

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

## Supplementary Online Material

Daniel L. Rabosky, Jonathan S. Mitchell, Jonathan Chang

February 13, 2017

### Contents

|          |   |           |
|----------|---|-----------|
| <b>1</b> | <b>Repeating analyses from MEA</b>  | <b>2</b>  |
| <b>2</b> | <b>Likelihoods and Rate-Shift Models: Outstanding Issues</b>                        | <b>3</b>  |
| 2.1      | Root Conditioning Used in MEA . . . . .   | 3         |
| 2.2      | MEA-Type Conditioning Fails for the Cetacean Dataset . . . . .                      | 3         |
| 2.3      | Comment on the “combineExtinctionAtNodes” setting used by MEA . . . . .             | 5         |
| 2.4      | Theoretical overview of the BAMM v2.5 likelihood . . . . .                          | 8         |
| 2.5      | History of the likelihood as implemented in BAMM . . . . .                          | 11        |
| <b>3</b> | <b>Rate-Shift Frequencies</b>   | <b>13</b> |
| 3.1      | BAMM Estimates of the Shift Rate are Correlated with the True Rate . . . . .        | 13        |
| 3.2      | Empirical Transition Rates for MEA Datasets . . . . .                               | 13        |
| <b>4</b> | <b>Extinction Probabilities and Unobserved Rate Shifts</b>                          | <b>16</b> |
| <b>5</b> | <b>Reanalysis of MEA’s “variable rates” dataset</b>                                 | <b>20</b> |
| 5.1      | Ascertainment biases in MEA’s simulation study . . . . .                            | 20        |
| 5.2      | Information Content of MEA Variable-Rate Phylogenies . . . . .                      | 22        |
| 5.3      | Analysis of MEA Trees with Time-Varying BAMM . . . . .                              | 23        |
| 5.4      | Robustness of Results to Priors and CombineExtinctionAtNodes settings . . . . .     | 27        |
| <b>6</b> | <b>Reliability of diversification analyses under BAMM v2.5 and earlier versions</b> | <b>31</b> |
| 6.1      | Simulation procedure . . . . .  | 31        |
| 6.2      | Results summary . . . . .   | 32        |

# 1 Repeating analyses from MEA

The directory `RerunningMEA` contains shell scripts to repeat all analyses as performed by MEA and with our modifications to their BMM settings. The MEA data files (trees, BMM control files) can be obtained from Dryad <http://datadryad.org/resource/doi:10.5061/dryad.mb0sd>. All BMM control files used by MEA end with the following line:

---

```
combineExtinctionAtNodes = random
```

---

which invokes a non-default option for BMM likelihood calculations. We consider the effects of this option in more detail below. In the shell scripts provided with this submission, output files from all BMM runs with MEA settings exactly as they provided them are prefixed with `MEA_random`. For comparison, we repeated all analyses from MEA after deleting the line specifying their `random` setting, thus restoring this setting to its default value (`if_different`). Output files from these BMM runs are prefixed with `BMM_defaults`. Finally, we repeated analyses with *time-constant* BMM, constraining the BMM model such that diversification rates within rate regimes are constant through time. This simply constrains the `lambdaShift` parameter to equal zero, such that speciation rates are not modeled with an exponential time-dependent process. Use of time-constant BMM was important for assessing MEA’s “variable rates” dataset, to ensure that the BMM analysis model exactly matched the generating model used by MEA; we also present analyses from time-varying BMM for this dataset, exactly as in MEA. To invoke time-constant BMM, we made the following modifications to MEA’s control files:

---

```
lambdaIsTimeVariablePrior = 0  
updateRateLambdaShift = 0  
updateRateLambdaTimeMode = 0  
lambdaShift0 = 0  
segLength = 1
```

---

**Note on the  $\gamma$  parameter in MEA:** In BMM, the *expected* number of rate shift (e.g., the prior) is specified with one of two settings: `expectedNumberOfShifts` and `poissonRatePrior`. The expected number of shifts is simply the inverse of the Poisson rate prior, so users can specify either of these options. In Rabosky (2014) and subsequent treatments of BMM, we have used the symbol  $\gamma$  (“gamma”) to denote the expected number of shifts. MEA use this parameter inconsistently with respect to previous usage in both their Dryad files and their manuscript. For example, all of their files prefixed with the title “gamma\_0.1” actually contain control files that specify an expected number of shifts of 10, or  $\gamma = 10$ . In our manuscript, the parameter  $\gamma$  always refers to the expected number of shifts. This is equivalent to the inverse of the rate hyperprior on the exponential distribution of Poisson rates that governs the compound Poisson process used by BMM. All of our discussion of MEA’s input files uses the interpretation of  $\gamma$  that is consistent with Rabosky 2014 (e.g.,  $\gamma$  is identical to  $1 / \text{poissonRatePrior}$ ).

## 2 Likelihoods and Rate-Shift Models: Outstanding Issues

### 2.1 Root Conditioning Used in MEA

In MEA, the likelihood is conditioned as described in the main text. Specifically, the probability of the tree is conditioned on the survival of the two crown lineages, given their speciation-extinction parameterization at the root of the tree. MEA do not correctly describe the calculation of the likelihood under their model, because their description ignores conditioning at the root (see eqn S11 from MEA’s Appendix, for example). The code distributed with MEA’s Dryad data package demonstrates that they have computed the extinction probability at the root without conditioning on the data augmentation. Their conditioning is found on lines 120-122 of their R file `likelihoodModel.R` that accompanies their Dryad submission. This function demonstrates that their root extinction probabilities are computed using only the current values of the process and that all downstream data augmentation is ignored for conditioning. From their likelihood function, `MonteCarloLikelihood(...)`, we see:

---

```
node_process <-  
  processes[getProcessForBranchAtTime(this_node, this_node,  
                                       start_time, processes, tree), ]  
log_likelihood <- log(new_d / (1 -  
  getExtinctionProbability(node_process, start_time))^2)
```

---

Here, `node_process` is the set of rate parameters at the root, or  $\Phi_R$ , and `new_d` is the unconditioned likelihood of the phylogeny given the data augmentation. This likelihood is divided by the probability that both crown lineages survive. Under an incorrect data augmentation scheme that considers only  $E(t)$  as a function of  $\Phi_R$ , the crown clade survives with probability

$$P_{survive} = (1 - E_R(t_0))^2 \tag{1}$$

where  $E_R(t_0)$  is the probability that a single lineage at the root of the tree with parameters  $\Phi_R$  goes extinct before the present, along with all of its descendant lineages. In MEA’s likelihood calculation, the call to `getExtinctionProbability` at the root (line 122) returns  $E_R(t_0)$ . This operation appears (but is not) mathematically identical to the conditioning in BAMM, which (from BAMM’s `SpExModel.cpp`) is given by

---

```
logLikelihood -= std::log(1.0 - E0);
```

---

which is performed on both crown lineages at the root, and where `E0` is the extinction probability for the corresponding lineage. However, as described in the main text (Figure 2), the BAMM extinction probabilities  $E(t)$  account for the downstream data augmentation (e.g., the shifts placed on the tree) and can thus be used immediately for conditioning at the root. The MEA extinction probabilities at the root are a strict function of the parameters at the root and lead to data probabilities that are unbounded on  $(0, 1)$ .

### 2.2 MEA-Type Conditioning Fails for the Cetacean Dataset

It is straightforward to construct rate-shift configurations that yield problematic likelihoods under the MEA likelihood calculator. Following the conditioning logic outlined in the main text, we

expect that the likelihood can become arbitrarily large under the following scenario:

- The parameters of the root process,  $\Phi_R$ , imply a high probability of crown clade extinction
- The data augmentation includes rate shifts near the root of the phylogeny that render extinction of the process unlikely

This set of conditions is not intended as an exhaustive description of the parameter space that will yield biased inference with MEA's likelihood function; here, we are simply concerned with demonstrating that MEA's likelihoods can become *infinite*. It is likely that the incorrect conditioning in their equation results in other biases, the exploration of which is beyond the scope of the present article. As an example, we will consider the cetacean dataset distributed with BAMM and BAMMtools. Consider a scenario where each basal branch undergoes a rate shift at  $t = 0.5$  time units after the crown divergence. Assume that both rate shifts have identical parameters, with  $\lambda = 0.13$  and  $\mu = 0.005$ ; these parameters are similar to the whole-tree estimates of  $\lambda$  and  $\mu$  for this dataset under a constant-rate birth-death process. Let the root speciation rate  $\lambda_R$  equal the speciation rates for the "shift" lineages ( $\lambda = 0.13$ ). The figure below shows the likelihood of the cetacean phylogeny as a function of the extinction rate at the root,  $\mu_R$ , holding all other parameters constant. The likelihood increases towards infinity as the root extinction rate increases; the calculations ultimately fail (yielding infinite or NaN values) when the extinction probability at the root becomes numerically indistinguishable from 1. Without a mathematically valid conditioning strategy, the MEA likelihood is unsuitable for inference.

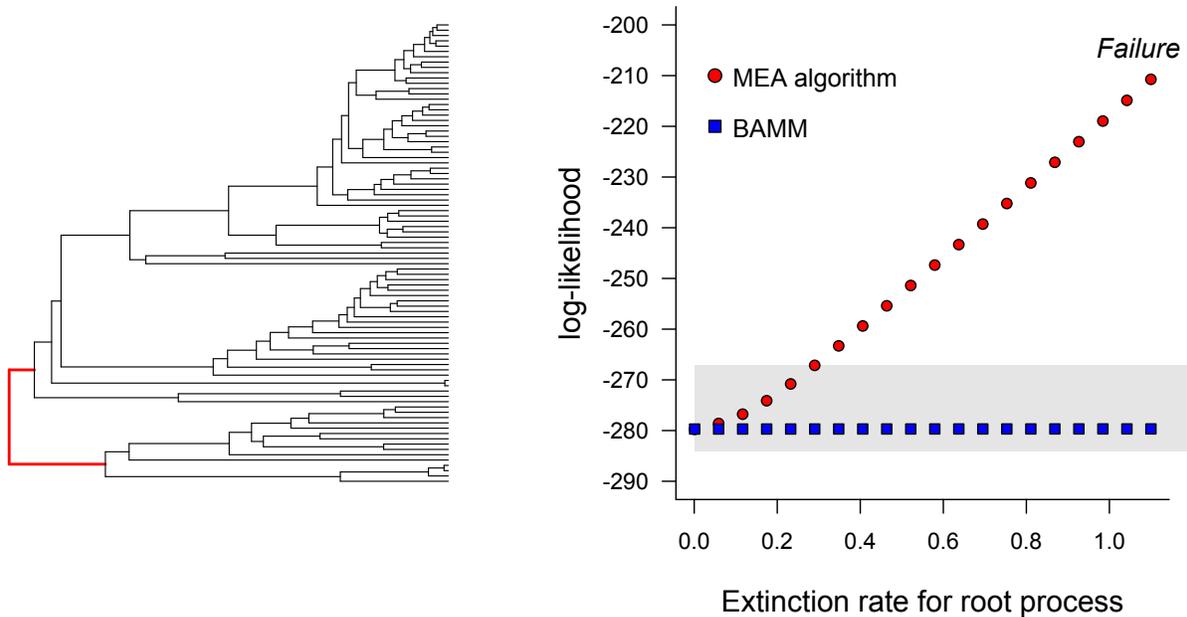


Figure S1: (Left) Phylogeny of cetaceans, highlighting the two basal branches (red) on which rate shifts have been placed. (Right) Likelihood of this shift configuration as a function of the extinction rate for the root rate regime,  $\mu_R$ , under the MEA algorithm (red) and BAMB (blue). Gray polygon denotes the range of log-likelihoods sampled using MCMC for this dataset with BAMB. Under MEA’s algorithm for computing the likelihood, the extinction probability at the root increases towards 1.0 as  $\mu_R$  increases; this leads to a corresponding inflation of the likelihood. In principle, the MEA conditioning strategy will yield infinite likelihoods for some shift configurations, but calculations typically fail before this point. “Failure” denotes the point where further increases in  $\mu_R$  resulted in numerical failure of the calculations; the failure point occurs when  $E(t)$  becomes numerically indistinguishable from 1.

### 2.3 Comment on the “combineExtinctionAtNodes” setting used by MEA

All analyses presented in MEA were obtained using a non-standard and developer-only option in BAMB v2.5. Specifically, MEA added the undocumented command

---

```
combineExtinctionAtNodes = random
```

---

to all of their BAMB control files, which can be viewed at <http://datadryad.org/resource/doi:10.5061/dryad.mb0sd>. Use of this option was not explained in their article, nor did MEA compare their results to those that would have been obtained with BAMB default settings. We repeated all BAMB analyses exactly as presented in their article (with the “random” setting), using their exact control files. We also performed a second set of analyses after removing the “random” line referenced above from their control files, thus restoring the BAMB v2.5 default settings.

The “random” option used by MEA is an alternative algorithm for computing the likelihood that differs from the default implementation in BAMB v2.5; the default option is explained in the main text (e.g., Fig. 2). Under “random”, the likelihood is unstable, because a given speciation-extinction parameterization is associated with multiple likelihood values.

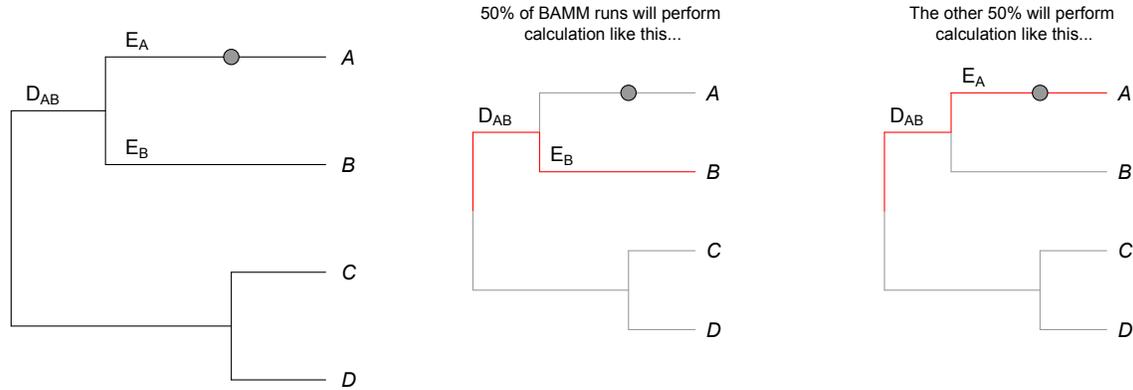


Figure S2: Likelihood calculation under the “random” option used by MEA for a tree with a single rate shift (on terminal branch leading to lineage A). With the “random” option, the likelihood of the branch segment ancestral to lineages A and B, or  $D_{AB}$ , is computed unpredictably in one of two ways: (1) the likelihood  $D_{AB}$  is computed with an extinction probability initialized to the value computed at the end of branch B,  $E_B$ ; or (2)  $D_{AB}$  is computed with an extinction probability initialized to the value computed at the end of branch A,  $E_A$ . These resolutions will yield different likelihoods; the resolution to be used is determined arbitrarily when a phylogeny is loaded into computer memory for a BAMM analysis. Thus, calculation of the likelihood with “random” is unstable: successive computations of the likelihood with identical parameters are not guaranteed to give identical values.

