. Hence, we have

$$P(x_0, x_1|n_T > 0) \approx [1 - E_0(t_0)]^{-1} \qquad (A.4)$$

and $E_0(t_0)$ depends only on the initial parameters (λ_0, μ_0) and the total age of the process, $T - t_0$. This is true even when the process undergoes an immediate shift to a new rate regime (e.g., t_s is close to t_0; Δt close to 0). Hence, $E_0(t)$ can approach unity and the corresponding probability $P(x_0, x_1|n_T > 0)$ is unbounded at 1 and can become infinite.

REFERENCES

Alfaro M.E., Holder M.T. 2006. The posterior and the prior in Bayesian phylogenetics. Ann. Rev. Ecol. Evol. Syst. 37:19–42.

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. U.S.A. 106:13410–13414.

Bailey N.T.J. 1964, The elements of stochastic processes with applications to the natural sciences. New York: Wiley.

Barraclough T.G., Vogler A.P. 2002. Recent diversification rates in North American tiger beetles estimated from a dated mtDNA phylogenetic tree. Mol. Biol. Evol. 19:1706–1716.

Beaulieu J.M., O'Meara B.C. 2015. Extinction can be estimated from moderately sized molecular phylogenies. Evolution 69:1036–1043.

Beaulieu J.M., O'Meara B.C. 2016. Detecting hidden diversification shifts in models of trait-dependent speciation and extinction. Syst. Biol. 65:583–601.

Brown J.M., Hedtke S.M., Lemmon A.R., Lemmon E.M. 2009. When trees grow too long: investigating the causes of highly inaccurate Bayesian branch-length estimates. Syst. Biol. 59:145–161.

Davis M.P., Midford P.E., Maddison W.P. 2013. Exploring power and parameter estimation of the BiSSE method for analyzing species diversification. BMC Evol. Biol. 13:38.

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

Dynesius M., Jansson R. 2014. Persistence of within-species lineages: a neglected control of speciation rates. Evolution 68:923–934.

Etienne R.S., Haegeman B. 2012. A conceptual and statistical framework for adaptive radiations with a key role for diversity dependence. Am. Nat. 180:E75–E89.

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.

Etienne R.S., Rosindell J. 2012. Prolonging the past counteracts the pull of the present: protracted speciation can explain observed slowdowns in diversification. Syst. Biol. 61:204–213.

FitzJohn R. 2010. Quantitative traits and diversification. Syst. Biol. 59:619–633.

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

FitzJohn R.G. 2012. Diversitree: comparative phylogenetic analyses of diversification in R. Methods in Ecology and Evolution 3: 1084–1092.

Heath T.A., Moore B.R. 2014. Bayesian inference of species divergence times. In Chen M.-H., Kuo L., Lewis P.O., editors. Bayesian phylogenetics: methods, algorithms, and applications. Boca Raton, Florida: Chapman & Hall/CRC.

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

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

Lewitus E., Morlon H. 2016. Characterizing and comparing phylogenies from their Laplacian spectrum. Syst. Biol. 65:495–507.

Maddison W.P., Midford P.E., Otto S.P. 2007. Estimating a binary character's effect on speciation and extinction. Syst. Biol. 56:701–710.

May M.R., Hohna S., Moore B.R. 2016. A Bayesian approach for detecting the impact of mass-extinction events on molecular phylogenies when rates of lineage diversification vary. Methods Ecol. Evol. 7:947–959.

Mitchell J.S., Rabosky D.L. 2016. Bayesian model selection with BAMM: effects of the model prior on the inferred number of diversification shifts. Methods Ecol. Evol. doi: 10.1111/2041-210X.12626.

Mooers A.O., Heard S.B. 1997. Evolutionary process from phylogenetic tree shape. Q. Rev. Biol. 72:31–54.

Moore B.R., Hohna S., May M.R., Rannala B., Huelsenbeck J.P. 2016. Critically evaluating the theory and performance of Bayesian analysis of macroevolutionary mixtures. Proc. Natl. Acad. Sci. U.S.A. 113:9569–9574.

Manceau M., Lambert A., Morlon H. 2015. Phylogenies support out-of-equilibrium models of biodiversity. Ecol. Lett. 18:347–356.

