Supporting Information

Contents

S1 A detailed description of the BAMM method

BAMM [1] employs a birth-death process model in which a phylogeny, Ψ, is assumed to diversify under a—theoretically infinite—mixture of time-dependent birth-death processes $\boldsymbol{\phi}$ ("regimes"). The objective is to infer the number and location of diversification-rate shifts across branches of the tree, and the diversification-rate parameters for every branch. The model is implemented in a Bayesian framework, such that inference is based on the joint posterior probability distribution of $\boldsymbol{\phi}$ given the observed phylogeny, Ψ, and prior information about the nature of diversification-rate variation, θ. The model has two primary components: 1) a *joint prior distribution* that reflects our belief about the nature of the diversification process *before* we observe the data, Ψ; and 2) a *likelihood function* that defines the probability of observing Ψ given $\boldsymbol{\phi}$, which updates our prior belief to provide the posterior estimate that reflects our belief in the parameter values *after* we observe the available data. The joint posterior distribution of $\boldsymbol{\phi}$ is given by Bayes' theorem:

$$P(\boldsymbol{\phi} \mid \Psi, \theta) \propto P(\Psi \mid \boldsymbol{\phi}, \theta)\, P(\boldsymbol{\phi} \mid \theta) \tag{S1}$$

We describe the details of the underlying model, the likelihood function, and joint prior distribution used by BAMM in the following sections, but introduce the notation here (see Table S1 for a summary). Following Rabosky [1], we refer to the countably infinite set of all possible diversification processes as $\boldsymbol{\phi}$. Each process has a location (lineage index and time), ξ, initial speciation rate, λ, extinction rate, μ, and time-dependence parameter, z; the i^{th} process is therefore specified as $\phi_i = \{\xi_i, \lambda_i, \mu_i, z_i\}$. We refer to the set of all processes as $\boldsymbol{\phi} = \{\boldsymbol{\xi}, \boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{z}\}$. The initial process at the root of the tree is called ϕ_R. The nature of diversification-rate variation is described by a vector of hyperparameters $\theta = \{\mu_\lambda, \mu_\mu, \mu_z, \sigma_z, \Lambda\}$: μ_λ and μ_μ specify the mean of the exponential priors for the speciation and extinction rates for a process, respectively; μ_z and σ_z specify the mean and the standard deviation of normal prior for the time-dependence parameter z, respectively; and Λ is the expected number of derived processes (*i.e.*, those beyond the root process) on the phylogeny.

For notational clarity, we introduce a set of convenience functions: $\lambda(i,t)$ is the time-dependent speciation rate of the process for lineage i at time t, and $\mu(i,t)$ is the extinction rate of the process for lineage i at time t. For example, if lineage i belongs to process j at time t, then these functions have values:

$$\begin{aligned}\lambda(i,t) &= \lambda_j \exp[z_j \times (t - \xi_j[2])] \\ \mu(i,t) &= \mu_j,\end{aligned}$$

where $\xi_j[2]$ is the initial time of the j^{th} process. The time-dependence function $\lambda(i,t)$ corresponds to Equation 1 of Rabosky (2014).

We note that Rabosky [1] parameterizes the expected number of non-root processes operating on the tree, Λ, which we refer to as the 'Λ-parameterization'. Under this parameterization, the diversification-process parameters are $\theta_\Lambda = \{\mu_\lambda, \mu_\mu, \mu_z, \sigma_z, \Lambda\}$. For reasons that we describe in Section S1.1, it is generally more natural to parameterize the rate at which new processes arise, η, which we refer to as the 'η-parameterization'. Under this parameterization, the diversification-process parameters are $\theta_\eta = \{\mu_\lambda, \mu_\mu, \mu_z, \sigma_z, \eta\}$.

S1.1 The compound-Poisson birth-death process

The rooted, ultrametric phylogeny, Ψ, as well as the processes operating on the tree, $\boldsymbol{\phi}$, are assumed to be generated by a compound-Poisson birth-death stochastic-branching process (CPBDP):

$$(\Psi, \boldsymbol{\phi}) \sim \text{CPBDP}(\theta_\eta)$$

The compound-Poisson birth-death process begins at time 0 with two lineages. A diversification process is drawn at the root of the tree, $\phi_R = \{\xi_R, \lambda_R, \mu_R, z_R\}$, with parameters chosen according to θ_η (*i.e.,* according to their prior distributions, see Section S1.1.1). Initially, each lineage belongs to the process ϕ_R. Each lineage i gives rise to exactly one new lineage with rate $\lambda(i,t)$, goes extinct with rate $\mu(i,t)$, and evolves a new diversification process with rate η. When a lineage speciates, both daughter lineages inherit the diversification-process parameters of their ancestor. When lineage i evolves a new diversification process at time t, a new diversification process, $\phi_j = \{\xi_j, \lambda_j, \mu_j, z_j\}$, is chosen according to θ_η. This process continues until the present at time T.

Table S1: BAMM model parameters and their interpretation

Parameter	Interpretation
Ψ	Phylogenetic tree with divergence times.
γ	Prior mean of the Poisson rate Λ.
Λ	Prior mean of the Poisson-distributed number k of shift events.
k	Number of diversification-rate shifts.
$\boldsymbol{\xi}$	Vector of locations of the diversification-rate shifts.
ξ_i	Location of the i^{th} diversification-rate shift.
$\boldsymbol{\lambda}$	Vector of speciation rates per process.
λ_i	Speciation rate of the i^{th} process.
$\boldsymbol{\mu}$	Vector of extinction rates per process.
μ_i	Extinction rate of the i^{th} process.
$\boldsymbol{z}$	Vector of the speciation-rate modifiers per process.
z_i	Speciation-rate modifier of the i^{th} process.

A realization of this branching process potentially contains extinct lineages; following Nee *et al.* [2], we refer to (Ψ, $\boldsymbol{\phi}$) outcomes of the "complete" compound-Poisson birth-death process (Figure S1A). Deleting the extinct lineages results in the so-called "reconstructed" compound-Poisson birth-death process (Figure S1B). Removing the extinct lineages also has the potential to remove entire diversification processes, which has important implications described below (see Section S1.1.2). As the study tree Ψ generally includes only extant species, we assume it is generated by the reconstructed compound-Poisson birth-death process.

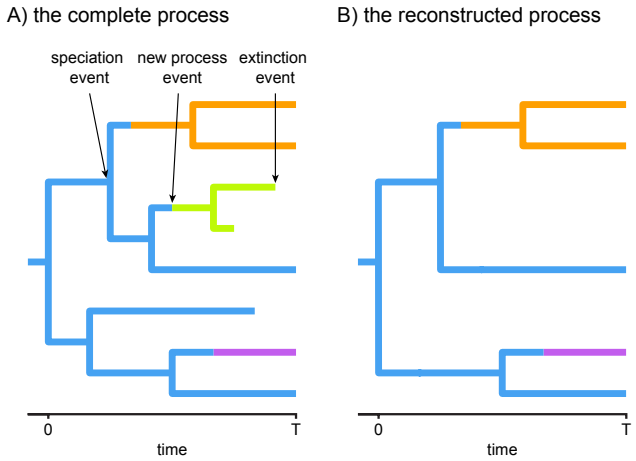

Figure S1: A realization of the complete (A) and reconstructed (B) compound-Poisson-birth-death process.