To illustrate the effects of the “random” option for a simple 4-taxon phylogeny, we will consider the tree  $((A:9, B:9):1, (C:9, D:9):1)$ . We will assume that lineages A and D have undergone rate shifts immediately after their origin, to “inert” parameters ( $\lambda = 0, \mu = 0$ ). We will also assume that the initial speciation rate at the root is  $\lambda_R = 1$ . Let  $D_{AB}$  and  $D_{CD}$  denote the likelihoods for branch segments ancestral to nodes AB and CD, and let  $E_X$  denote the extinction probability computed at the rootwards end of the branch segment leading to lineage X (e.g.,  $E_A$ ). Given that there are two rate shifts on this phylogeny (on lineages A and D), there are four possible sequences of calculations that can occur to compute the likelihood under the “random” option (and BAMM would choose among these at random):

1.  $D_{AB}$  initialized with  $E_A$ , and  $D_{CD}$  initialized with  $E_C$
2.  $D_{AB}$  initialized with  $E_B$ , and  $D_{CD}$  initialized with  $E_C$
3.  $D_{AB}$  initialized with  $E_A$ , and  $D_{CD}$  initialized with  $E_D$
4.  $D_{AB}$  initialized with  $E_B$ , and  $D_{CD}$  initialized with  $E_D$

Because parameters for the shift regimes on lineages A and D are identical, cases (1) and (4) will give identical likelihoods. There are 3 possible likelihoods we can obtain for any value of  $\mu_0$ . The following figure illustrates the likelihood surface for the 4-taxon tree as a function of  $\mu_0$ :

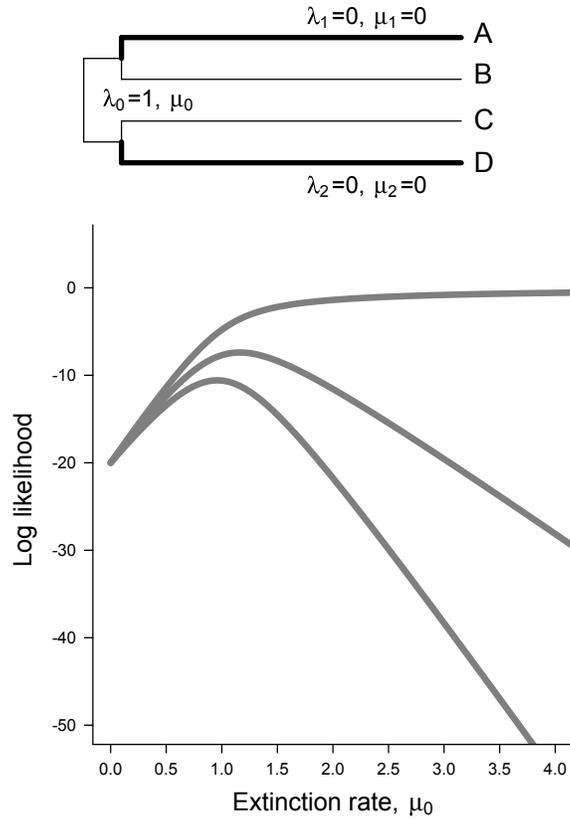


Figure S3: The “random” option used by MEA is mathematically unstable: when rate shifts are present, any speciation-extinction parameterization is associated with multiple likelihood values. Likelihood curves above represent the likelihood surface for the focal 4-taxon phylogeny with two rate shifts (at base of lineages A and D). The profile likelihood for  $\mu_0$  is a composite, because the “random” option associates multiple likelihoods with a given set of parameters. The specific value returned by any BAMM calculation using “random” is arbitrary and will potentially change every time a phylogeny is loaded into computer memory.

Due to the mathematical instability associated with “random”, we cannot recommend the use of this setting with the BAMM software. However, we present results in our article using both the MEA “random” setting as well as BAMM v2.5 defaults.

## 2.4 Theoretical overview of the BAMM v2.5 likelihood

The focus of the present article has been on BAMM v2.5, as this is the version that was used in MEA. This section will clarify the theoretical implications of the calculations as performed in BAMM v2.5. In the main text, we stated that BAMM v2.5 effectively conditions the likelihood calculations on the existence of the subtree leading to rate shifts (Figure 2, main text). The key features of the BAMM likelihood in v2.5 are that (1) the extinction probabilities  $E(t)$  are passed rootwards up the tree (*pass-up algorithm*), and (2) they are multiplied at internal nodes when the descendant lineages differ in their downstream shift histories (that is, when one or both descendants of a node - or any of their descendants - has a mapped rate shift). We consider these calculations to be an approximation of the true likelihood, and our conclusion that the approximation is useful follows from our analysis of performance in practice.

It should be stressed that the  $E(t)$  probabilities in the likelihood equation are merely a component of the likelihood and, considered on their own, imply nothing about whether given phylogeny was shaped by a high or low extinction process. In other words,  $E(t)$  values might differ between BAMM v2.5, MEA, and earlier versions of BAMM, yet all might yield similar inferences about evolutionary rates on an observed phylogeny. BAMM v2.5 generally yields small values of  $E(t)$  at the root of the phylogeny, but this does not mathematically imply that extinction rates were low at the root. Inferred extinction rates for the root process can be high even if the root  $E(t)$  value is very small.

Imagine that we have mapped a set of one or more rate shifts to a phylogeny (the data augmentation), and denote the portion of the phylogeny between all rate shifts and the root as the *backbone* of the phylogeny. The BAMM v2.5 calculations assume that the generating stochastic process can only yield the observed backbone. We will start by considering the state-space logic that underlies the probability calculations for the constant-rate birth-death process. As clearly illustrated in Maddison et al's (2007) derivation of the BiSSE model, the likelihood of a given phylogenetic branching pattern is computed by enumerating the set of events that can occur on a given interval of time  $\Delta t$  while yielding a clade like the observed clade. The following figure illustrates this set of events for a birth-death process where no rate shifts have been placed on the phylogeny:

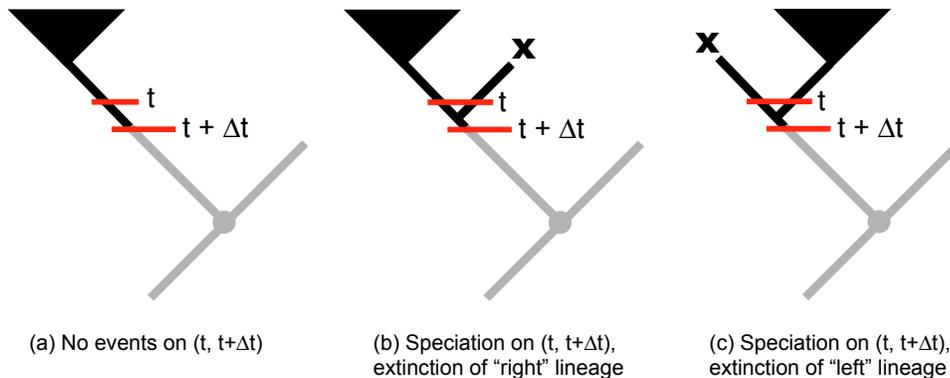


Figure S4: The set of scenarios by which a lineage at time  $t + \Delta t$  might yield a clade identical to the observed clade under a birth-death process. Labeling of "right" and "left" descendant lineages is arbitrary. Speciation is assumed to generate a symmetric diversification process (b, c), such that extinction of either progeny lineage (and all of its descendants) can still yield a clade identical to the observed clade.

Let  $D(t)$  denote the probability density that a lineage at time  $t$  gives rise to an evolutionary history identical to the observed history, and let  $E(t)$  denote the probability that a lineage at time  $t$  goes extinct before the present (along with all descendant lineages). We can write down a pair of ordinary differential equations that specify the changes in  $D(t)$  and  $E(t)$  as a function of time, given the observed phylogenetic tree.

$$\begin{aligned}\frac{dD}{dt} &= -(\lambda + \mu)D(t) + 2\lambda E(t)D(t) \\ \frac{dE}{dt} &= \mu - (\lambda + \mu)E(t) + \lambda E(t)^2\end{aligned}$$

The equation for  $D(t)$  accounts for the set of events listed in the above figure. The corresponding  $E(t)$  equation is derived in similar fashion, but must account for the fact that a speciation event on  $\Delta t$  must entail the extinction of *both* progeny lineages and their descendants such that the process goes extinct before the present. The probability of two lineages, identical in state, going extinct before the present, is simply the product of extinction probabilities (given that the events are independent),  $E(t)^2$ .

**Implications of the pass-up algorithm:** Imagine that we have mapped a rate shift to a phylogeny (the data augmentation), such that a shift occurs at time  $t_s$ . Let the speciation-extinction parameterization of the parent process be denoted by  $\Phi_0$  and let the corresponding parameters of the shift process (after  $t_s$  on the focal branch) be denoted by  $\Phi_1$ . In BAMM v2.5, the  $D$  calculation for the branch segment immediately rootwards of the shift, under parameter  $\Phi_0$ , is initialized with  $E(t|\Phi_1)$ , which is the probability that a lineage at time  $t$  goes extinct before the present given a shift to parameters  $\Phi_1$  at time  $t$ . This initialization implies the following symmetry to the underlying process:

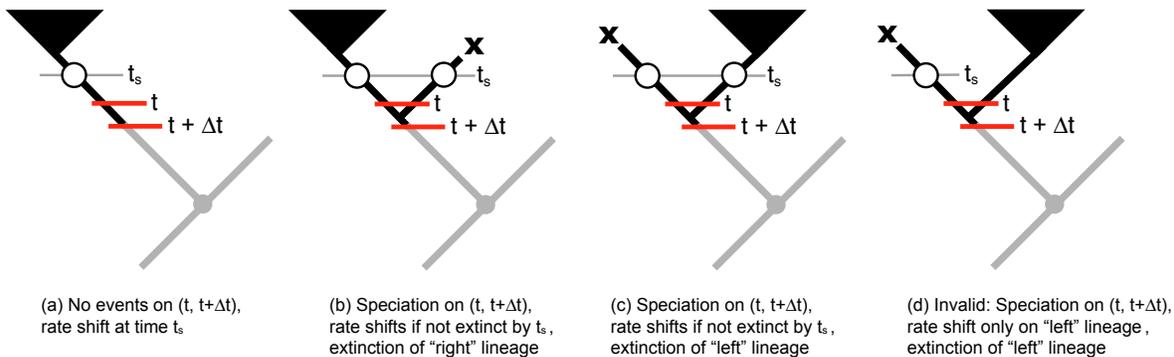


Figure S5: The set of scenarios on the time interval  $(t, t + \Delta t)$  by which a single lineage with a mapped rate shift can give rise to the observed data (a, b, c). For the process to yield descendant lineages consistent with the observed data, a rate shift must occur at time  $t_s$ ; if no rate shift occurs, the descendant lineages may be of incorrect type. If speciation occurs on  $(t, t + \Delta t)$ , any descendant lineages must all undergo rate shifts at time  $t_s$  provided they survive between  $t$  and  $t_s$  (b, c). If speciation generates lineages on  $(t, t + \Delta t)$  that do not undergo a rate shift, then extinction of either the "right" or "left" lineage defines an evolutionary process that is inconsistent with the observed data, as the surviving lineage may be of the wrong type (d).

The enumeration of these scenarios should not be taken to imply that two independent lineages have indeed undergone simultaneous rate shifts (Figure S5b, S5c). These are merely the scenarios

that are permitted to occur under the model. Any speciation events that have occurred on the rootwards portion of the branch between the shift at time  $t_s$  and the next rootwards node in the reconstructed phylogeny are assumed to yield a process that is identical to the observed branch - namely, that it will also have undergone a parallel shift at time  $t_s$  (if it survives to this time). However, note that any side lineages generated between the rate shift and the next rootwards node must go extinct before the present, or the process will not be consistent with the observed data.

**Implications of multiplying extinction probabilities at nodes:** BAMM v2.5 (default settings) multiplies  $E(t)$  probabilities at nodes that differ in their downstream shift histories, such that one or both descendant branches has a rate shift somewhere in their descendant subtrees. This  $E(t)$  value is then used to initialize the calculation on the parent branch. If there are no downstream shifts from a particular node, such that all lineages are of the same type, no multiplication is performed. Consider an interval of time  $(t, t + \Delta t)$  that is the branch subtending a node whose descendant lineages differ in their downstream shift history. In Figure S6, we have a branch leading to a clade P which contains a rate shift, and a branch leading to clade Q, which does not. Denote the most-recent common ancestor of P and Q as node PQ and denote this time as  $t_{PQ}$ . At time  $t_{PQ}$ , on rootwards side, the  $E(t)$  calculations are initialized with  $E_P(t_{PQ})E_Q(t_{PQ})$ . The effect of this multiplication is to assume that any speciation events occurring on  $(t, t + \Delta t)$  will generate a symmetric pair of diversification processes, as illustrated in Figure S6. However, because there is only one observed lineage (clade PQ), any such speciation events on this interval must be accompanied by a corresponding extinction: either the left or the right progeny lineage must go extinct to generate a subtree consistent with the observed data.

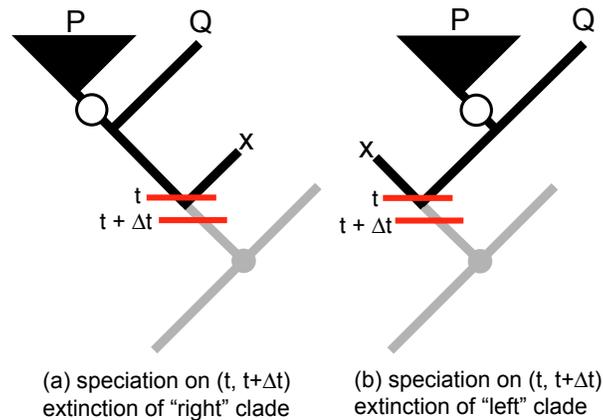


Figure S6: Alternative scenarios by which a speciation event on time interval  $(t, t + \Delta t)$  can give rise to a node that subtends lineages that differ in type. The reconstructed tree includes a speciation event that subtends lineages of type P (with rate shift) and Q (no rate shift). Any unobserved speciation events that occur on the internal branch subtending this node must generate a symmetric pair of diversification processes; this condition is necessary for extinction of the right (a) or left (b) descendant lineage to yield a present-day clade with descendants of type P and Q.

**The recompute algorithm for extinction probabilities:** The likelihood calculations in MEA differ from those in BAMM v2.5. For any interval of time  $(t_x, t_x + \Delta t)$  during which the diversification process has parameters  $\Phi_x$ , the extinction probability is the probability that an independent lineage with parameters  $\Phi_x$  originating at time  $t_x$  goes extinct before the present (along with all

descendants). Importantly, BAMM v2.3 and BAMM v2.4 also used the recompute algorithm. Because  $E(t)$  depends solely on the current parameters and  $t$ , there is no combining (multiplying or otherwise) of extinction probabilities at internal nodes.

## 2.5 History of the likelihood as implemented in BAMM

The likelihood calculations in BAMM have changed several times during the history of the program. We have performed extensive internal testing of the program following each change to ensure continued reliability of the program.

- **BAMM v2.2 and earlier (initial release through March 08, 2015):** The initial release of BAMM used the pass-up algorithm described in the previous section, with arbitrary inheritance of  $E(t)$  values through nodes.
- **BAMM v2.2 and earlier (March 08 2015 through Nov 9, 2015):** This release of BAMM implemented the *recompute* algorithm, exactly as MEA did in their article for their independent likelihood calculator. The version of BAMM available when MEA submitted their manuscript (listed as September 21, 2015) used the recompute algorithm, a fact that is documented on BAMM’s GitHub repository. Specifically, lines 548 - 595 in file SpExModel.cpp (from BAMM v2.3) and lines 633 - 691 (from BAMM v2.4) clearly demonstrate that BAMM v2.3 and BAMM v2.4 used the recompute algorithm. Thus, MEA’s decision to use the “combineExtinctionAtNodes = random” option did not recover the behavior of the most recent stable version of the program. Rather, their performance assessment of BAMM partially mimicked the behavior of an earlier version of the software (v2.2 and earlier).
- **BAMM v2.5 (released Nov 9, 2015):** BAMM v2.5 re-implemented the pass-up algorithm and introduced the multiply algorithm for handling  $E(t)$  values at nodes as described in the previous section.

**Why did BAMM abandon the *recompute* algorithm in v2.5?** The recompute algorithm as implemented in MEA is theoretically flawed. The problem arises due to inappropriate conditioning at the root. Strictly speaking, this is not a problem associated with *recompute* per se, but is a problem with the interpretation of  $E(t)$  at the root of the tree. As we have demonstrated in the main text, Appendix, and in previous sections, the inappropriate conditioning scheme used by MEA can yield infinite likelihoods. The most severe problems we have seen with MEA-type conditioning arise under the following topological mapping of rate shifts:

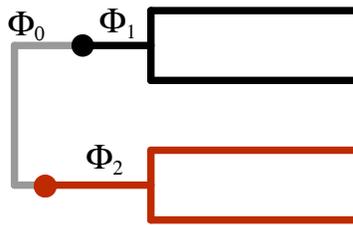


Figure S7: A rate-shift configuration with three diversification rate regimes: one at the root ( $\Phi_0$ ), plus one distinct regime on each of the two basal branches ( $\Phi_1$  and  $\Phi_2$ ).

Denote the mapping of rate shifts in the preceding figure as a basal-branches shift configuration. With a basal-branches shift configuration, the effects of incorrect survival conditioning (as in MEA) are acute. Specifically, the likelihood can increase to arbitrarily large values when the parameters at the root,  $\Phi_0$ , imply a high probability of extinction for an independent lineage between the root and the present day (e.g.,  $E(t|\Phi_0) \rightarrow 1$ ). In BAMM v2.4 and earlier versions, a design feature of the software prohibited the basal-branches shift configuration from arising. This was a feature that we had included in the earliest development versions of the software for bookkeeping purposes and was invoked by the default (hidden) command `validateEventConfiguration = 1`. With `validateEventConfiguration` set to 1, the program would automatically reject any MCMC moves during simulation of the posterior that placed rate shifts on both basal branches. Even though BAMM v2.3 and v2.4 used the recompute algorithm (as in MEA), BAMM still performed well in practice, because the program did not allow the basal-branches shift configuration to arise.