Morlon H. 2014. Phylogenetic approaches for studying diversification. Ecol. Lett. 17:508–525.

Morlon H., Lewitus E., Condamine F.L., Manceau M., Clavel J., Drury J. 2016. RPANDA: an R package for macroevolutionary analyses on phylogenetic trees. Methods Ecol. Evol. 7:589–597.

Morlon H., Parsons T.L., Plotkin J.B. 2011. Reconciling molecular phylogenies with the fossil record. Proc. Natl. Acad. Sci. U.S.A. 108:16327–16332.

Morlon H., Potts M.D., Plotkin J.B. 2010. Inferring the dynamics of diversification: a coalescent approach. Plos Biol. 8:e1000493.

Nee S., May R.M., Harvey P.H. 1994. The reconstructed evolutionary process. Philos. Trans. R Soc. Lond. B Biol. Sci. 344:305–311.

O'Meara B.C. 2012. Evolutionary inferences from phylogenies: a review of methods. Ann. Rev. Ecol. Evol. Syst. 43:267–285.

Phillimore A.B., Price T.D. 2008. Density dependent cladogenesis in birds. Plos Biol. 6:e71.

Pickett K.M., Randle C.P. 2005. Strange Bayes indeed: uniform topological priors imply non-uniform clade priors. Mol. Phylo. Evol. 34:203–211.

Pybus O.G., Harvey P.H. 2000. Testing macro-evolutionary models using incomplete molecular phylogenies. Proc. R. Soc. Lond. B 267:2267–2272.

Rabosky D.L. 2013. Diversity dependence, ecological speciation, and the role of competition in macroevolution. Ann. Rev. Ecol. Evol. Syst. 44:481–502.

Rabosky D.L. 2014. Automatic detection of key innovations, rate shifts, and diversity-dependence on phylogenetic trees. PLoS ONE 9:e89543.

Rabosky D.L. 2015. Challenges in the estimation of extinction from moelcular phylogenies: a response to Beaulieu and O'Meara. Evolution. doi:10.1111/evo.12820.

Rabosky D.L., Donnellan S.C., Grundler M., Lovette I.J. 2014a. Analysis and visualization of complex macroevolutionary dynamics: an example from australian scincid lizards. Syst. Biol. 63: 610–627.

Rabosky D.L., Grundler M., Anderson C., Title P., Shi J., Huang H., Brown J.W. et al. 2014b. BAMMtools: an R package for the analysis of evolutionary dynamics on phylogenetic trees. Methods Ecol. Evol. 5:701–707.

Rabosky D.L., Lovette I.J. 2008. Explosive evolutionary radiations: decreasing speciation or increasing extinction through time? Evolution 62:1866–1875.

Rabosky D.L., Santini F., Eastman J.M., Smith S.A., Sidlauskas B., Chang J., Alfaro M.E. 2013. Rates of speciation and morphological evolution are correlated across the largest vertebrate radiation. Nat. Commun. 4. doi: 10.1038/ncomms2958.

Raup D.M. 1985. Mathematical models of cladogenesis. Paleobiology 11:42–52.

Ricklefs R.E. 2007. Estimating diversification rates from phylogenetic information. Trends Ecol. Evol. 22:601–610.

Rosenblum E.R., Sarver B.A., Brown J.W., Des Roches S., Hardwick K.M., Hether T.D., Eastman J.M., Pennell M.W., Harmon L.J. 2012. Goldilocks meets Santa Rosalia: an ephemeral speciation model explains patterns of diversification across time scales. Evol. Biol. 39: 255–261.

Silvestro D., Antonelli A., Salamin N., Quental T.B. 2015. The role of clade competition in the diversification of North American canids. Proc. Natl. Acad. Sci. U.S.A. 28:8684–8689.

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.

Stadler T. 2011. Mammalian phylogeny reveals recent diversification rate shifts. Proc. Natl. Acad. Sci. U.S.A. 108:6187–6192.

Stadler T. 2013. Recovering speciation and extinction dynamics based on phylogenies. J. Evol. Biol. 26:1203–1219.