S1.1.1 Sampling new diversification processes

At certain times during the compound-Poisson-birth-death process (specifically, when the process is initiated and each time a new diversification process arises), new diversification-process parameters must be chosen according to θ_η. When this occurs, the parameters for diversification process i are sampled according to the following rules:

$$\lambda_i \sim \text{Exponential}(1/\mu_\lambda)$$
$$\mu_i \sim \text{Exponential}(1/\mu_\mu)$$
$$z_i \sim \text{Normal}(\mu_z, \sigma_z)$$

The values of ξ_i are determined by the branch and time on which the new process arises.

S1.1.2 Extinct diversification processes

Importantly, the deletion of extinct species can result in the deletion of entire diversification processes (*e.g.,* the green process in Figure S1A), which confounds the interpretation of the expected number of diversification processes, Λ. Indeed, the Λ-parameterization does not yield a well-defined stochastic process that can be used as a generative model (*i.e.,* we cannot simulate data under this parameterization of the process) because there is no obvious way to transform θ_Λ into θ_η. It is tempting to expect that $\Lambda = \eta L$, where L is the length (*i.e.*, the sum of all branch lengths) of the reconstructed tree, under the assumption that new diversification processes arise as a (homogeneous) Poisson process over the reconstructed tree. However, this is incorrect. For example, consider a process that begins at $t = 0$ with $\lambda_R = 1, \mu_R = 0$ (a pure-birth, or Yule, process). Further, assume that each time a new diversification process ϕ_j arises, $\lambda_j = 0$ and $\mu_j \to \infty$ (an extreme pure-death process). In this case, each time a new diversification process arises, the lineage bearing it will immediately go extinct (since $\mu_j \to \infty$). As a result, the reconstructed process in this case can contain no diversification processes besides ϕ_R; here it is obvious that $\Lambda = 0 \neq \eta L$. Unfortunately, the relationship between θ_η and the expected number of diversification processes is complex and unknown, and there is no obvious way to specify θ_η such that the expected number of processes is Λ.

While seemingly innocuous, the inability to simulate data under the parameterization used by `BAMM` precludes the application of useful software diagnostics that depend on simulating data from the prior [*e.g.,* 3], and obscures characterization of the frequentist properties of the joint posterior distribution [*e.g.,* 4]. More critically, the deletion of diversification processes leads to a serious error in computing the likelihood, which we discuss in more detail in Section S1.2.3.

S1.2 Likelihood function

We introduce the likelihood function by analogy to the `BiSSE` (Binary State Speciation and Extinction) model developed in Maddison et al. [5], from which `BAMM` borrows heavily. We note that the diversification processes of `BAMM` are equivalent to the state-specific branching processes of the `BiSSE` model, and that the birth-death process model used in `BAMM` is merely a more elaborate version of the `BiSSE` branching process. Specifically, the `BiSSE` model assumes a finite number of processes (two; one for each of the binary states), where the process to which each extant species belongs is known (indeed, it is an observed discrete trait), whereas the `BAMM` model assumes a countably infinite number of processes, where the membership of species to processes is unknown.

The general approach adopted by both `BiSSE` and `BAMM` models is to derive a set of ordinary differential equations (ODEs) that describe how the probability of observing a descendant clade changes along a branch in the observed phylogeny. Each equation in this set describes how the probability of observing the clade changes through time if it is in a particular process over that time period; collectively, these equations are called $\frac{\mathrm{d}D_{N,i}(t)}{\mathrm{d}t}$, where the subscript N refers to the descendant clade and the subscript i refers to i^{th} process.

Computing the likelihood proceeds by establishing an initial value problem: in principle, if we know the probabilities of observing a lineage at some specific time (*e.g.*, the present), and know how those probabilities change over time (described by the ODEs), then we can compute the probabilities of observing those lineages at some earlier time (*e.g.*, the root). Assuming that there are exactly k possible processes, we initialize k probabilities at each tip in the phylogeny; we then compute how each of those k probabilities changes down each branch in the tree using the above set of k ODEs. At each node in the tree, we take the product of each of the k probabilities for the descendants of that node (multiplied by the instantaneous speciation rate for each of the k processes to account for the observed speciation event at the node) as the initial values for the ancestral branch subtending that node. Proceeding in this way down the tree results in a set of k probabilities at the root; these k probabilities represent the probability of observing the phylogeny conditional on the root being in each of the processes (*i.e.*, the i^{th} conditional probability is the probability of observing the tree given that the root is in process i). The overall likelihood of the tree is a weighted average of the k probabilities at the root, where the weighting scheme represents the assumed probability that the root was in each of the k processes.

As with all birth-death process models, special care must be taken to account for the possibility of extinction. Specifically, the above ODEs must accommodate lineages that may arise along each branch in the tree that subsequently go extinct before the present (and so are unobserved). This requires a second

set of k ODEs, $\frac{\mathrm{d}E_i(t)}{\mathrm{d}t}$, which define how the probability of extinction in process i changes over time. These ODEs must be solved to compute the differential equations $\frac{\mathrm{d}D_{N,i}(t)}{\mathrm{d}t}$, as we will demonstrate when we derive both sets of equations in the following sections.

This framework therefore requires four distinct pieces of information to compute the likelihood of the data:

1. A set of ordinary differential equations describing how the probability of the data (observed lineages) changes through time, $\frac{\mathrm{d}D_{N,i}(t)}{\mathrm{d}t}$.
2. A set of ordinary differential equations describing how the extinction probability of unobserved (extinct or unsampled) lineages changes through time, $\frac{\mathrm{d}E_i(t)}{\mathrm{d}t}$.
3. An appropriate set of initial conditions.
4. An appropriate weighting scheme for the root probabilities.

In the following sections we detail how each of these components is determined for increasingly complex birth-death process models.

S1.2.1 Finite state space of diversification processes

Consider a time-independent birth-death process with two possible processes, $\phi_0 = \{\lambda_0, \mu_0\}$ and $\phi_1 = \{\lambda_1, \mu_1\}$. Further, assume that a lineage changes its process at rate q (for simplicity, we will assume an equal rate of change between the two process). We define $D_{N,0}(t)$ as the probability of observing lineage N descending from a particular branch at time t, given that the process at that point is ϕ_0 (with rate parameters λ_0, and μ_0). To compute the probability of observing the lineage at some earlier point, $D_{N,0}(t + \Delta t)$, we enumerate all possible events that could occur within the interval Δt. Assuming that Δt is small—so that the probability of any two events occurring in the interval is negligible—there are four possible scenarios within the interval: (1) nothing happens in the interval; (2) the process changes $0 \rightarrow 1$; (3) a speciation event occurs and the left descendant subsequently goes extinct before the present, or; (4) a speciation event occurs and the right descendant subsequently goes extinct before the present. We can thus compute $D_{N,0}(t + \Delta t)$ as (see Maddison et al. [5] and FitzJohn et al. [6] for a more complete elucidation):

$$\begin{aligned} D_{N,0}(t+\Delta t) = (1-\mu_0\Delta t)\times & \quad \text{in all cases, no extinction of the observed lineage} \\ [(1-q\Delta t)(1-\lambda_0\Delta t)D_{N,0}(t) & \quad \text{Case (1) nothing happens} \\ + q\Delta t(1-\lambda_0\Delta t)D_{N,1}(t) & \quad \text{Case (2) process change but no speciation} \\ + (1-q\Delta t)\lambda_0\Delta t E_0(t)D_{N,0}(t) & \quad \text{Case (3) no process change, speciation, extinction} \\ + (1-q\Delta t)\lambda_0\Delta t E_0(t)D_{N,0}(t)] & \quad \text{Case (4) no process change, speciation, extinction} \end{aligned} \tag{S2}$$

A matching equation can be written down for $D_{N,1}(t + \Delta t)$.