During the development of BAMM v2.5, we performed a number of exploratory analyses with `validateEventConfiguration = 0` and noted that this could lead to severe pathologies when likelihoods were computed with the recompute algorithm. We determined that the problem was attributable to incorrect survival conditioning, as discussed in the main text. To obtain a mathematically consistent definition of  $E(t)$  that could be used for conditioning at the root, we implemented the pass-up and multiply algorithms in BAMM v2.5. BAMM v2.5 was only released following extensive internal testing, during which we found no evidence that these design choices impacted performance of the software. Indeed, BAMM v2.5 performs at least as well as any previous version of BAMM for all performance metrics that we have explored; we have documented the performance of BAMM v2.5 relative to previous versions in this supplementary document and the main text.

**Future releases of the program:** The multiply and pass-up algorithms entail several unusual assumptions about the generating process (though we yet to observe any problems in practice due to these assumptions). The likelihood calculations in BAMM may change in the future. At present, there is no consensus on the correct way to compute the likelihood of a mapped set of diversification parameters on a phylogenetic tree.

### 3 Rate-Shift Frequencies

#### 3.1 BMM Estimates of the Shift Rate are Correlated with the True Rate

MEA correctly note that the rate at which rate shifts occur,  $\eta$ , is not necessarily equal to the rate at which rate shifts are observed across the extant portion of a phylogeny. In the extreme, they suggest a scenario whereby every rate shift leads to extinction; in such a case,  $\eta$  would not be estimable from the extant tree because all shifts would be unobserved. Here, we demonstrate that estimates of  $\eta$  from the extant portion of the phylogeny are generally correlated with the generating value. We simulated 200 phylogenies under two prior parameterizations for speciation and extinction, using the `SimulateCBDPtree` function distributed with MEA’s Dryad submission. For each simulation, we sampled a transition rate  $\eta$  from a uniform (0.001, 0.1) distribution, then simulated an instance of the compound Poisson birth-death process as described in MEA (MEA S3.1.2, p. 18). We used the same general parameterization as MEA: simulated trees were of the same age as the cetacean dataset (35.5 time units); speciation and extinction prior distributions were centered on their empirical values; and the transition rate distribution was loosely centered on the value they used for their variable rate simulations ( $\eta = 0.006$ ). The two prior parameters were (i) mean value of  $\lambda = 0.15$ , mean  $\mu = 0.05$  (exactly as in MEA variable rate simulations), and (ii) mean value of  $\lambda = 0.15$ , mean  $\mu = 0.15$ .

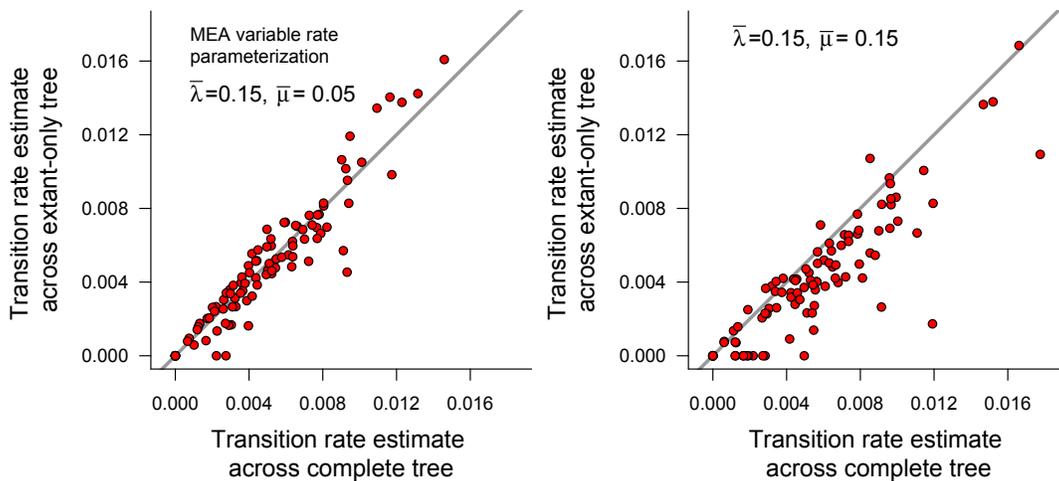


Figure S8: Estimated  $\eta$  values for the complete tree (e.g., total number of shifts / sum of branch lengths) versus the estimated transition rate  $\eta$  across the extant-only tree (number of shifts in extant tree / sum of branch lengths in extant tree) for two speciation-extinction prior parameterizations. In both parameterizations, the inferred values of  $\eta$  from the extant-only data are reasonably correlated with those that would have been estimated from the complete data (with extinct lineages). This suggests that  $\eta$  as estimated with BMM from extant-only data is a reasonable approximation to the true value, at least under prior parameterizations where extinction rates are not markedly higher than speciation rates.

#### 3.2 Empirical Transition Rates for MEA Datasets

MEA’s Figure S5 illustrates the effects of the transition rate (“shift rate”)  $\eta$  on extinction probabilities for five parameterizations of the  $\eta/\lambda$  ratio, a quantity we refer to as the *transition-to-speciation* ratio. In Figure S9, we illustrate the transition-to-speciation ratio, as inferred with BMM, across

all empirical datasets considered by MEA. In Figure S10, we perform a similar exercise but use speciation rates as inferred under a constant-rate birth-death (CRBD) process.

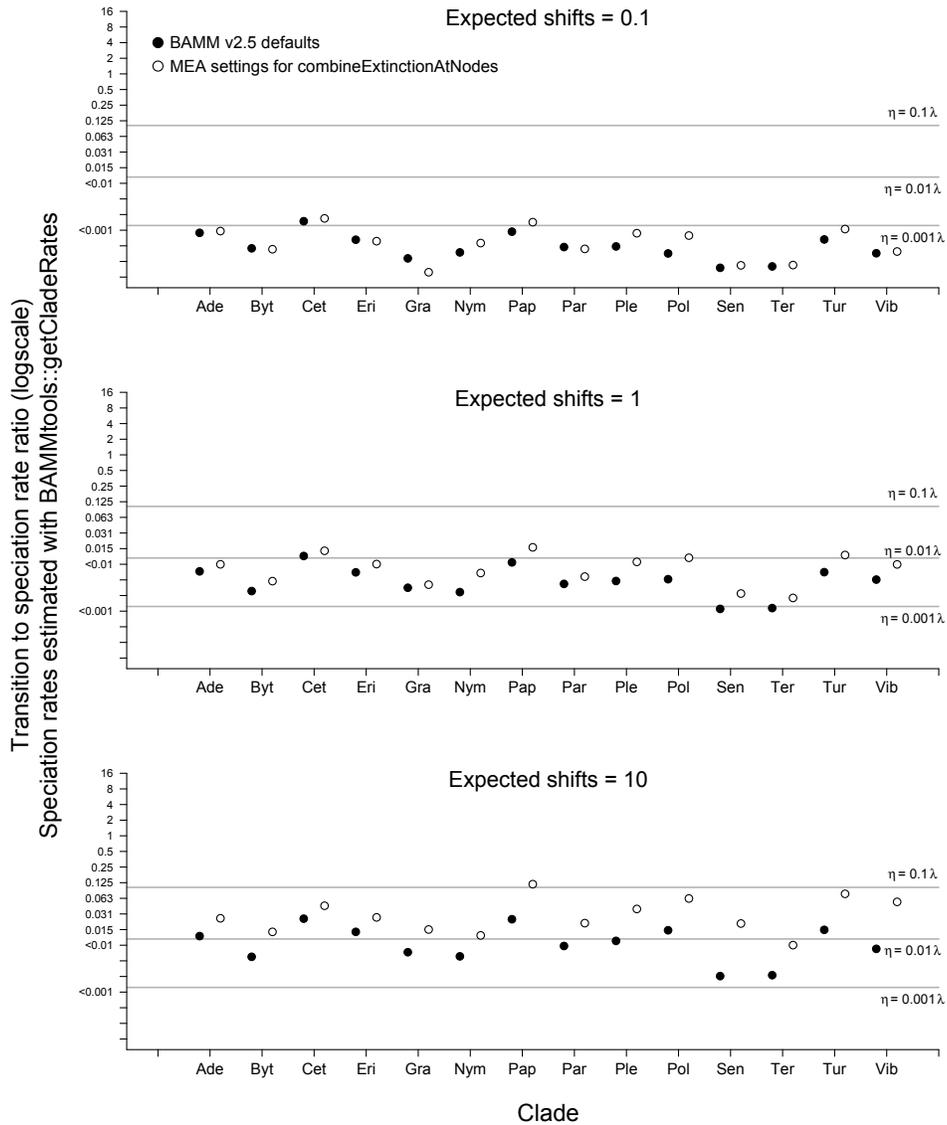


Figure S9: Estimated ratio of  $\eta$  to  $\lambda$  across 14 empirical datasets considered by MEA, across the range of  $\gamma$  priors (“expectedNumberOfShifts”) considered by MEA ( $\gamma = 0.1, 1, 10$ ). The transition rate  $\eta$  was inferred using BAMM exactly as used by MEA; mean speciation rates across each phylogeny were obtained using the function `getCladeRates` from the R package `BAMMtools`. Open and filled circles denote estimates obtained using the options `combineExtinctionAtNodes = random` (as in MEA) and `combineExtinctionAtNodes = if_different` (BAMM v2.5 default value). In all cases, the empirical shift-to-speciation ratio is far lower than the values considered in MEA; the minimum parameterization illustrated by MEA is  $\eta = 0.5\lambda$ . Note results plotted on logarithmic scale.

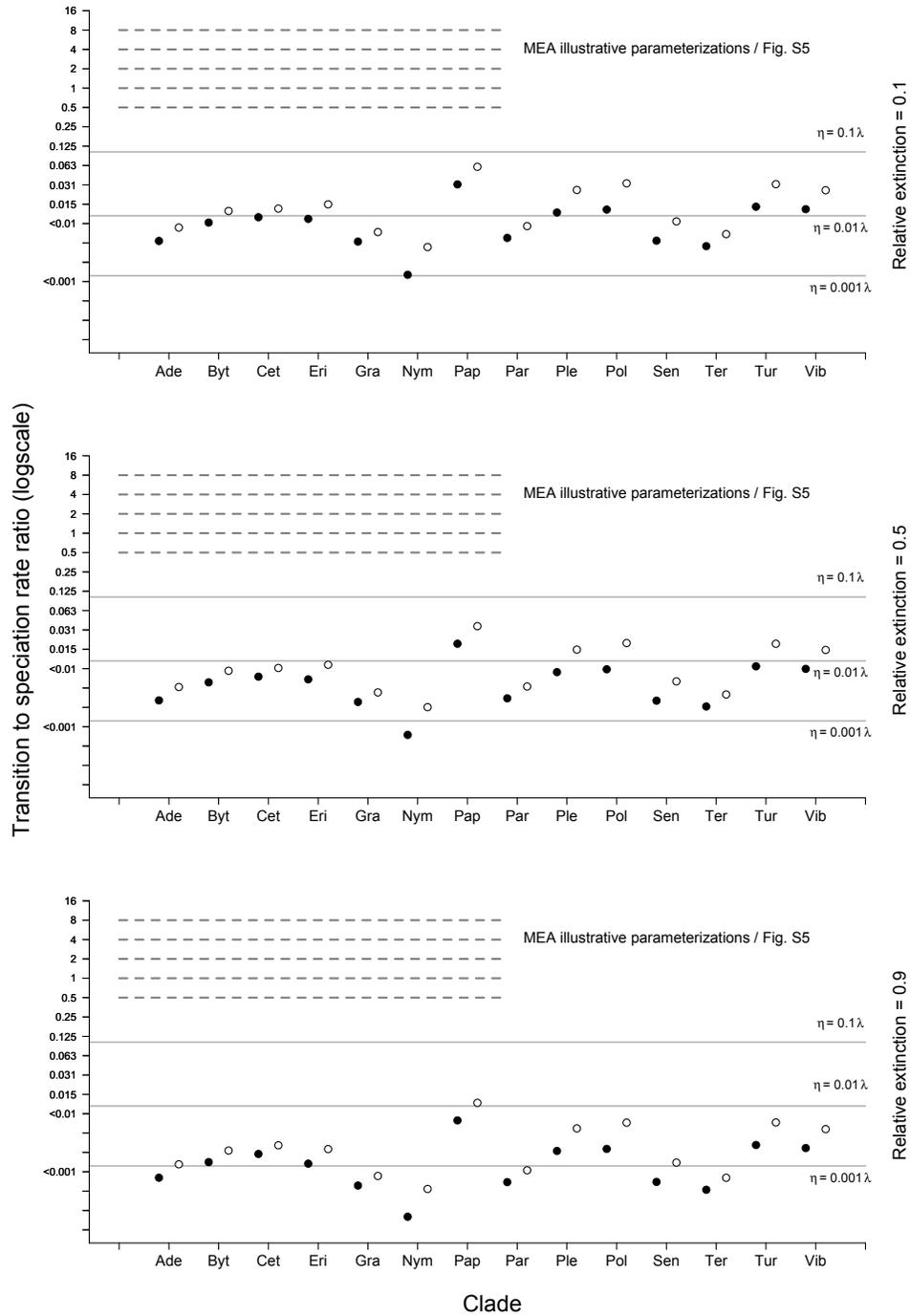


Figure S10: Estimated ratio of  $\eta$  to  $\lambda$  across 14 empirical datasets considered by MEA, where  $\eta$  was estimated with BAMM but the speciation rate  $\lambda$  was estimated under a constant-rate birth-death process. Prior on number of shifts was  $\gamma = 1$  for this analysis. Speciation rates were estimated under three relative extinction rates ( $\mu/\lambda$ ). Open and filled circles denote estimates obtained using the options `combineExtinctionAtNodes = random` (as in MEA) and `combineExtinctionAtNodes = if_different` (BAMM v2.5 default value). The five parameterizations analyzed in MEA's Fig. S5 are presented for comparison. In general, the empirical shift-to-speciation ratio is approximately 100 to 1000 times lower than the values considered in MEA. Note results plotted on logarithmic scale.

## 4 Extinction Probabilities and Unobserved Rate Shifts

MEA describe a computationally-intensive method for simulating extinction probabilities under a BAMM-type model of diversification, where lineages undergo rate shifts to new speciation-extinction parameters with rate  $\eta$ . It is important to recognize that their approach does not actually infer a true extinction probability using data: rather, it simulates extinction probabilities under prior assumptions about the distributions of  $\lambda$  and  $\mu$ . Typically, in Bayesian inference, the joint posterior probability of the model is proportional to the product of the likelihood and the prior. In MEA, the likelihood itself is a function of the data, the parameters, and the prior. As such, the likelihood itself varies as a direct function of whatever prior parameterization the researchers specifies to describe the distributions of unobserved rate shifts. By definition, these distributions are not observed and are not informed by data. The Monte Carlo method of simulating extinction probabilities has two major drawbacks:

- As noted above, the approach is largely *data-free*: the method can only account for the effects of unobserved rate shifts on the extinction probability under strong prior assumptions about the (true) distributions of speciation and extinction rates.
- The approach is computationally-intensive and increases the time required to compute the likelihood by many orders of magnitude. This cost is due to the fact that extinction probabilities must be simulated; when unobserved rate shifts are ignored, as in BAMM, the extinction probability can be computed analytically.

Neither of these points is directly addressed by MEA, who assume that accounting for unobserved rate shifts under prior assumptions is better than assuming  $\eta = 0$  (as in BAMM, MEDUSA, and all other rate-shift models; Table 2). Because we have no information about the true distribution of  $\lambda$  and  $\mu$  for unobserved rate shifts, it is not obvious how we can test whether the Monte Carlo extinction estimates correspond to reality. In fact, as such rate shifts are (by definition) unobserved, it is not clear that any data could be collected, even in principle, that would inform us as to the validity of MEA's simulated extinction probabilities.

Despite this concern, we can still assess whether the extinction probabilities as estimated by MEA's Monte Carlo procedure have empirical consequences for inference. As described in the previous section, we estimated the transition rate  $\eta$  ("rate-shift rate") for all 14 empirical datasets analyzed by MEA, by repeating their analyses with the BAMM control files they provided with their Dryad submission to accompany their article.

For each of the 14 empirical phylogenies considered by MEA, we:

- We first rescaled each phylogeny to a crown age of 1.0. All results below are mathematically independent of this rescaling; it is simply performed to place all phylogenies on a common (0,1) relative divergence timescale
- We estimated the distribution of  $\eta$  from the BAMM-estimated posterior on the number of rate shifts (dividing the whole-tree event rate  $\Lambda$  for given MCMC sample by the sum of branch lengths in the tree)
- We estimated rates of speciation and extinction for the clade under a constant-rate birth-death process using crown clade age (here, 1.0) and total species richness (after Magallon and Sanderson 2001)

- Speciation and extinction rates were estimated under 3 relative extinction rates:  $\mu/\lambda = 0.1$ , 0.5, and 0.9
- We used MEA’s Monte Carlo extinction simulator with these rates of  $\eta$ ,  $\lambda$ , and  $\mu$  to simulate 50,000 processes beginning with a single lineage at time  $t = 0$  and ending at time  $t = 1$  (as noted above, all trees were rescaled to a relative timescale)
- We parameterized the prior distributions for  $\lambda$  and  $\mu$  based on the constant-rate birth-death (CRBD) estimates. The exponential priors for both  $\lambda$  and  $\mu$  were set to have a mean equal to the CRBD estimate of  $\lambda$
- These simulations enabled us to reconstruct cumulative extinction-through-time curves for each empirical dataset (e.g., we could reconstruct the probability of complete lineage extinction by time  $t = 0.50$ )

These Monte Carlo extinction probabilities represent that a single lineage originating at the crown age of a particular empirical dataset goes extinct before the present (along with all of its descendant lineages). We compared these Monte Carlo extinction estimates to the corresponding probability under a constant-rate birth-death process, where the probability that a single lineage leaves zero descendants after some time  $t$ ,  $P_{0,t}$ , is computed as

$$P_{0,t} = \frac{\mu(e^{(\lambda-\mu)t})}{\lambda e^{(\lambda-\mu)t} - \mu} \quad (2)$$

after Bailey (1964), Raup (1985), and others. We also compared extinction probabilities for each clade to the illustrative parameterizations considered by MEA (their Figure S5). MEA illustrated the effects of unobserved rate shifts on lineage extinction probabilities as a function of  $\eta/\lambda$  ratio (e.g., the ratio of the rate-shift rate to the speciation rate). Given our estimates of  $\lambda$  for each clade, we simply computed transition rates to match the MEA parameterizations of  $\eta/\lambda = 0.5, 1, 2, 4$ , and 8.

Figure S11 compares the probability of extinction using MEA’s Monte Carlo simulator to those obtained under the simple analytical approximation given above. Using MEA’s computationally-intensive inference scheme does not appreciably change the probability of lineage extinction relative to what would be obtained under a constant-rate birth-death process, at least when biologically plausible parameterizations of the shift rate  $\eta$  are used. Hence, there appears to be little reason at present to incorporate these computationally-demanding “unknown unknowns” into empirical inference. Figure S12 illustrates the same extinction trajectories from Figure 3 (main text), but after activating the “random” option used by MEA; this option has no discernible effect on inference.

These results are a simple consequence of the fact that, when  $\eta$  is low relative to  $\lambda$ , unobserved shifts cannot appreciably influence the chance of lineage extinction. Indeed, all of the estimated shift rates  $\eta$  from MEA’s analyses suggest that the frequency of rate shifts is very low relative to speciation rates, perhaps 100 - 1000x less common; see Figure S9 and Figure S10. The illustrative parameterizations considered by MEA (their Figure S5) imply considerable significance for these unobserved rate shifts, but this perception is due to extremely high ratios of  $\eta/\lambda$  they considered. The  $\eta/\lambda$  ratios considered by MEA are approximately 2-3 orders of magnitude higher than the values they estimated for their empirical datasets.

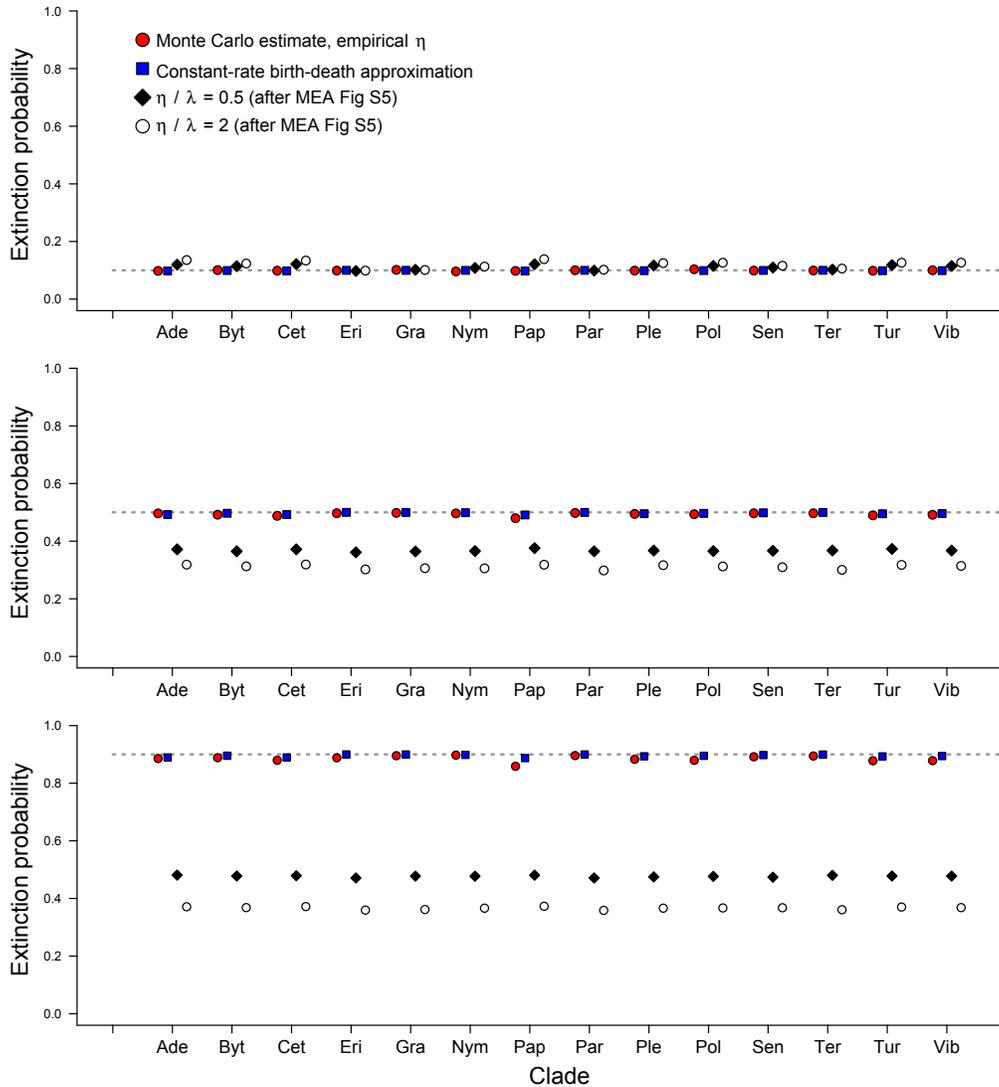


Figure S11: MEA Monte Carlo extinction estimates are virtually identical to those that would be obtained under a constant-rate birth-death process for biologically-relevant parameterizations of the diversification process. Each point represents the cumulative probability that a lineage originating at the root node for each clade goes extinct before the present, for each of 14 empirical datasets considered by MEA. Rows correspond to relative extinction rates of 0.1 (top), 0.5 (middle), and 0.9 (bottom). Sets of points for each clade compare extinction probabilities from MEA’s computationally-intensive Monte Carlo simulator (red circles) to those obtained under a simple (analytical) constant-rate birth-death approximation (blue squares). For comparison, we also present extinction probabilities under the minimum  $\eta/\lambda$  ratio considered in MEA’s illustrative Figure S5 ( $\eta/\lambda = 0.5$ ) and for  $\eta/\lambda = 2$ . In each case, the Monte Carlo extinction estimates are nearly identical to those from the analytical probability that assumes  $\eta = 0$ . However, these empirical extinction probabilities are far higher (at least for relative extinction = 0.5 and 0.9) than those that would be obtained under MEA’s illustrative parameterizations (black diamonds; open circles). This discrepancy is due to the fact that MEA’s illustrative parameterizations specify  $\eta/\lambda$  ratios that are approximately 100 - 1000x greater than the empirical values estimated for any of their focal datasets.

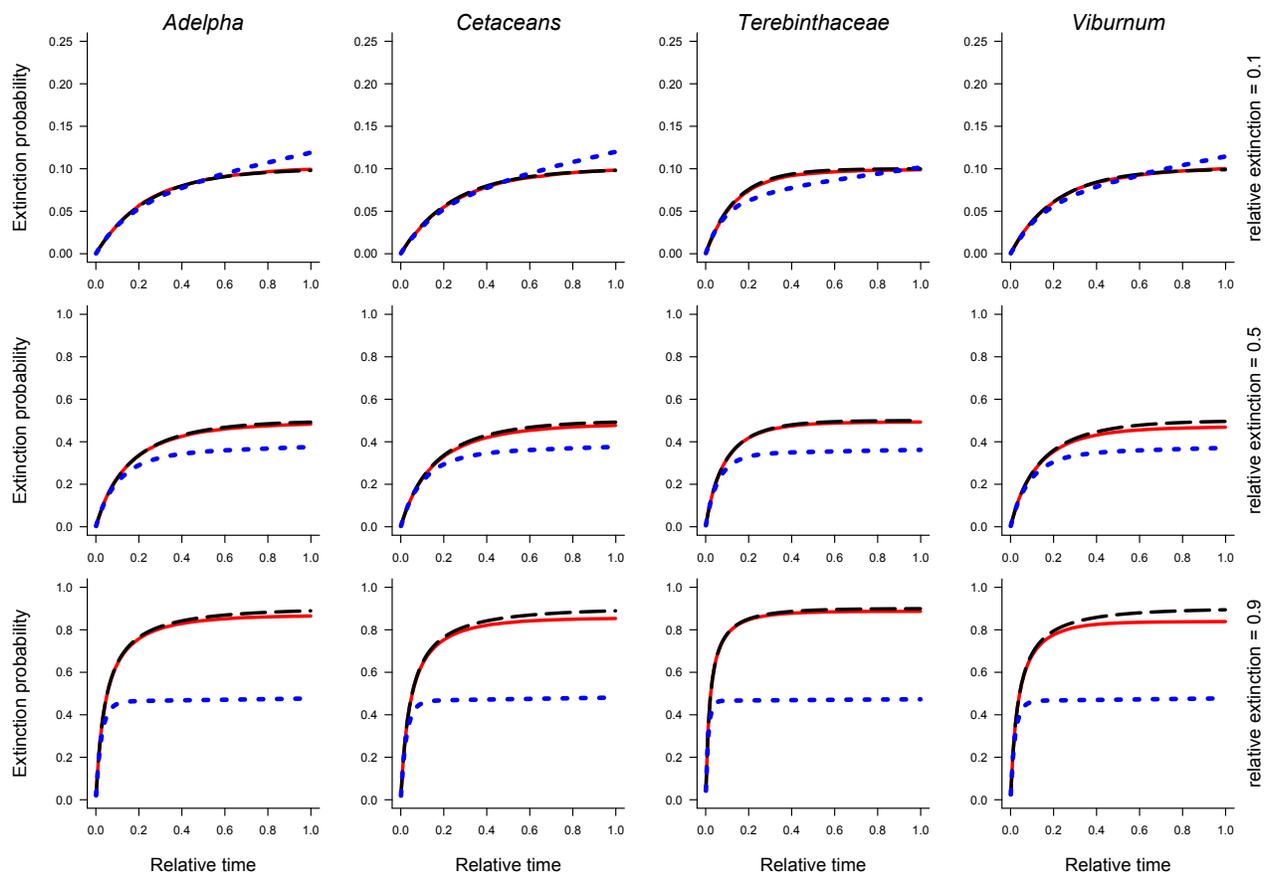


Figure S12: The chance of lineage extinction as a function of relative time from the root node for four empirical datasets analyzed by MEA (columns) and under three different relative extinction rates (rows). Figure is identical to Figure 3 from the main text but presents analyses with the “random” option used by MEA. 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)  $\eta$ . 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 (their Fig. S5). 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 $\times$  greater than for the empirical datasets considered in their article.

## 5 Reanalysis of MEA’s “variable rates” dataset

### 5.1 Ascertainment biases in MEA’s simulation study

Our simulations (see Section 6.1), and MEA’s, rely on the same underlying Poisson process model of rate variation: in any given interval of time  $\Delta t$ , a lineage may undergo speciation, extinction, or a rate shift (with total rate  $\lambda + \mu + \eta$ ). Following a rate shift, a lineage is assigned new diversification parameters drawn from exponential distributions with means  $\psi_\lambda$  and  $\psi_\mu$ . MEA selected their phylogenies to contain at least one rate shift and a number of tips between 50 and 150 extant species. As we have shown (main text Figure 7), MEA’s phylogenies contain very small rate regimes: for 58% of their phylogenies, the second-largest rate regime contained 5 or fewer tips. For 80% of their phylogenies, the second-largest regime contained fewer than 20 tips. The low statistical power associated with MEA’s trees results largely from ascertainment bias that discarded many phylogenies of large size and/or large shift regimes that were expected to be common under their simulation parameters (Fig. 7c). To demonstrate this sampling effect, we simulated 100 phylogenies that contained at least one rate shift, using their code and input parameters ( $\eta = 0.006$ ,  $\psi_\lambda = 0.15$ ,  $\psi_\mu = 0.05$ ,  $T_{max} = 35.5$ ) and found that tree sizes and shift counts are much larger in the unconstrained simulations. It was still necessary for us to abort simulations that generated very large ( $N > 2000$  tip) phylogenies, and the results presented here are thus conservative. **Key points of bias in MEA’s simulations relative to the Poisson process they purported to simulate include:**

- MEA report median and mean tree sizes for their “variable rates” dataset of 84 and 89.1, respectively. In our sample of 100 trees, we find corresponding values of 77 and 342. If we impose MEA’s lower bound rejection criterion (thus discarding trees with fewer than 50 tips), the median and mean tree sizes rise to 239 and 534, respectively. These results are shown in Figure S13.
- MEA report that the mean and max number of diversification rate shifts in trees in their dataset were 4.35 and 10, respectively. We find that an unbiased sample from the distribution implied by their parameters is characterized by a larger mean (10.02 shifts) and max (56 shifts). *The expected number of shifts under their parameters is thus larger than the largest value that was observed in their dataset.* This mean only increases as one excludes tiny trees from the sample (e.g., excluding trees with fewer than 25 tips gives a mean of 14.2 shifts per tree).
- From our set of 100 simulated trees, we find that MEA’s rejection criteria would have excluded 38% of simulated phylogenies of (extant) size less than 50 tips.
- We found that an additional 43% of simulations exceeded their upper bound acceptance criterion ( $N = 150$ ) and would thus have been excluded from their sample. Fully 24 of 100 simulations yielded trees larger than 500 tips under MEA’s parameters.
- Of the total 1002 rate shifts with extant descendants that were generated during this simulation of 100 trees, only 74 (7.4%) would have been accepted under MEA’s criteria.

By rejecting all simulations that included more than 150 extant tips, MEA excluded most simulated phylogenies that contained statistically-detectable rate shifts. This ascertainment bias selected for simulated phylogenies that included only small rate regimes; phylogenies with larger

(detectable) rate regimes would have frequently exceeded the size selection threshold ( $N = 150$  tips). Moreover, by selecting only small phylogenies, MEA generally discarded outcomes where trees experienced nested shifts in diversification rate. This is a common expectation under the Poisson process model of rate variation. Under the Poisson process, an increase in net diversification rates anywhere in the tree increases the expected number of rate shifts across the tree: if a lineage undergoes an increase in rates, then it is expected to leave more descendant lineages. Larger subclades have more lineages that can potentially undergo rate shifts (each with independent rate  $\eta$ ); for this reason, diversification rate increases may frequently yield compound and nested patterns of rate shifts in practice. MEA's simulations would have rejected this frequent outcome of the Poisson process, by severely restricting tree sizes. This property of MEA's simulations is demonstrated in Fig. 7c (main text) and in Figure S13, which illustrates that most outcomes from the Poisson process simulated by MEA would have been rejected under their stringent criteria for accepting a simulation as valid.