Define $E_0(t)$ as the probability that a lineage in ϕ_0 at time t goes extinct before the present. To determine the extinction probability at an earlier point, $E_0(t + \Delta t)$, we can again enumerate all the possible events in the interval Δt: (1) the lineage goes extinct within the interval; (2) the lineage neither goes extinct nor speciates, resulting in a single lineage that must eventually go extinct before the present; (3) the lineage neither goes extinct nor speciates, but there is a process change, resulting in a single lineage that must go extinct before the present, or; (4) the lineage speciates in the interval, resulting in *two* lineages that must eventually go extinct before the present.

$$\begin{aligned} E_0(t+\Delta t) = \mu_0\Delta t + & \quad \text{Case (1) extinction in the interval} \\ (1-\mu_0\Delta t)\times & \quad \text{no extinction in the interval and ...} \\ [(1-q\Delta t)(1-\lambda_0\Delta t)E_0(t) & \quad \text{Case (2) nothing happens, but subsequent extinction} \\ + q\Delta t(1-\lambda_0\Delta t)E_1(t) & \quad \text{Case (3) process change and subsequent extinction} \\ + (1-q\Delta t)\lambda_0\Delta t E_0(t)^2] & \quad \text{Case (4) speciation and subsequent extinctions} \end{aligned} \tag{S3}$$

Again, a matching equation $E_1(t+\Delta t)$ can be written down.

We can expand the `BiSSE` model to accommodate an arbitrary number of processes, k, by writing a set of k difference equations $D_{N,0}(t+\Delta t), D_{N,1}(t+\Delta t), \ldots, D_{N,k}(t+\Delta t)$:

$$\begin{aligned} D_{N,i}(t+\Delta t) = (1-\mu_i\Delta t)\times & \\ [(1-\sum_{j\neq i}^{k} q\Delta t)(1-\lambda_i\Delta t)D_{N,i}(t) & \\ + (1-\lambda_i\Delta t)\sum_{j\neq i}^{k} q\Delta t D_{N,j}(t) & \\ + 2(1-\sum_{j\neq i}^{k} q\Delta t)\lambda_i\Delta t E_i(t) D_{N,i}(t)] & \end{aligned} \tag{S4}$$

along with $E_0(t+\Delta t), E_1(t+\Delta t), \ldots, E_k(t+\Delta t)$:

$$\begin{aligned} E_i(t+\Delta t) = \mu_i\Delta t + & \\ (1-\mu_i\Delta t)\times & \\ [(1-\sum_{j\neq i}^{k} q\Delta t)(1-\lambda_i\Delta t)E_i(t) & \\ + (1-\lambda_i\Delta t)\sum_{j\neq i}^{k} q\Delta t E_j(t) & \\ + (1-\sum_{j\neq i}^{k} q\Delta t)\lambda_i\Delta t E_i(t)^2] & \end{aligned} \tag{S5}$$

It is possible to derive differential equations from the difference equations S4 and S5 (see Maddison et al. [5] for the two-process case and FitzJohn [7] for the k-process case). For the general k-process case, the differential equations are:

$$\begin{aligned} \frac{\mathrm{d}D_{N,i}(t)}{\mathrm{d}t} &= -\left(\lambda_i + \mu_i + \sum_{j\neq i}^{k} q\right) D_{N,i}(t) + 2\lambda_i E_i(t) D_{N,i}(t) + \sum_{i\neq j}^{k} q D_{N,j}(t) \\ \frac{\mathrm{d}E_i(t)}{\mathrm{d}t} &= -\left(\lambda_i + \mu_i + \sum_{j\neq i}^{k} q\right) E_i(t) + \lambda_i E_i(t)^2 + \mu_i + \sum_{i\neq j}^{k} q E_j(t) \end{aligned}$$

Initial probabilities are assigned according to the observed discrete states: if species i has state j, then $D_{i,j}(0) = 1$ for the observed state, and $D_{i,j}(0) = 0$ for all other ($\neq j$) states. If the state is not observable, then $D_{i,j}(0) = 1$ for all j, since all states have probability 1 of producing the observation; this is analogous to the treatment of missing or ambiguous states in conventional phylogenetic likelihood calculation, *c.f.*, [8]. Initial extinction probabilities are set to 0 (since there is no time for extinction to occur at the present). Root probabilities are either weighted using equal probabilities (uniformly), by a vector of pre-defined root stationary probabilities (informative), or by the stationary distribution of the model, to compute the overall likelihood of the data.

S1.2.2 Infinite state space of diversification processes

Rabosky [1] extends the above logic to the case where the number of processes is infinite, and the membership of each extant species to processes is unknown. In this case, $k = \infty$, and the rate of change away from the current process (and towards any other process) is η. In principle, we might extend equations S4 and S5 to integrate over all possible processes by setting $k = \infty$ and proceed in the usual manner:

$$\begin{aligned} D_{N,i}(t+\Delta t) = (1-\mu_i\Delta t)\times & \\ & [(1-\eta\Delta t)(1-\lambda_i\Delta t)D_{N,i}(t) \\ & + (1-\lambda_i\Delta t)\sum_{j\neq i}^{\infty}\eta\Delta t D_{N,j}(t) \\ & + 2(1-\eta\Delta t)\lambda_i\Delta t E_i(t)D_{N,i}(t)] \end{aligned} \tag{S6}$$

$$\begin{aligned} E_i(t+\Delta t) = \mu_i\Delta t + & \\ (1-\mu_i\Delta t)\times & \\ [(1-\eta\Delta t)(1-\lambda_i\Delta t)E_i(t) & \\ + (1-\lambda_i\Delta t)\sum_{j\neq i}^{\infty}\eta\Delta t E_j(t) & \\ + (1-\eta\Delta t)\lambda_i\Delta t E_i(t)^2] & \end{aligned} \tag{S7}$$

However, computing the infinite sums in equations S6 and S7 is intractable.

S1.2.3 Infinite state space of diversification processes, with data-augmented histories

Rabosky [1] circumvents the issues associated with an infinite number of possible processes by using data augmentation. In the data-augmentation scheme, likelihood computations are performed assuming the history of diversification processes generating the phylogeny are known. Such a history has a finite number $(k+1)$ of processes, $\boldsymbol{\phi} = \{\phi_R, \phi_1, \ldots, \phi_k\}$, with $\phi_i = \{\xi_i, \lambda_i, \mu_i, z_i\}$. Histories, as well as their associated locations and diversification parameters, are then sampled with reversible-jump Markov chain Monte Carlo (described in Rabosky [1]), which effectively (numerically) integrates over all possible histories, visiting each history in proportion to its posterior probability under the model. The data-augmentation approach has a number of advantages:

1. A single probability, $D_{N,i}(0) = 1$, needs to be initiated at each tip N in the tree (*i.e.,* conditional on the history, the process that each tip belongs to is known).
2. A single probability, $D_{N,i}(t)$, needs to be integrated down each branch of the tree.
3. A single probability, $D_{N,R}(\text{root})$, exists at the root of the tree, so $D_{N,R}(\text{root})$ is the likelihood of the entire tree, conditional on the history.
4. The MCMC approximates posterior estimates of the number of diversification processes, as well as their locations and branch-specific diversification-rate parameters.

Since the distribution of processes over the tree is assumed known, we can condition the terms in equation S6 on zero additional processes by setting $\eta\Delta t = 0$:

$$\begin{aligned} D_{N,i}(t+\Delta t) = (1-\mu_i\Delta t)\times & \\ [(1-\lambda_i\Delta t)D_i(t) + 2\lambda_i\Delta t E_i(t)D_i(t)] & \end{aligned} \tag{S8}$$

The resulting differential equation is:

$$\frac{\mathrm{d}D_{N,i}(t)}{\mathrm{d}t} = -(\lambda_i+\mu_i)D_{N,i}(t) + 2\lambda_i E_i(t)D_{N,i}(t), \tag{S9}$$

which is equation (10) from Rabosky [1].

Unfortunately, the history of diversification processes on the tree *does nothing* to simplify equation S7: the data-augmented processes $\boldsymbol{\phi}$ do not describe how diversification processes have varied on extinct side-branches of the tree (*i.e.*, realizations of the process obviously cannot be mapped onto unobserved lineages). It is not possible to perform the infinite sum in equation S7 analytically, although the extinction probability can be approximated by computationally expensive Monte Carlo methods (we describe this approach in Section S2). Rather than use a correct extinction probability that corresponds to the model, Rabosky [1] assumes that extinct lineages evolve at constant speciation- and extinction-rates (*i.e.,* they cannot generate new diversification processes). The resulting (incorrect) quantity $\mathrm{d}E_i(t)/\mathrm{d}t$ used by `BAMM` is:

$$\frac{\mathrm{d}E_i(t)}{\mathrm{d}t} = \mu_i - (\mu_i + \lambda_i)E_i(t) + \lambda_i E_i(t)^2, \tag{S10}$$

which implies that extinct lineages evolve under the diversification process they had at time t (*i.e.*, thereby disallowing subsequent diversification-rate shifts along these extinct lineages).

`BAMM` uses numerical integration to compute equations S8 and S10 down each branch in the observed phylogeny; at the root of the tree, the probability of the data, $D_{N,R}(\text{root})$ is taken as the likelihood of the data:

$$P(\Psi \mid \boldsymbol{\phi}, \theta) = D_{N,R}(\text{root}), \tag{S11}$$

which is the likelihood function used by equation S1.

S1.3 Joint prior distribution

Computing the conditional distribution of diversification processes given the phylogeny Ψ requires both a likelihood function (described above), as well as a prior distribution for each parameter in the model. Accordingly, we need to specify a prior distribution for the set of diversification processes, $\boldsymbol{\phi}$ (technically, this is a *joint* prior probability distribution, since each element of $\boldsymbol{\phi}$ is multivariate).

Rabosky [1] assumes the joint prior distribution on diversification processes is a compound Poisson process (CPP) prior model. The CPP model specifies the prior distribution on the number of diversification processes, k, the locations of the diversification-rate shifts, $\boldsymbol{\xi}$, and the parameters of the birth-death process model, $\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{z}$.

Under the CPP prior model, the diversification process is inherited identically over ancestor-descendant branches between events. These events—diversification-rate shifts—arise along the tree according to a Poisson process with rate η (*i.e.,* the waiting times between diversification-rate shifts are exponentially distributed with rate η). Under a Poisson process with rate η evolving over a tree of length L, the number of processes (not counting the root process), k, over the entire tree follows a Poisson distribution with rate $\Lambda = \eta L$. Additionally, Rabosky [1] assumes that rate shifts are uniformly distributed across the entire tree:

$$\begin{aligned} k &\sim \text{Poisson}(\Lambda) \\ \xi_i &\sim \text{Uniform}(0, L) \end{aligned}$$

When a diversification-rate shift occurs, a new birth-death process is specified by drawing diversification-rate parameters from their corresponding prior distributions. By default, those prior distributions are:

$$\begin{aligned} \lambda_i &\sim \text{Exponential}(1/\mu_\lambda) \\ \mu_i &\sim \text{Exponential}(1/\mu_\mu) \\ z_i &\sim \text{Normal}(\mu_z, \sigma_z), \end{aligned}$$

with $\mu_z = 0$ by default.

To relax the assumption that diversification-rate shifts occur at a fixed rate, Λ, Rabosky [1] constructs a hierarchical model in which the parameter Λ is itself a random variable. Specifically, Λ is drawn from an exponential distribution with an expected value of γ:

$$\Lambda \sim \text{Exponential}(1/\gamma)$$

S1.3.1 Analytical solution for the prior number of diversification-rate shifts

Under the constructed hierarchical model, the number of events, k, and the expected number of events, Λ, are both random variables. Therefore, the prior probability of k events (which `BAMM` uses to compute Bayes factors for hypothesis testing) depends on γ, $P(k \mid \gamma)$. The prior probability distribution for number of events k is approximated by simulation in `BAMM` (these numerical methods are described in Section S1.3.2). Here, we present an analytical solution for prior number of diversification-rate shifts.

Conditional on the expected number of diversification processes, Λ, the number of diversification processes, k, follows a Poisson distribution with rate Λ:

$$k \mid \Lambda \sim \text{Poisson}(\Lambda)$$

The number of expected events itself follows an exponential distribution with expectation γ. For simplicity, we replace the expectation of the exponential distribution with the rate parameter, β ($\beta = 1/\gamma$):

$$\Lambda \sim \text{Exponential}(\beta = 1/\gamma)$$

Additionally, we can replace the exponential distribution with a gamma distribution with shape parameter $\alpha = 1$:

$$\Lambda \sim \text{Gamma}(\alpha = 1, \beta)$$

The joint prior distribution of k and Λ is thus:

$$\begin{aligned} P(k, \Lambda \mid \beta) &= P(k \mid \Lambda) P(\Lambda \mid \beta) \\ &= \frac{\Lambda^k \mathrm{e}^{-\Lambda}}{k!} \frac{\beta^\alpha}{\Gamma(\alpha)} \Lambda^{\alpha-1} \mathrm{e}^{-\beta\Lambda} \end{aligned}$$

The distribution $P(k \mid \beta)$ is obtained by integrating over Λ:

$$\begin{aligned} P(k \mid \beta) &= \int_0^\infty \frac{\Lambda^k e^{-\Lambda}}{k!} \frac{\beta^\alpha}{\Gamma(\alpha)} \Lambda^\alpha e^{-\beta\Lambda} d\Lambda \\ &= \binom{k+\alpha-1}{k} \left(\frac{1}{1+\beta}\right)^k \left(\frac{\beta}{1+\beta}\right)^\alpha. \end{aligned}$$

This is the negative binomial, or gamma-Poisson, distribution [9]. When $\alpha = 1$ (the exponential distribution used in `BAMM`), this simplifies to:

$$\begin{aligned} P(k \mid \beta) &= \left(\frac{1}{1+\beta}\right)^k \left(\frac{\beta}{1+\beta}\right) \\ P(k \mid \gamma) &= \left(\frac{\gamma}{1+\gamma}\right)^k \left(\frac{1}{1+\gamma}\right), \end{aligned}$$

which implies

$$k \sim \text{Geometric}\left(\frac{1}{1+\gamma}\right).$$

The mean of this geometric distribution is

$$E(k) = \frac{1-p}{p} = \frac{1 - \frac{1}{1+\gamma}}{\frac{1}{1+\gamma}} = \gamma.$$

Interestingly, while $E(k) = \gamma$, the mode$(k) = 0$ for all γ: the prior mode on the number of events is 0 regardless of the value of γ. For this reason, the MAP (= mode of the posterior probability distribution) is an unfortunate choice for assessing the sensitivity of inferences to the assumed γ prior when the data were generated by a constant-rate birth-death process (*c.f.*, Figures 5, S18).