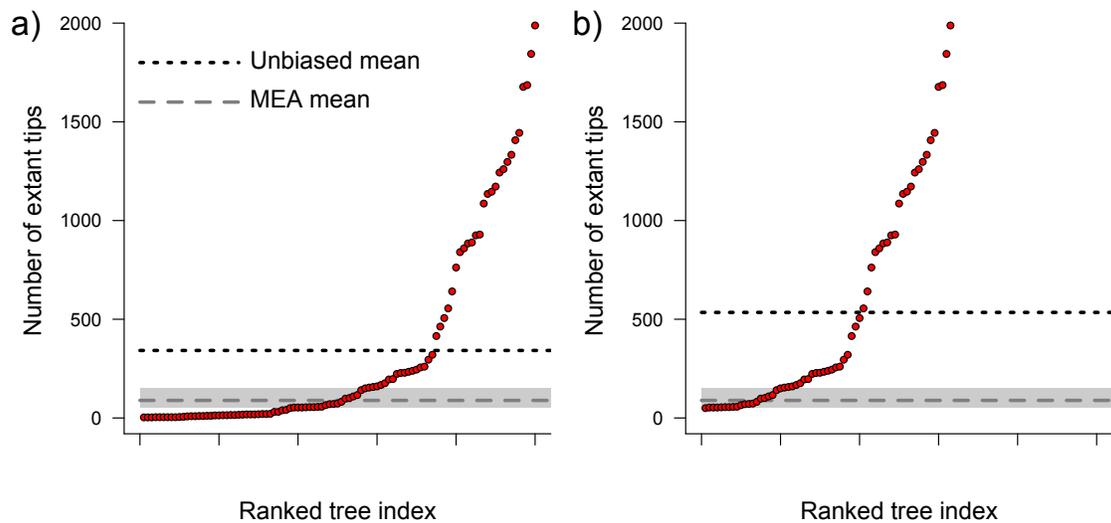


Figure S13: Distribution of tree sizes from parameters used to generate MEA's "variable-rates" dataset. MEA rejected all simulations that yielded trees with more than 150 or fewer than 50 tips. The gray polygon denotes the the MEA acceptance zone: all simulated phylogenies falling outside of this zone would have been rejected. When these simulations are repeated using their simulation code and parameters, but without imposing this strict rejection criterion, we see that MEA excluded a large fraction of the expected simulation outcomes under their parameterization. All trees with at least two surviving lineages are shown in (a), and (b) illustrates a truncated distribution after imposing MEA's lower-bound rejection criterion ( $N = 50$ ).

## 5.2 Information Content of MEA Variable-Rate Phylogenies

In the main text, we present an analysis of the maximum theoretical information content associated with each rate shift in the MEA dataset,  $\Delta\text{Log}L$ . This information content was computed as the difference in likelihoods for each shift regime under the true (generating) parameters and under the likelihood that would have been computed if we simply used parameters from fitting a constant-rate birth-death process to the tree as a whole. Here, we provide additional analysis where we compute  $\Delta\text{Log}L$  by finding maximum likelihood speciation and extinction parameters for each rate partition (e.g., rate regimes  $r_1$  and  $r_2$  in Figure 7b) under the analytical approximation to the likelihood; we consider the analytical approximation with  $\eta = 0$  to be reasonable given that the computationally-intensive Monte Carlo likelihood from MEA had no discernible effect on inference (Fig. 8b). We believe that the  $\Delta\text{Log}L$  statistic for information content should be computed with the true parameters, as this is the comparison made in MEA (e.g., MEA compare BAMM estimates to true estimates of rates; they do not compare BAMM estimates to independent ML rate estimates). However, we provide these summaries for comparison, which yield virtually identical results to those provided in the main text but with one difference:  $\Delta\text{Log}L$  cannot be less than zero, because the ML parameter estimates for each partition considered separately will always be better than the whole-tree CRBD approximation. Importantly, if MEA's variable-rates phylogenies are compromised by ascertainment biases, we might expect the ML-optimized parameter sets to yield substantially different performance relative to the true parameter values; such biases are possible because MEA rejected all simulated phylogenies with fewer than 50 or more than 150 tips, thus selecting only a narrow range of simulated outcomes for analysis. As shown in Figure S14, values of  $\Delta\text{Log}L$  are highly correlated regardless of whether the true or ML parameter estimates are used, and as reported in the main text, the information content of most phylogenies is low in both cases.

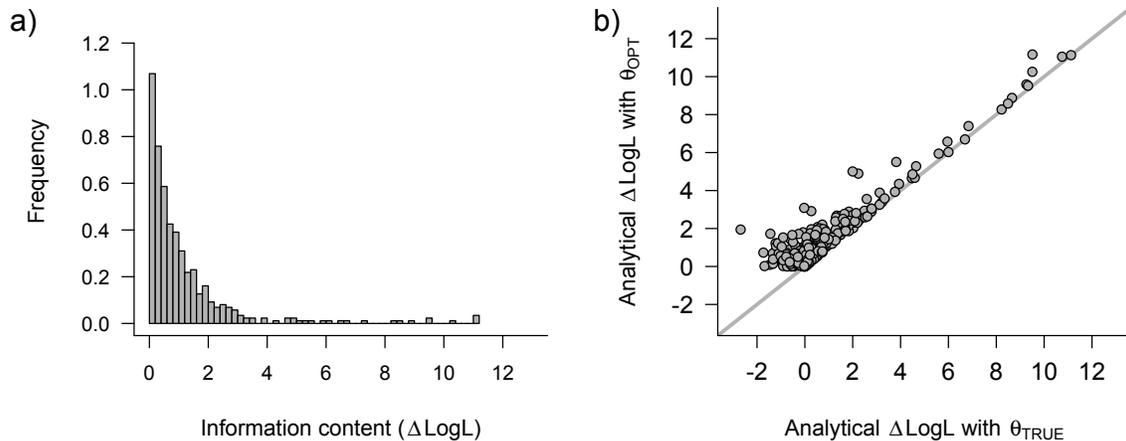


Figure S14: (a) Histogram of  $\Delta\text{Log}L$  values using the ML estimates for each shift regime; this is similar to Figure 8a but with  $\Delta\text{Log}L > 0$ . Even here, we find 404 of 435 rate shifts showing  $\Delta\text{Log}L < 3$ . (b) shows the corresponding correlation between  $\Delta\text{Log}L$  computed with the true parameters versus  $\Delta\text{Log}L$  computed with ML estimates. While the ML estimates are better, this effect is only manifest for rate shifts of very small to zero effect ( $\Delta\text{Log}L < 3$ ).  $\theta_{\text{TRUE}}$  and  $\theta_{\text{OPT}}$  refer to true and ML-estimated parameters, respectively.

### 5.3 Analysis of MEA Trees with Time-Varying BAMM

MEA simulated datasets under a process where diversification rates are constant through time within rate regimes. However, they analyzed those datasets under a more complex BAMM model that allowed rates to vary through time within rate regimes. In the main text, we presented only the results using *constant-rate* BAMM, constraining speciation and extinction rates for the  $i$ 'th regime such that  $\lambda_i(t) = \lambda_i$  and  $\mu_i(t) = \mu_i$ . In this section, we provide the comparable analyses using time-varying rate regimes in BAMM, exactly as performed by MEA. We demonstrate that virtually identical results are obtained even when time-varying BAMM is used to analyze time-constant datasets, although there is more noise in the analyses due to the overparameterization associated with inclusion of an unnecessary exponential change parameter (`lambdaShift`). Figure S15 is virtually identical to Figure 9 in the main text but is based on time-varying rate regimes. Figure S16 illustrates the consequences of MEA's low-power trees for the regression analyses presented in their article and is similar to Figure 10 in the main text. Figure S18 summarizes BAMM's performance across individual rate regimes under multiple prior parameterizations expected number of shifts ( $\gamma$ ); results are virtually identical to those shown in Fig. 12 in the main text.

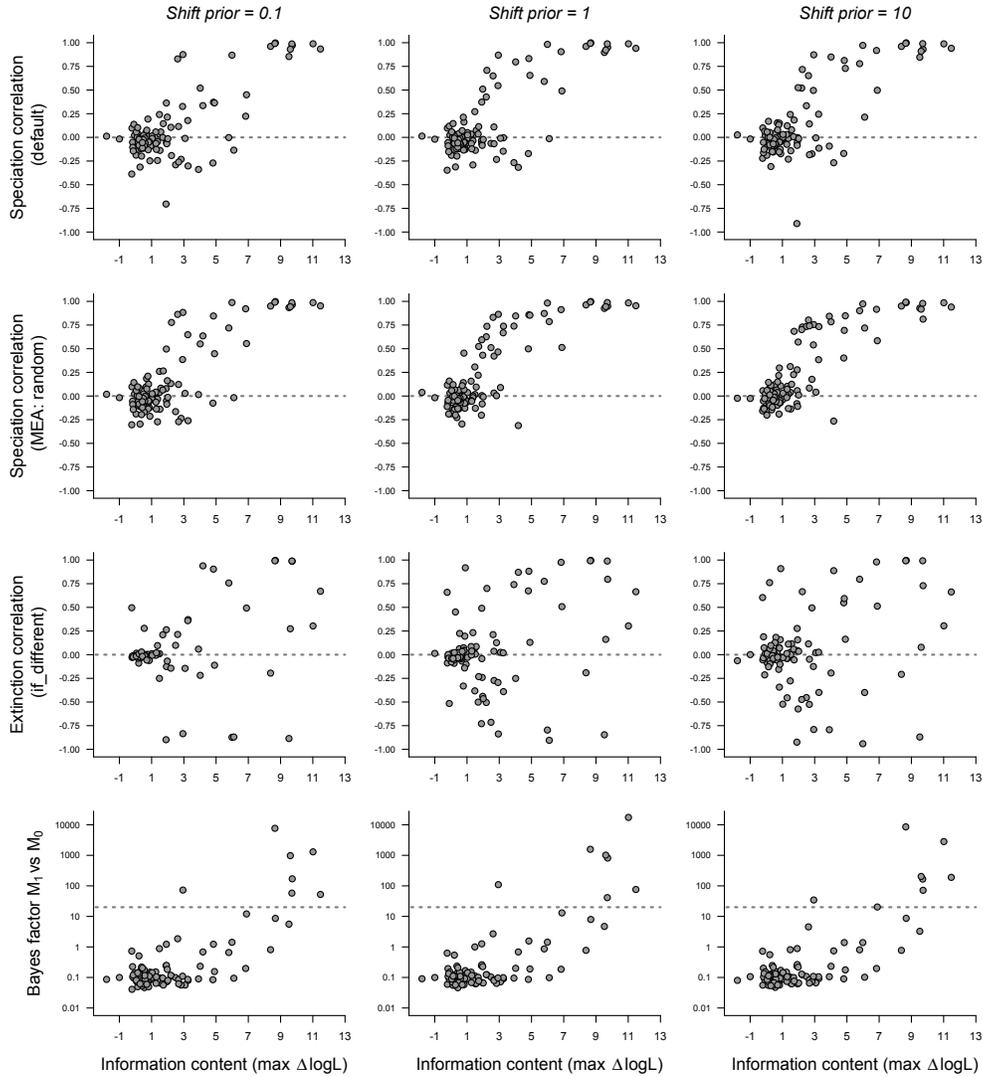


Figure S15: Summary attributes of BAMM analysis of all trees in MEA “variable rates” dataset (their Fig. 6) analyzed under three priors (columns), plotted by the information content in each tree. Information content for each tree is the maximum  $\Delta\text{Log}L$  value across all shifts occurring 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 ( $\text{max } \Delta\text{Log}L > 3$ ). 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 dataset, 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  $\Delta\text{Log}L$  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 ( $\text{BF} = 20$ ). For trees with maximum  $\Delta\text{Log}L < 3$  (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). Results here use BAMM with time-varying rate regimes and are virtually identical to those presented in Fig. 9 in the main text.

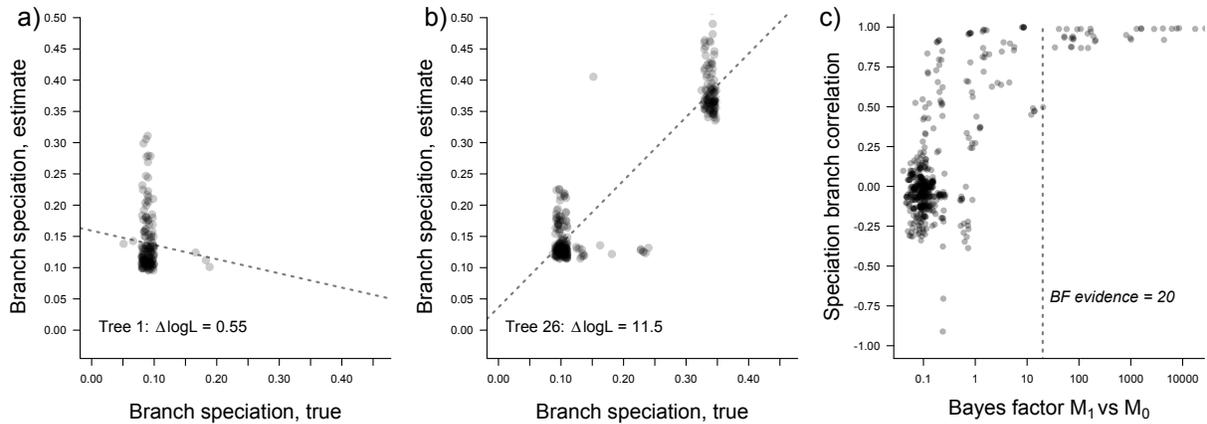


Figure S16: Low statistical power confounds MEA's assessment of BMM speciation rate reliability. (a) Regression analysis of BMM-estimated speciation rates as a function of the true rates for the first tree in MEA's variable-rates dataset, which is characterized by low information content ( $\Delta\log L_{max} = 0.55$ ; only 1 or 2 tips per rate shift). Slope of fitted regression line is -0.23 (dotted line) and not significantly different from 0 ( $p = 0.41$ ,  $r^2 = 0.004$ ). Points have been jittered to reduce overplotting. (b) Comparable analysis for a tree with high information content (tree 26). Slope of fitted regression line is 1.01 ( $p < 0.001$ ;  $r^2 = 0.88$ ). (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). Results here based on BMM with time-varying rate regimes and are virtually identical to those presented in Fig. 12 in the main text.

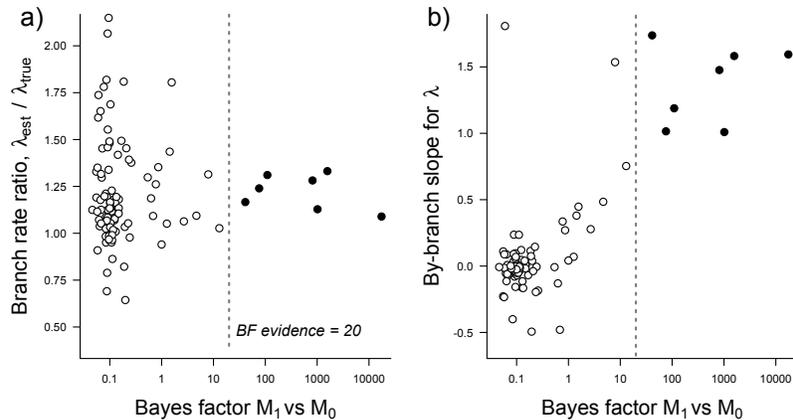


Figure S17: Branch-specific estimates of speciation rates obtained with BAMM are more accurate when trees contain evidence for rate variation (a) Proportional accuracy of branch-specific speciation rate estimates as a function of the Bayes factor evidence for rate heterogeneity in the tree; each point 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 1.37. 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 ( $BF = 20$ ). Analyses shown here are similar to Figure 11 in the main text but were obtained using BAMM with time-varying rate regimes.

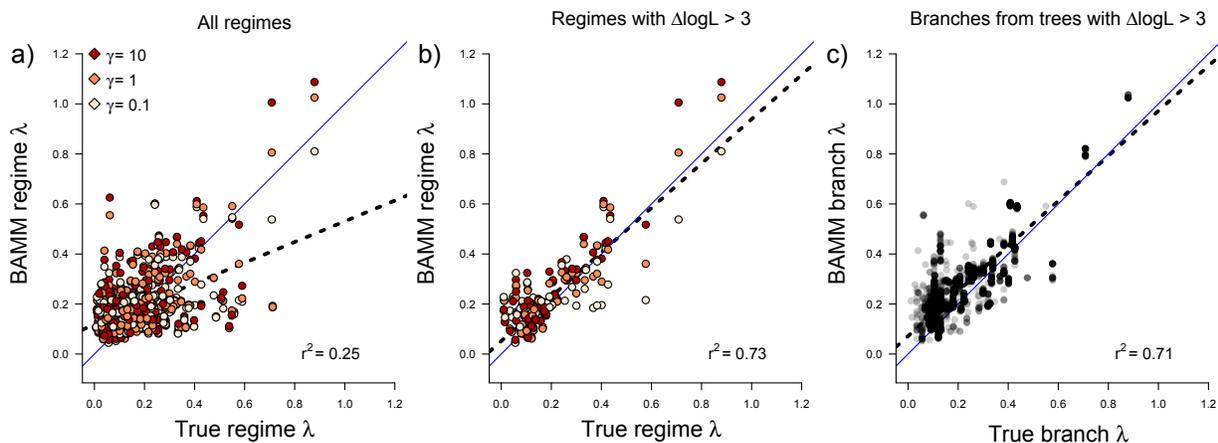


Figure S18: (a) Relationship between true speciation rates and BAMM-estimated rates across all 535 rate regimes in the MEA variable-rates dataset, estimated under three different prior parameterizations (colors;  $\gamma = 0.1, 1,$  and  $10$ ). Correlation is positive; dashed line is fitted regression line (slope =  $0.41$ ,  $r^2 = 0.25$ ). Blue line denotes theoretical 1:1 expectation. (b) Same as (a), but where all rate regimes with low information content ( $\Delta \text{Log}L < 3$ ) are excluded. Slope of fitted regression line is  $0.89$  ( $r^2 = 0.732$ ). (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 ( $\Delta \text{Log}L < 3$ ) from analyses with  $\gamma = 1$  (23 trees; 4374 branches total). Slope of fitted regression line is  $0.900$ ; results for other prior parameterizations are virtually identical to those shown in (c). Results here based on BAMM with time-varying rate regimes and are virtually identical to those presented in Fig. 12 in the main text.

## 5.4 Robustness of Results to Priors and CombineExtinctionAtNodes settings

In this section, we demonstrate that speciation rate estimates obtained with BAMM are robust to both (i) the prior on the number of shifts,  $\gamma$ , as well as (ii) the specific option used for `combineExtinctionAtNodes` (e.g., the default BAMM v2.5 setting, or “random” as used by MEA). Figure S19 is identical to Figure 12 from the main text, but uses `combineExtinctionAtNodes = if_different` (BAMM default value). Figure S20 and Figure S21 are identical to Fig 12c but demonstrate BAMM’s performance across both `combineExtinctionAtNodes` options and across three priors on the expected number of rate shifts ( $\gamma = 0.1, 1,$  and  $10$ ). We used only the MEA “variable rates” phylogenies for these analyses; performance would presumably improve if the target phylogenies were more informative (see main text figures 7-8).

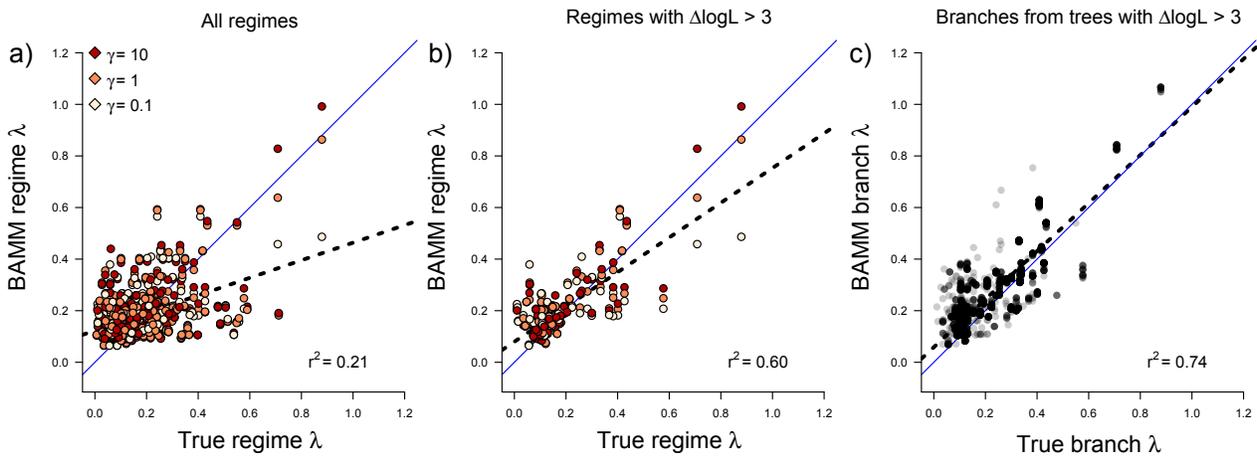


Figure S19: (a) Relationship between true speciation rates and BAMM-estimated rates across all 535 rate regimes in the MEA variable-rates dataset, estimated under three different prior parameterizations (colors;  $\gamma = 0.1, 1,$  and  $10$ ), and using the BAMM v2.5 default setting for `combineExtinctionAtNodes`. Dashed line is fitted regression line (slope = 0.34). Blue line denotes theoretical 1:1 expectation. (b) Same as (a), but where all rate regimes with low information content ( $\Delta \log L < 3$ ) are excluded. Slope of fitted regression line is 0.67. (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 ( $\Delta \log L > 3$ ) 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). Compare to Figure 12 from main text.

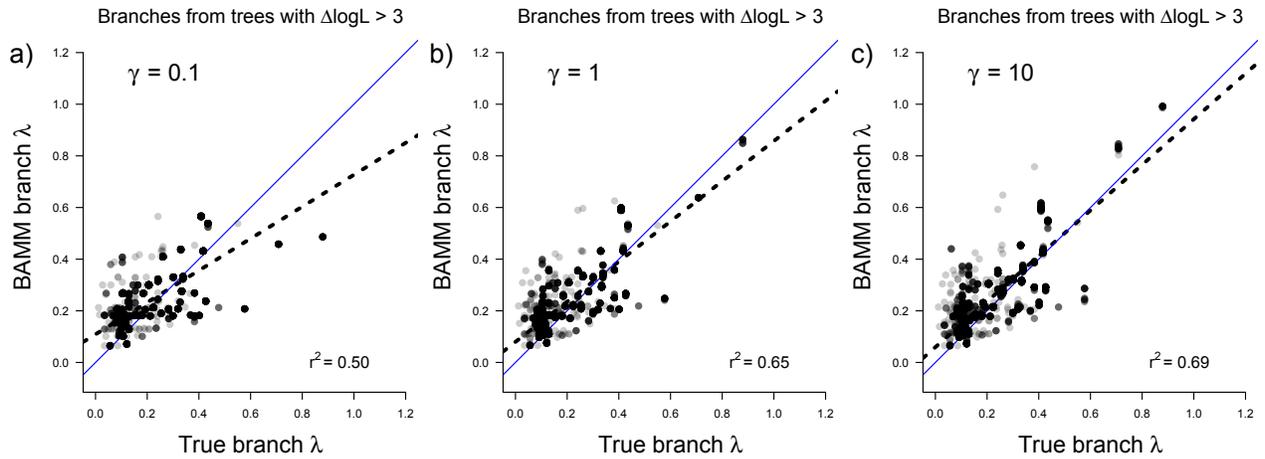


Figure S20: Branch-specific estimates of speciation against the true values for all trees all trees where at least one shift had non-trivial information content ( $\Delta\text{Log}L > 3$ ). These results were obtained with BMM v2.5 default settings and are illustrated for shift priors of (a)  $\gamma = 0.1$ , (b)  $\gamma = 1.0$ , and (c)  $\gamma = 10$ . Blue line is perfect 1:1 relationship; dotted line is fitted regression line for each set of points. Regression slopes for a, b, and c are 0.62, 0.78, and 0.88, respectively. A total of 23 phylogenies and 4374 branches are included in the results shown here.

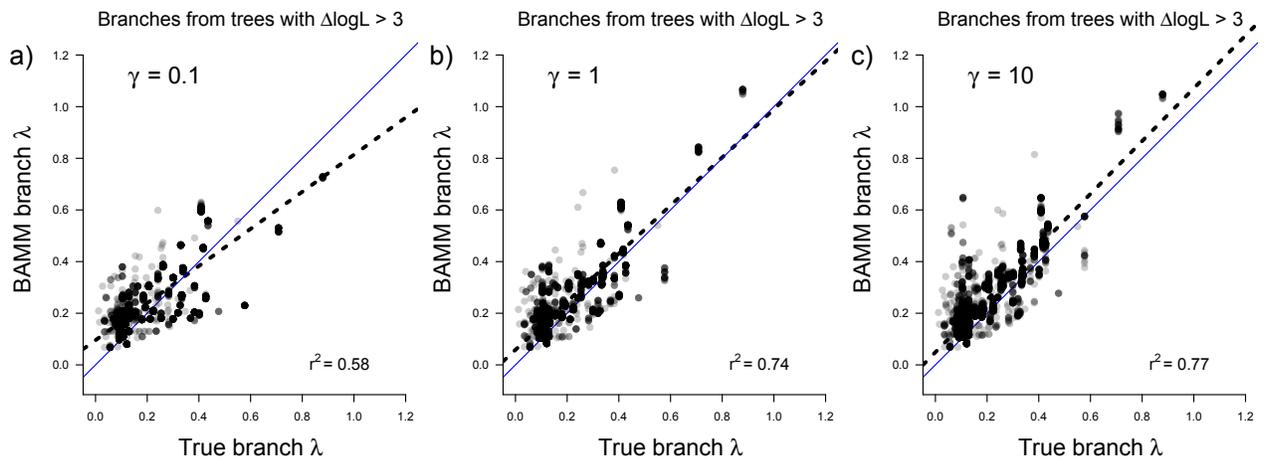


Figure S21: Branch-specific estimates of speciation against the true values for all trees all trees where at least one shift had non-trivial information content ( $\Delta\text{Log}L > 3$ ). These results were obtained with the "random" option as used by MEA and are illustrated for shift priors of (a)  $\gamma = 0.1$ , (b)  $\gamma = 1.0$ , and (c)  $\gamma = 10$ . Blue line is perfect 1:1 relationship; dotted line is fitted regression line for each set of points. Regression slopes for a, b, and c are 0.71, 0.93, and 1.02, respectively. A total of 23 phylogenies and 4374 branches are included in the results shown here.

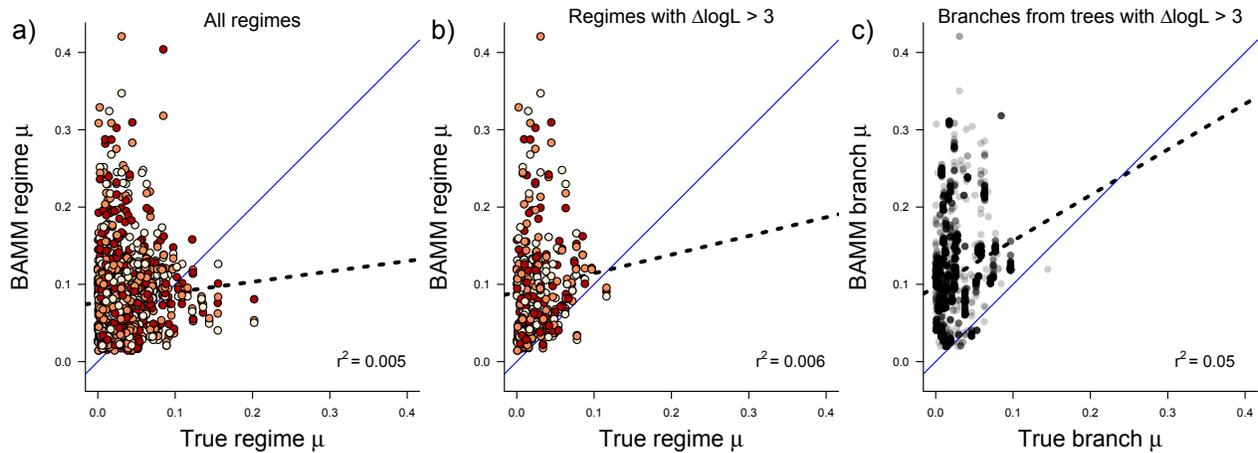


Figure S22: Extinction rates are poorly estimated for individual rate regimes from the MEA “variable rates” phylogenies. (a) Relationship between true extinction rates and BAMM-estimated rates across all 535 rate regimes in the MEA variable-rates dataset, estimated under three different prior parameterizations (colors;  $\gamma = 0.1, 1, \text{ and } 10$ ), and using the BAMM v2.5 default setting for `combineExtinctionAtNodes`. 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 ( $\Delta\text{Log}L < 3$ ) are excluded. (c) Plot of all branch-specific estimates of extinction against the true values for all trees where at least one shift had non-trivial information content ( $\Delta\text{Log}L > 3$ ) from analyses with  $\gamma = 1$  (23 trees; 4374 branches total).

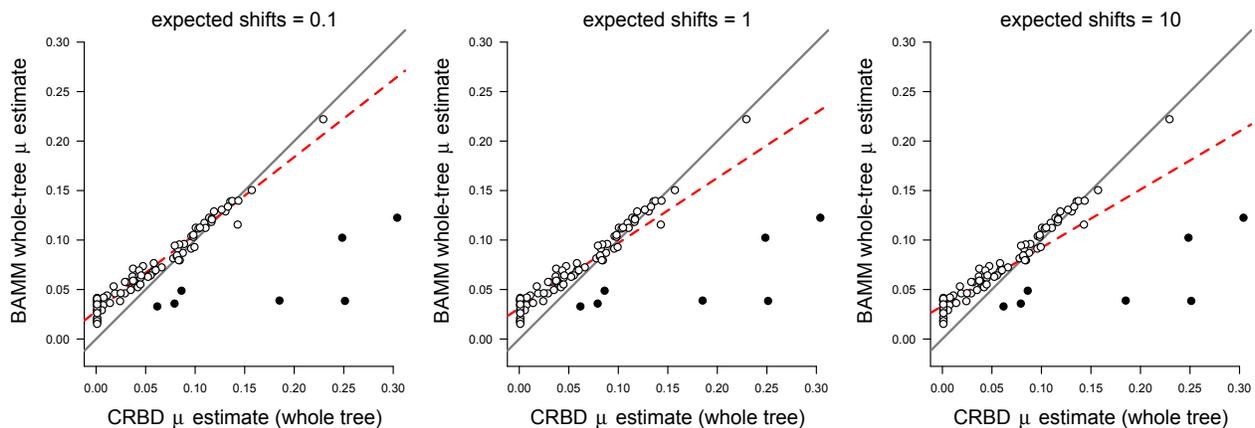


Figure S23: Re-analysis of MEA “variable rates” phylogenies: BAMM extinction estimates are highly correlated with corresponding estimates under a constant-rate birth-death process, across three prior parameterizations (a, b, c) and with the default BAMM v2.5 setting for `combineExtinctionAtNodes`. Whole-tree estimates of the extinction rate were obtained using the BAMMtools function `getCladeRates`, which computes the mean rate across the observed portion of the phylogeny. Open circles denote the 92 phylogenies where BAMM did not detect significant evidence for rate shifts (Bayes factor evidence  $< 20$ ); closed circles denote the 8 “rate variable” phylogenies from MEA where BAMM recovered strong evidence for shifts. As for speciation rates (Main text Figure 11), whole-tree rate estimates with BAMM are highly correlated with the constant-rate estimates when rate heterogeneity is not detected; red line is fitted regression for regimes where rate heterogeneity was not inferred. Gray line denotes perfect isometric relationship.

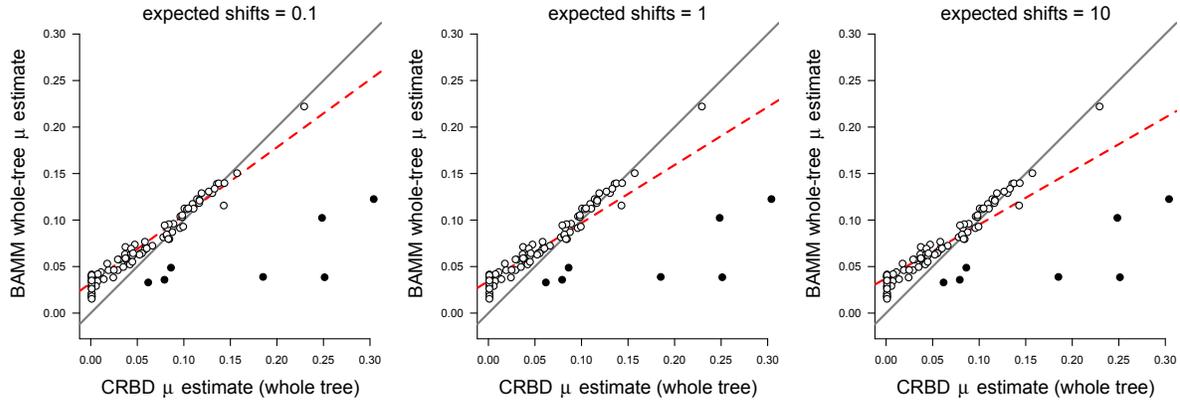


Figure S24: Re-analysis of MEA “variable rates” phylogenies: BAMM extinction estimates are highly correlated with corresponding estimates under a constant-rate birth-death process, across three prior parameterizations (a, b, c) and using the “random” setting for `combineExtinctionAtNodes`, as used by MEA. Whole-tree estimates of the extinction rate were obtained using the BAMMtools function `getCladeRates`, which computes the mean rate across the observed portion of the phylogeny. Open circles denote the 92 phylogenies where BAMM did not detect significant evidence for rate shifts (Bayes factor evidence < 20); closed circles denote the 8 “rate variable” phylogenies from MEA where BAMM recovered strong evidence for shifts. As for speciation rates (Main text Figure 11), whole-tree rate estimates with BAMM are highly correlated with the constant-rate estimates when rate heterogeneity is not detected; red line is fitted regression for regimes where rate heterogeneity was not inferred. Gray line denotes perfect isometric relationship.