S1.3.2 Numerical approximation of the prior number of diversification-rate shifts

The ability to derive an analytical solution for the prior distribution on the number of diversification-rate shifts (see Section S1.3.1) provides two main benefits: (1) an analytical solution provides an efficient means for calculating the exact prior probabilities, and so obviates the need to approximate these probabilities using numerical methods (MCMC simulation), and; (2) an analytical solution allows us to validate the numerical methods (MCMC algorithms) implemented in `BAMM` that are used to estimate the target (*i.e.*, the joint prior and posterior probability) distributions. This latter point is particularly important; the analytical prior provides a means to establish that the MCMC algorithms implemented in `BAMM` are correct, and therefore allows us to rule out simple implementation errors (*i.e.*, programming bugs) as the cause of the pathologies that we have identified with the `BAMM` method.

`BAMM` provides two methods for approximating the marginal prior distribution of the number of diversification processes, k. The first approach, which is invoked using the `sampleFromPriorOnly` command, relies on the same MCMC algorithm used by `BAMM` to approximate the joint posterior distribution of the parameters, but simply forces the log-likelihood function to always return 0 so that it simulates the prior. The second approach, which is invoked using the `fastSimulatePrior` command, relies on an entirely separate, stand-alone MCMC algorithm that simply samples the number of diversification processes under the prior.

The marginal prior distribution of the number of diversification processes estimated using these two algorithms conforms precisely to the analytical prior (Figure S2). The correspondence of the simulated and analytical priors confirms that the MCMC algorithms implemented in the (modified version of) `BAMM` can correctly target the intended joint (prior and posterior) probability distribution. However, we show below that the intended joint prior probability distribution is statistically incoherent (Section S1.3.3).

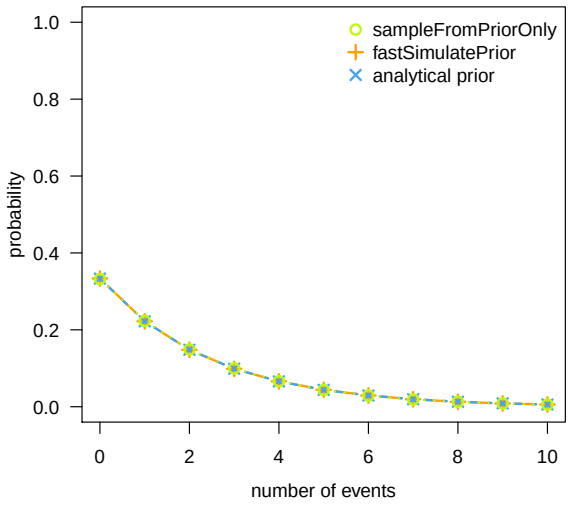

Figure S2: Simulated (numerically estimated) and analytical prior distributions of k under the `BAMM` model. `BAMM` provides several approaches for estimating the prior distribution on k (*i.e.*, the prior number of diversification-rate shifts). The first (and conventional) approach for estimating the prior distribution uses the same MCMC algorithm used to estimate the posterior distribution. This approach is invoked using the `sampleFromPriorOnly` command (green symbols). The second approach for estimating the prior distribution on k relies on a stand-alone MCMC algorithm (`fastSimulatePrior`; orange symbols); prior probabilities used to compute branch-specific Bayes factors (to identify significant diversification-rate shifts) are estimated using the `fastSimulatePrior` algorithm. These two numerical estimates of the prior distribution on the number of diversification-rate shifts are compared to the analytical prior (blue symbols; described in Section S1.3.1). The correspondence between the estimated and analytical priors confirms that the MCMC algorithm implemented in (the corrected version of) `BAMM` is able to successfully target the intended joint prior probability distribution. Accordingly, the pathologies that we have identified in this study cannot be attributed to prosaic implementation errors.

S1.3.3 The distribution of the location of diversification-rate shifts is not uniform

Under a birth-death process, events on different lineages occur independently. Thus, to show that the distribution of diversification-rate shifts is not uniform when conditioning on the tree length (the sum of all branch lengths for a given tree), we need only show that the distribution for a single branch is not uniform when conditioning on the branch length. Moreover, the Yule process [10] is a special case of the birth-death process (with $\mu = 0$); accordingly, demonstrating that the distribution is not uniform for the Yule process demonstrates that it is not uniform in general for the birth-death process. We define η as the rate at which speciation-rate shifts occur in a Yule process, where y is the number of speciation-rate shifts after time t. The initial speciation rate of the Yule process is λ_0 and the speciation rate after speciation-rate shift i is λ_i. The joint probability that exactly one speciation-rate shift ($y = 1$) from λ_0 to λ_1 occurs at time x on a branch of length t and no speciation events occur on $(0, t)$ is:

$$f(x, y = 1) = \eta \mathrm{e}^{-(\lambda_0+\eta)x} \mathrm{e}^{-(\lambda_1+\eta)(t-x)},$$

where we assume that η, λ_0 and λ_1 are known parameters. The marginal probability density of $y = 1$ is

$$f(y = 1) = \int_0^t \eta \mathrm{e}^{-(\lambda_0+\eta)x} \mathrm{e}^{-(\lambda_1+\eta)(t-x)} \mathrm{d}x = \frac{\eta \mathrm{e}^{-(\lambda_1+\eta)t}(1 - \mathrm{e}^{-(\lambda_0-\lambda_1)t})}{\lambda_0 - \lambda_1},$$

and the conditional density of x is

$$f(x|y = 1) = \frac{f(x, y = 1)}{f(y = 1)} = \frac{\mathrm{e}^{-(\lambda_0-\lambda_1)x}(\lambda_0 - \lambda_1)}{1 - \mathrm{e}^{-(\lambda_0-\lambda_1)t}}. \tag{S12}$$

It is easily shown that this is a legitimate density function for x on $(0, t)$ since

$$\int_0^t \frac{\mathrm{e}^{-(\lambda_0-\lambda_1)x}(\lambda_0 - \lambda_1)}{1 - \mathrm{e}^{-(\lambda_0-\lambda_1)t}} \mathrm{d}x = 1,$$

and the limit $\lambda_0 \to \lambda_1$ exists such that

$$\lim_{\lambda_0 \to \lambda_1} f(x|y = 1) = \frac{1}{t}.$$

Thus, the density of speciation-rate shifts on the branch is uniform only in the case that $\lambda_0 = \lambda_1$, which is the trivial case that the prior on rates is degenerate so that no actual speciation-rate shifts occur at each event. Figure S3 illustrates the probability density of x on a branch of length 1 with $\lambda_0 < \lambda_1$, in which case the probability density of a rate shift increases with time. Conversely, if $\lambda_0 > \lambda_1$, it decreases with time. This shows that the uniform prior on diversification-rate shifts assumed by Rabosky [1] is not a reasonable model if diversification-rate shifts occur according to a stochastic process along lineages of the tree.