## 6 Reliability of diversification analyses under BAMM v2.5 and earlier versions

### 6.1 Simulation procedure

We assessed BAMM’s performance across four releases of the software (v1.0, v2.0, v2.3.1, and v2.5). Given that MEA’s simulated rate-variable phylogenies contain little information with which to infer among-lineage rate variation (see also Section 5.1), we simulated a new set of phylogenies under a forward-time Poisson process with rate shifts, as described in MEA. Our simulator allowed rate shifts to occur on lineages that subsequently became extinct, and phylogenies containing such unobserved shifts were retained for analysis. All simulations thus incorporated a potential source of bias that currently cannot be accommodated in BAMM or other rate-shift models. If the assumption that rate shifts cannot occur on extinct lineages is problematic for inference, we expect to observe biased estimates of evolutionary rate parameters. As in MEA, we assumed that lineages undergo shifts to new diversification rate regimes with parameter  $\eta$ . For simplicity, all simulations assumed that rates of speciation and extinction within a regime were constant in time (e.g., within regimes,  $\lambda(t) = \lambda$ ). Our simulations also encompassed a range of  $\eta : \psi_\lambda$  ratios (0.01 and 0.1), where  $\psi_\lambda$  is the ratio of the Poisson shift rate to the mean of the exponential distribution from which speciation rates for new regimes are drawn (parameter  $\mu_\lambda$  in MEA’s terminology). The minimum ratio we considered ( $\eta = 0.01\psi_\lambda$ ) exceeds estimated  $\eta : \lambda$  ratios for all empirical datasets assessed by MEA. The ratio used by MEA in their “variable rates” simulations was  $\eta = 0.04\psi_\lambda$ ; as such, our analyses bracketed the range of parameterizations used in their analyses (e.g., we explored scenarios where rateshifts were less common and more common than MEA did). To summarize, 200 phylogenies were generated under each of two  $\eta : \psi_\lambda$  ratios; each phylogeny was then analyzed under three different model priors ( $\gamma = 0.1, 1, \text{ and } 100$ ) and under four BAMM versions.

The rate shift simulator is programmed in C++ and is available at <https://github.com/macroevolution/simtree>. Each simulation was initialized with two lineages and with  $\lambda$  and  $\mu$  drawn from exponential distributions with  $\psi_\lambda = 0.1$  and  $\psi_\mu = 0.09$ , respectively. Upon experiencing a rate shift (which occurs with rate  $\eta$ ), the lineage was assigned new  $\lambda$  and  $\mu$  parameters drawn from the same exponential distributions. Each simulation was performed for 100 time units, and we retained trees that included between 10 and 5000 tips. We recognize that any restrictions on tree size can lead to unexpected ascertainment biases (see Fig 7c and Section 5.1 for illustration of ascertainment bias in the MEA variable-rates dataset), but we expect this broad range of acceptable sizes (10, 5000) to have substantially lower ascertainment bias relative to MEA’s simulations (Section 5.1). We analyzed each tree using BAMM versions 1.0, 2.0, 2.3.1 and 2.5 using three different model priors for each version (`expectedNumberOfShifts = 0.1, 1, and 100`). We assumed time-constant BAMM (e.g.,  $\lambda(t) = \lambda$  within regimes). All priors on rate parameters were set using the function `setBAMMpriors(...)` from the `BAMMtools` package. All other parameters were held constant across all analyses and assumed their default values. Each MCMC simulation was performed for 20 million generations, and the first 10% of each simulation was discarded as burnin.

To summarize analyses, we computed mean estimates of speciation and extinction rates for each rate regime under each set of analysis conditions, as in Fig. 12a. We then computed the Pearson correlation between the estimated and true rates across all rate regimes for each analysis condition (BAMM version and  $\gamma$ ). Because we expect BAMM to have low power to infer rates on

small rate regimes, we assessed this correlation explicitly with respect to the number of tips in each rate regime. We also computed a linear regression of the estimated rates on the true rates across regimes, with the slope of this relationship taken as another measure of accuracy. We also tested, for each dataset, whether the best-supported shift scenarios as inferred using either the model with the maximum *a posteriori* probability (MAP model; Rabosky 2014) or selected using Bayes factors (Mitchell and Rabosky 2016) differ between BAMM v2.5 and three earlier versions of the program (v1.0, v2.0, v2.3.1). For each version, we computed the following statistic:

$$\Delta S = M_{v2.5} - M_{vX}$$

where  $M$  is the number of rate shifts inferred using either MAP or Bayes factor-based model selection, and  $vX$  denotes the corresponding BAMM version. Hence, for  $x = v1.0$ , a value of  $\Delta S = 2$  would indicate that BAMM v2.5 selected a model that contained two additional shifts relative to the best-fit model selected by BAMM v1.0. Positive values thus indicate a tendency for BAMM v2.5 to favor more complex models than a specified previous version, and negative values indicate that BAMM v2.5 favors less complex models. A value of  $\Delta S = 0$  indicates that the best model was identical for both BAMM versions.

## 6.2 Results summary

Results are shown in Figure S25 - S30. For the most widely-used prior parameterization ( $\gamma = 1$ ), we found little variation in performance across all four BAMM versions, despite significant differences in the underlying likelihood calculations. When we consider all regimes together from the  $\gamma = 1$  analyses, including root regimes (thus pooling across BAMM versions and  $\eta : \psi_\lambda$  ratios), the overall correlation between true and estimated  $\lambda$  for rate regimes is 0.49. However, this calculation is somewhat misleading, because it assigns small and large shift regimes equal weight (e.g., a rate shift leading to a single tip carries the same weight as a rate shift leading to 50 tips). When we consider only the set of rate regimes with 10 or more tips, this correlation rises to 0.81 and the OLS slope between true and estimated regime rates is 0.783. When we consider only regimes with at least 25 tips, the correlation between true and inferred rates is 0.89, and the corresponding slope is 0.91. Interestingly, BAMM results are more accurate when we exclude root regimes and consider only regimes resulting from rate shifts. For such shift regimes with at least 10 tips, the correlation rises to 0.80 and the slope rises to 0.86. With 25 or more tips, the slope of the relationship for  $\lambda$  is 0.98. Extinction rates are correlated with the generating values, but these correlations are substantially weaker than the corresponding results for speciation, although BAMM v2.5 consistently performs better with a high prior mean on the number of shifts ( $\gamma = 100$ ) relative to earlier versions. The best model selected by BAMM v2.5 is usually identical to the best model selected by all previous versions of BAMM under the most widely used prior parameterization (Figure S29 - S30); there is no evidence that BAMM v2.5 tends to select models of greater or lower complexity than previous versions under  $\gamma = 1$  or when model selection is performed using Bayes factors.

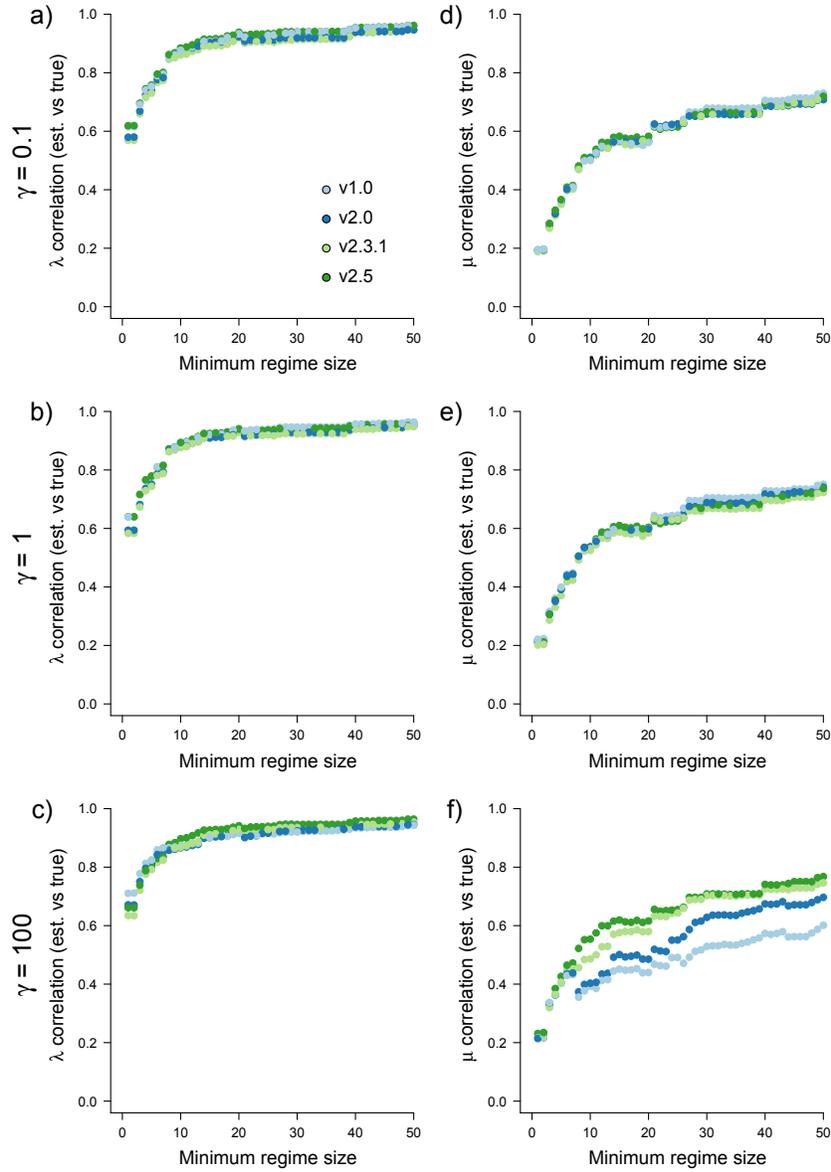


Figure S25: Reliability of diversification rates across four BAMM versions (v1.0, v2.0, v2.3.1, v2.5) when rate shifts are rare relative to speciation events ( $\eta = 0.01\psi_\lambda$ ): Pearson correlations between true and estimated rates across rate regimes. Each panel shows the results of BAMM analyses of 200 phylogenies that collectively included 527 rate regimes, under each of 3 priors ( $\gamma = 0.1, 1, 100$ ). Correlations were computed for all regimes, and then for the set of regimes larger than a particular minimum regime size. For example, with  $x$  (minimum regime size) = 25, the correlation is only computed for the set of regimes that include at least 25 tips. Results for speciation under the three priors are shown in (a, c, d); extinction results are shown in (b, d, f). Speciation correlations are consistently high: for regimes with at least 10 tips, the mean  $\lambda$  correlation is 0.89 ( $\mu = 0.53$ ) and rises to 0.93 for regimes with 25 or more tips (0.63 for  $\mu$ ). We find no differences between versions except for  $\gamma = 100$ , where BAMM v2.5 performs best for  $\mu$ .

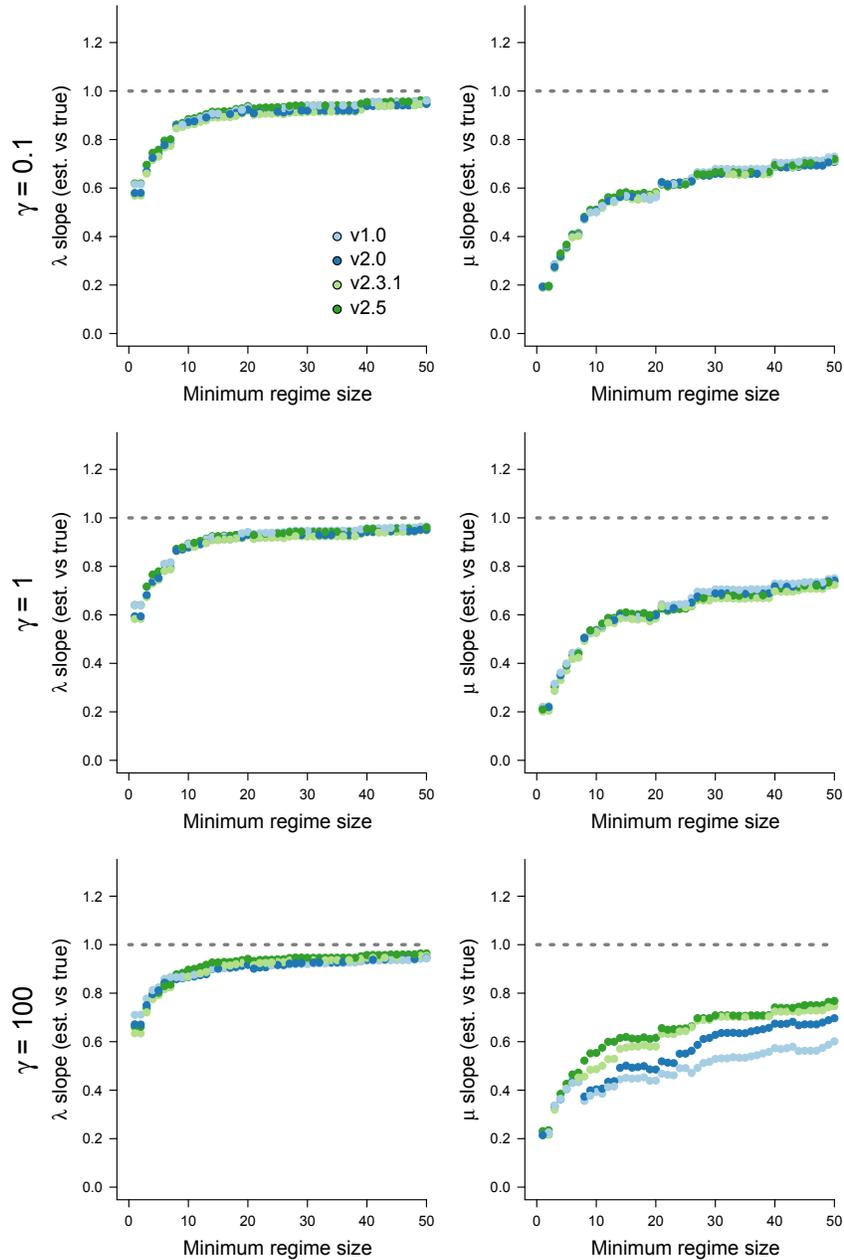


Figure S26: Reliability of diversification rates across four BAMM versions (v1.0, v2.0, v2.3.1, v2.5) under  $\eta = 0.01\psi_\lambda$ : shown are slopes from OLS regressions between true and estimated rates across rate regimes. Same data and summary procedure as in Figure S25. Results for speciation under the three priors are shown in (a, c, d); extinction results are shown in (b, d, f). Speciation slopes are close to 1 for all BAMM versions: for regimes with at least 10 tips, the mean  $\lambda$  slope (true vs. estimated) is 0.97. We find no differences between versions except for  $\gamma = 100$ , where BAMM v2.5 performs slightly better for  $\lambda$  and  $\mu$ .

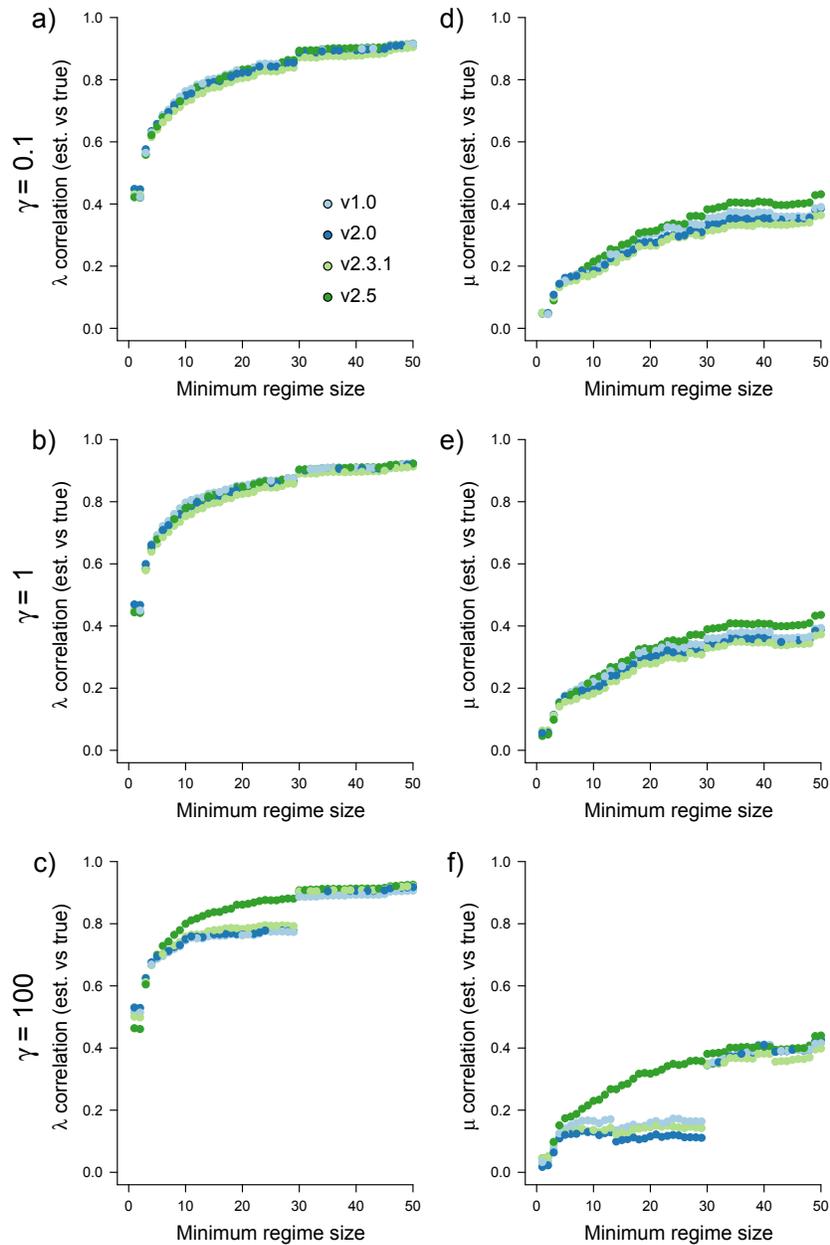


Figure S27: Reliability of diversification rates across four BMM versions (v1.0, v2.0, v2.3.1, v2.5) when rate shifts are infrequent relative to speciation events ( $\eta = 0.1\psi_\lambda$ ): Pearson correlations between true and estimated rates across rate regimes. Each panel shows the results of BMM analyses of 200 phylogenies that collectively included 2108 rate regimes, under each of 3 priors ( $\gamma = 0.1, 1, 100$ ). Correlations were computed for all regimes, and then for the set of regimes larger than a particular minimum regime size. Hence, with  $x$  (minimum regime size) = 25, the correlation is only computed for the set of regimes that include at least 25 tips. Results for speciation under the three priors are shown in (a, c, d); extinction results are shown in (b, d, f). Speciation correlations are consistently high: for regimes with at least 10 tips, the mean  $\lambda$  correlation is 0.78 ( $\mu = 0.20$ ) and rises to 0.86 for regimes with 25 or more tips (0.32 for  $\mu$ ). BMM v2.5 performs better than previous versions for  $\gamma = 100$ , but performance overall is similar across the most widely-used prior parameterization ( $\gamma = 1$ ).

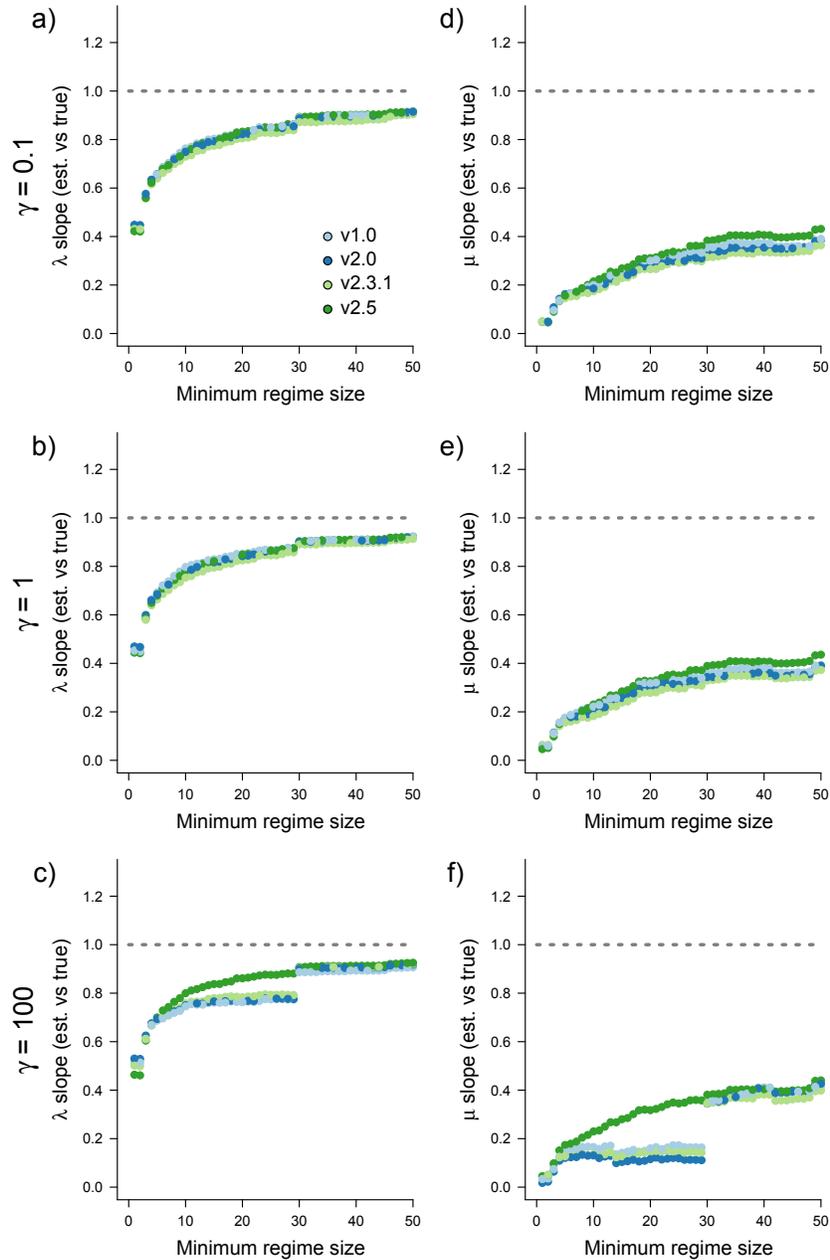


Figure S28: Reliability of diversification rates across four BAMM versions (v1.0, v2.0, v2.3.1, v2.5) when rate shifts are infrequent relative to speciation events ( $\eta = 0.1\psi_\lambda$ ): shown are slopes from OLS regressions between true and estimated rates across rate regimes. Same data and summary procedure as in Figure S27. Results for speciation under the three priors are shown in (a, c, d); extinction results are shown in (b, d, f). For regimes with at least 10 tips, the mean  $\lambda$  slope (true vs. estimated) is 0.71; this slope rises to 0.84 for regimes with 25 or more tips. BAMM v2.5 performs better than previous versions for  $\gamma = 100$ , but performance overall is similar across the most widely-used prior parameterization ( $\gamma = 1$ )

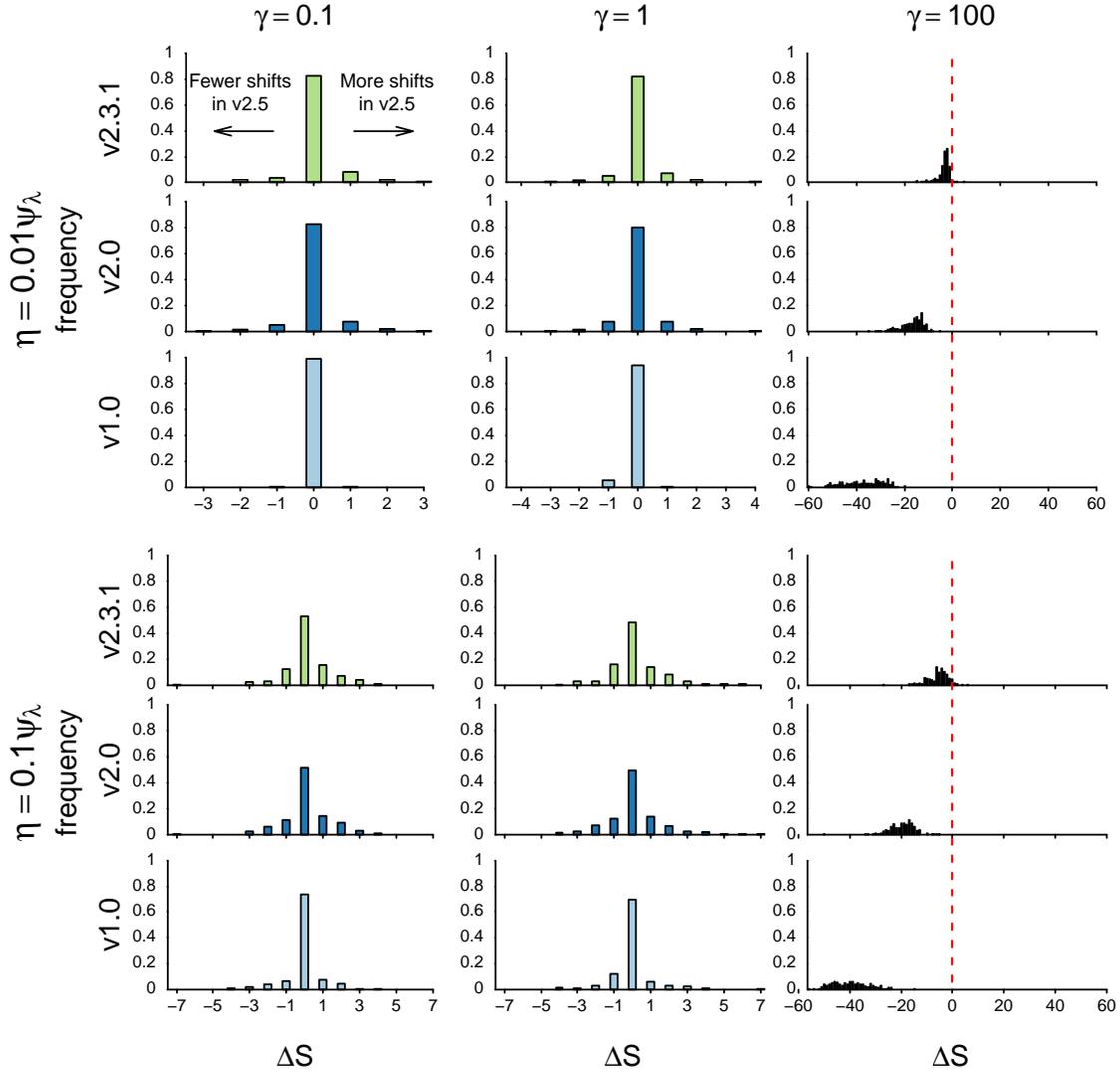


Figure S29: Consistency of inferences about the number of rate shifts across four major versions of BMM, when the model with the highest posterior probability (MAP model) is selected as the overall best model. Each panel shows a set of histograms of the  $\Delta S$  statistic, which is simply the difference in the number of shifts between the model favored by BMM v2.5 relative to the model favored by an earlier BMM version.  $\Delta S > 0$  implies that BMM v2.5 favored a model with more shifts relative to a specified early version, and  $\Delta S < 0$  implies that BMM v2.5 favors a model with fewer shifts. With  $\Delta S = 0$ , the preferred models are identical. Each stacked column of 3 histograms provides the  $\Delta S$  statistics for a given prior  $\gamma$  and shift rate  $\eta$ . Inferences about the number of shifts are generally consistent ( $\Delta S = 0$ ) for all versions and shift frequencies ( $\eta = 0.01\psi_\lambda$ ,  $\eta = 0.1\psi_\lambda$ ) under the most widely-used prior ( $\gamma = 1$ ). However, high prior means ( $\gamma \gg 1$ ) had a strong effect on inferences with BMM versions prior to v2.5. This is attributable to an implementation error that affected BMM versions prior to v2.5 and which magnified the effects of the prior on the posterior; see Mitchell and Rabosky (2016) for details. **These results indicate that empirical results on the number of shifts are unlikely to have been affected by changes to the BMM software, given that  $\gamma = 1$  has been used in the overwhelming majority of published studies.**

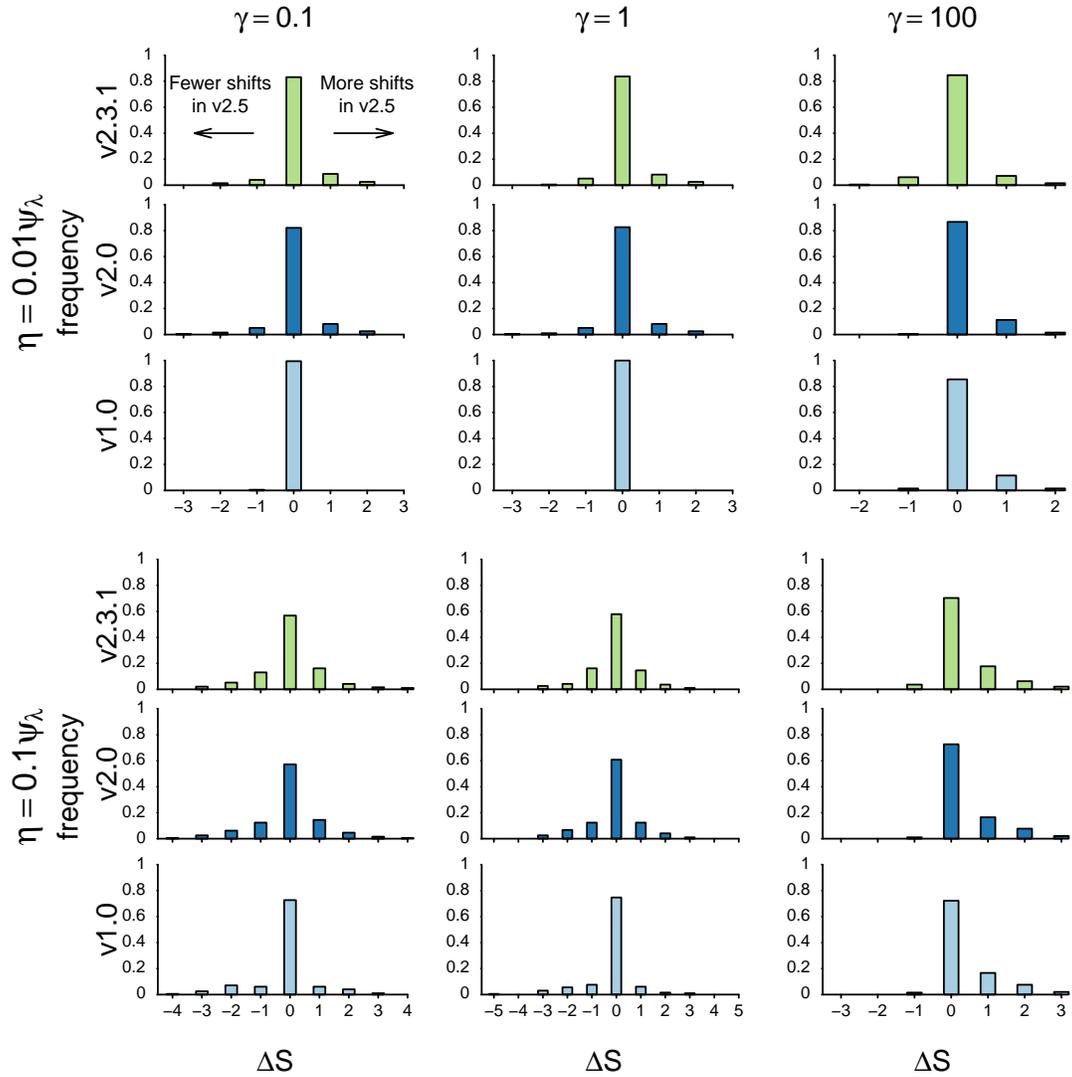


Figure S30: Consistency of inferences about the number of rate shifts across four major versions of BMM, when Bayes factors are used to select the best model. See caption to Figure S29 for details. Inferences about the number of shifts have generally been consistent across BMM versions. There is a slight tendency for BMM v2.5 to find more shifts relative to earlier versions under  $\gamma = 100$ .