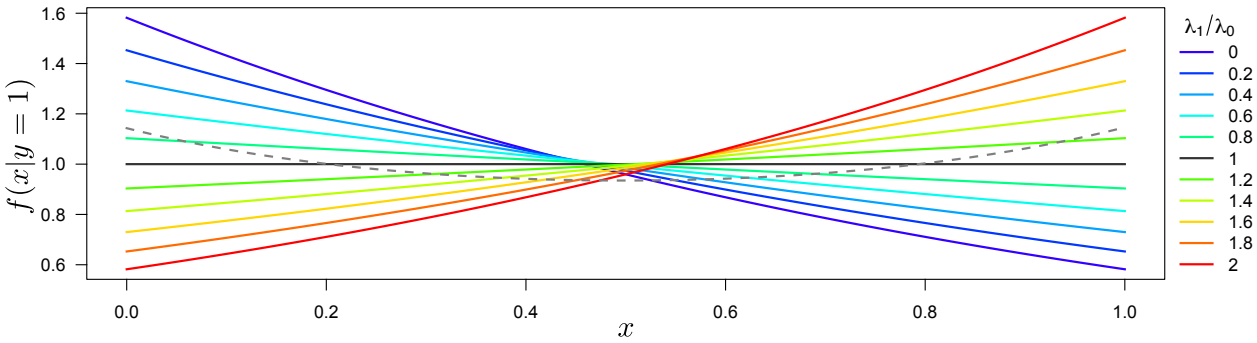

Figure S3: The probability density of a single event through time on a branch. We plot the conditional density function of an event as a function of time (described by equation S12) for $\lambda_0 = 1$ and various values of $\lambda_1 = \{0, 0.2, 0.4, \ldots, 2\}$. The probability density for the event time is only uniform in the special case where $\lambda_0 = \lambda_1$; otherwise, the probability density of an event increases in time when $\lambda_1 > \lambda_0$, whereas the probablity density of an event decreases in time when $\lambda_1 < \lambda_0$. When λ_0 and λ_1 are assumed to come from an exponential distribution (with rate 1), and the probability density function is (numerically) integrated over λ_0 and λ_1, the resulting density function remains non-uniform (dashed grey line).

S1.3.4 The distribution of the number of diversification-rate shifts is not Poisson

Consider a continuous-time pure-birth process that alternates between two instantaneous speciation rates, λ_0 and λ_1. Rate changes occur along a branch from either λ_1 to λ_0 or vice-versa with instantaneous rate η. The process terminates for a particular branch when a speciation event occurs. The process can be described using the instantaneous rate matrix

$$\begin{pmatrix} -(\eta+\lambda_0) & \eta & \lambda_0 \\ \eta & -(\eta+\lambda_1) & \lambda_1 \\ 0 & 0 & 0 \end{pmatrix},$$

where the indices of columns/rows correspond to the following states: (1) process has rate λ_0; (2) process has rate λ_1, and; (3) process has terminated due to a speciation event. The distribution of the number of events (speciation-rate changes) on a branch of length t under this process can be derived by integrating over the positions of the events. If the process is initiated with rate λ_0 the probabilities of 0, 1 or 2 rate-change events and no speciation event are

$$\begin{aligned} p_0 &= \mathrm{e}^{-(\lambda_0+\eta)t} \\ p_1 &= \int_0^t \eta \mathrm{e}^{-(\lambda_0+\eta)y} \mathrm{e}^{-(\lambda_1+\eta)(t-y)} dy \\ &= \mathrm{e}^{-(\lambda_1+\eta)t} \left(1 - \mathrm{e}^{-(\lambda_0-\lambda_1)t}\right) \frac{\eta}{\lambda_0-\lambda_1} \\ p_2 &= \int_0^t \int_0^{y_1} \eta^2 \mathrm{e}^{-(\lambda_0+\eta)y_1} \mathrm{e}^{-(\lambda_1+\eta)y_2} \mathrm{e}^{-(\lambda_0+\eta)(t-y1-y2)} dy_2 dy_1 \\ &= \mathrm{e}^{-(\lambda_0+\lambda_1+\eta)t} \left(\mathrm{e}^{\lambda_0 t} + \mathrm{e}^{\lambda_1 t}[\lambda_1 t - \lambda_0 t - 1]\right) \left(\frac{\eta}{\lambda_0-\lambda_1}\right)^2. \end{aligned}$$

The probabilities (conditioned on no speciation event) converge to those of the Poisson distribution for $\lambda_1 \to \lambda_0$ as expected

$$\begin{aligned} \left(\lim_{\lambda_1\to\lambda_0} p_0\right) \frac{1}{\mathrm{e}^{-\lambda_0}} &= \mathrm{e}^{-\eta t}, \\ \left(\lim_{\lambda_1\to\lambda_0} p_1\right) \frac{1}{\mathrm{e}^{-\lambda_0}} &= \mathrm{e}^{-\eta t}\eta t, \\ \left(\lim_{\lambda_1\to\lambda_0} p_2\right) \frac{1}{\mathrm{e}^{-\lambda_0}} &= \frac{\mathrm{e}^{-\eta t}(\eta t)^2}{2}. \end{aligned}$$

If a distribution is Poisson with expectation Λt, then in general

$$\frac{q_i}{q_{i-1}} = \left(\frac{\Lambda t}{i}\right),$$

and in particular

$$\frac{q_1}{q_0} = \Lambda t.$$

However, for the process outlined above

$$\frac{p_1}{p_0} = \frac{(1 - \mathrm{e}^{-(\lambda_1-\lambda_0)t})\eta}{\lambda_1-\lambda_0}.$$

In general, for any choice of function $\Lambda = f(\lambda_0, \lambda_1)$ that is independent of t, we have $p_1/p_0 \neq q_1/q_0$ except in the trivial case where $\lambda_1 = \lambda_0$, demonstrating that the distribution of the number of events on a branch of length t under the two-rate process is not Poisson (Figure S4). Note that it is not necessary to condition on the branch length in this proof because we are using ratios of probabilities conditioned on the same branch length, t. If the distribution of events is not Poisson for the two-rate pure-birth process studied here, it is not strictly Poisson for the birth-death process with a countably infinite distribution of rates, as the two-rate model is a special case of this more general process.

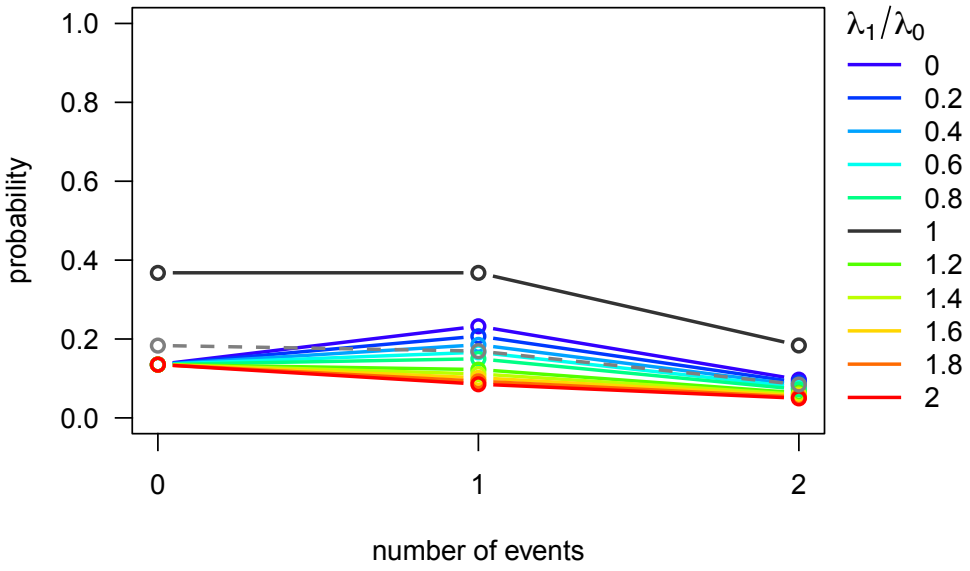

Figure S4: The distribution of the number events on a branch is not Poisson. We plot the probability of k events occuring on a branch of length 1 when $\lambda_0 = 1$ and for various values of $\lambda_1 = \{0, 0.2, \ldots, 2\}$. The probability depends on the ratio λ_1/λ_0. The distribution is only Poisson in the special (and trivial) case where $\lambda_0 = \lambda_1$. When λ_0 and λ_1 are assumed to come from an exponential distribution (with rate 1), and the probability mass function is (numerically) integrated over λ_0 and λ_1, the resulting probability mass function remains non-Poisson (dashed grey line).

S2 Monte Carlo simulation of extinction probabilities

We developed a Monte Carlo simulation approach for approximating the probability that a single lineage evolving under the compound-Poisson birth-death process initiated at time t goes extinct before reaching the present. For simplicity, we only implemented the time-independent version of the CPBDP (*i.e.,* $z = 0$).

We simulate under the time-independent compound-Poisson birth-death process as described Algorithm S1. Briefly, we initialize the process at time t with speciation and extinction rates λ_{init} and μ_{init}, respectively. The process continues forward in time, with the diversification process changing on each lineage at rate η. The process terminates when the number of species drops to 0, or it exceeds a pre-defined stopping point, $t_{\max}$ (the present). We record the termination time of each process, t_{end}, and repeat this simulation for N replicates. For the N replicate simulations, the Monte Carlo estimate of the extinction probability at time t, $\widehat{E(t)}$, is simply computed as the fraction of simulations that terminated before time $t_{\max}$:

$$\widehat{E(t)} = \frac{1}{N}\sum_{i=1}^{N} I(t^i_{\text{end}}),$$

where

$$I(t^i_{\text{end}}) = \begin{cases} 1 & \text{if } t^i_{\text{end}} < t_{\max} \\ 0 & \text{otherwise.} \end{cases}$$

The simulator is implemented in `R` and `C++` using `Rcpp` [11, 12], and is included elsewhere as an archived file (link: https://www.dropbox.com/sh/cua2p8i77kwsdmi/AAD_LdqYMZdBDFA4VOIQQpuea?dl=0).

S2.1 Extinction probability validation

We conducted a simulation study to characterize the behavior of the Monte Carlo simulator. We simulated extinction probabilities under a variety of parameter settings (Table S2). For each combination of parameter settings, we performed 50,000 Monte Carlo replicates and estimated the extinction probability, $\widehat{E(t)}$.

Table S2: Parameters used for Monte Carlo simulation study

Parameter	Values
t	0.0
$t_{\max}$	$\{0.0, 0.3, 0.6, \ldots, 3.0\}$
λ_{init}	1.0
μ_{init}	$\{0.1, 0.2, 0.3, \ldots, 0.9\}$
η	$\{0.0, 0.5, 1.0, 2.0, 4.0, 8.0\}$
μ_λ	1.0
μ_μ	0.5

We note that the constant-rate birth-death process (CRBDP) is a special case of the compound-Poisson birth-death process when $\eta = 0$. For the constant-rate birth-death process, the extinction probability is known analytically [13]:

$$E(t) = 1 - \frac{\lambda - \mu}{\lambda - \mu \exp\left[(\mu - \lambda)t_{\max}\right]}. \tag{S13}$$

To validate our Monte Carlo simulator, we compared the extinction probabilities estimated by Monte Carlo simulation to the analytical extinction probabilities under the constant-rate birth-death process. Our simulations show good agreement between the analytical constant-rate extinction probability (Figure S5, grey lines) and the Monte Carlo estimated extinction probability when $\eta = 0$ (Figure S5, dark blue dots).

As an additional validation, we analyzed the whale tree using `BAMM` under a constant-rate birth-death process model ($z = 0, \Lambda = 0$). We assumed complete species sampling (to avoid potentially confounding

Algorithm S1 Compound-Poisson birth-death process simulator. The process begins at time t with speciation and extinction rates λ_{init} and μ_{init}, respectively. New diversification parameters evolve at rate η. When new diverisification rate parameters evolve, new values for λ and μ are sampled from exponential distributions with expected value μ_λ and μ_μ, respectively. The process ends when the total number of species is 0, or when the time exceeds the pre-defined stop time, $t_{\max}$. The time of the end of the process is then returned.

Inputs:
  t: the starting time of the process.
  $t_{\max}$: the maximum time to run the process.
  λ_{init}: the speciation rate at the start of the process.
  μ_{init}: the extinction rate at the start of the process.
  η: the (per-lineage) rate at which new processes arise.
  μ_λ: the expected value of expontially distributed speciation rates.
  μ_μ: the expected value of expontially distributed extinction rates.
Initialize:
  $t_{\text{current}} \leftarrow t$
  $np \leftarrow 1$
  $\phi_1 \leftarrow \{n_1 = 1, \lambda_1 = \lambda_{\text{init}}, \mu_1 = \mu_{\text{init}}\}$
  $\boldsymbol{\phi} \leftarrow \{\phi_1\}$
while $\sum n_i > 0$ **do** // while the total number of species is greater than 0
  for $\phi_i \in \boldsymbol{\phi}$ **do**
    $\Delta t_i \sim \text{Exponential}(n_i(\eta + \lambda_i + \mu_i))$ // generate a waiting time for each extant process
  end for
  $x \leftarrow \text{argmin}_i(\Delta t_i)$ // get the index for the process with the minimum event time
  $t_{\text{current}} \leftarrow t_{\text{current}} + \Delta t_x$ // increment t_{current}
  if $t_{\text{current}} > t_{\max}$ **then** // if t is greater than $t_{\max}$, stop the process
    break
  else
    $u \sim \text{Uniform}(0, 1)$
    if $u \leq \eta/(\eta + \lambda_x + \mu_x)$ **then** // a new process arises
      $np \leftarrow np + 1$ // increment the number of processes
      $n_x \leftarrow n_x - 1$ // decrement the number of species in process x
      $\lambda_{np} \sim \text{Exponential}(1/\mu_\lambda)$ // get the new speciation rate
      $\mu_{np} \sim \text{Exponential}(1/\mu_\mu)$ // get the new extinction rate
      $\phi_{np} \leftarrow \{n_{np} = 1, \lambda_{np}, \mu_{np}\}$ // make the new process
      $\boldsymbol{\phi} \leftarrow \boldsymbol{\phi} \cup \phi_{np}$ // add the new process to $\boldsymbol{\phi}$
    else if $u \leq (\eta + \lambda_x)/(\eta + \lambda_x + \mu_x)$ **then** // the process has a speciation event
      $n_x \leftarrow n_x + 1$ // increment the number of species in process x
    else // the process has an extinction event
      $n_x \leftarrow n_x - 1$ // decrement the number of species in process x
    end if
    if $n_x = 0$ **then** // if the process has no species left, remove it from $\boldsymbol{\phi}$
      $\boldsymbol{\phi} \leftarrow \boldsymbol{\phi} \setminus \phi_x$
    end if
  end if
end while
return t_{current}

effects of approaches used to accommodate incomplete sampling). The MCMC was run for 100,000 generations, sampling every 100 generations. We modified `BAMM` to output the extinction probability at the root of the tree for each sample. We compared these extinction probabilities to those computed by Monte Carlo simulation, $\widehat{E(t)}$, and to the analytical constant-rate extinction probability using equation S13. All three methods yield identical extinction probabilities (within the tolerance of Monte Carlo error, see Figure S6).

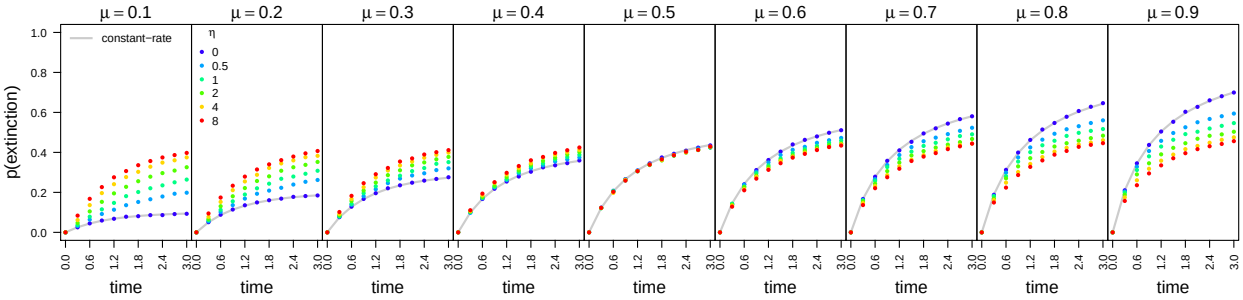

Figure S5: Monte Carlo estimates of extinction probabilities. We simulated extinction probabilities by Monte Carlo simulation under the compound-Poisson birth-death process (CPBDP). We simulated under a variety of transition rates, $\eta = \{0, 0.5, 1, 2, 4, 8\}$, a variety of initial extinction rates, $\mu = \{0.1, 0.2, \ldots, 0.9\}$. In all simulations, the speciation rate was $\lambda = 1$ and remained constant, $z = 0$, and the duration of the simulation was three time units. In each panel, we compare the analytical extinction probabilities under a constant-rate birth-death process (grey line) to the extinction probabilities estimated by Monte Carlo simulation under the CPBDP when $\eta = 0$ (which corresponds to a process without diversification-rate shifts, *i.e.,* a constant-rate birth-death process; dark blue dots).

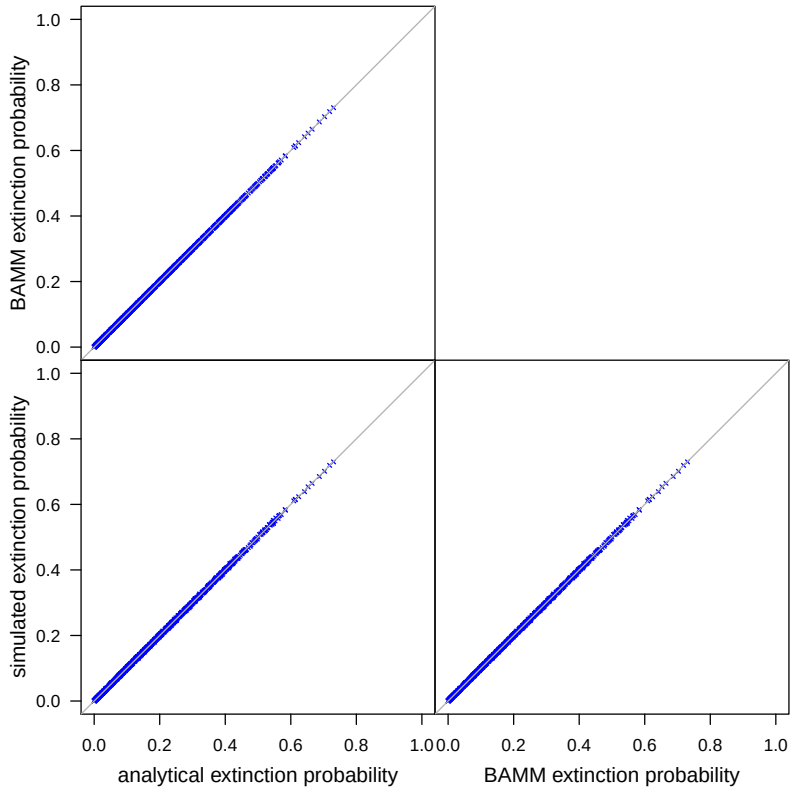

Figure S6: Validating extinction-probability estimates for the whale tree under a constant-rate birth-death process. We analyzed the whale tree using `BAMM`, assuming no diversification-rate shifts, $\eta = 0$ and time-homogeneous speciation rates, $z = 0$. For each sample from the posterior distribution, we estimated the extinction probability at the root of the tree using Monte Carlo simulation. Additionally, we computed the extinction probability analytically using equation S13. The extinction probability for each of these methods are in close agreement, validating our Monte Carlo estimator for the extinction probability.

S2.2 Likelihood calculation

We implemented a likelihood function that computes the probability of the data by substituting $\widehat{E(t)}$ for $E(t)$ in equation S8. While computationally intense, this approach offers an approximately correct means of computing the probability of the data under the model. The likelihood function is implemented in `R` and `C++` using `Rcpp` [11, 12], and is included as an archived file (see `supplementary_data.zip`).

We validated the likelihood function by analyzing the whale tree using `BAMM` under a time-homogeneous model ($z = 0$). We assumed complete species sampling (to avoid potentially confounding effects of approaches used to accommodate incomplete sampling). The MCMC was run for 100,000 generations, sampling every 100 generations. We then computed the likelihood of the whale tree using our likelihood function with the speciation- and extinction-rate parameters estimated by `BAMM`, and with $\eta = 0$ (*i.e.,* disallowing new diversification processes on extinct side-branches). With $\eta = 0$, our likelihood function should compute likelihoods nearly identical to those of `BAMM` (allowing minor differences due to numerical aspects of the algorithms). The likelihoods of the whale tree under `BAMM` and our likelihood function are identical, suggesting that our likelihood function is correctly implemented (Figure S7).

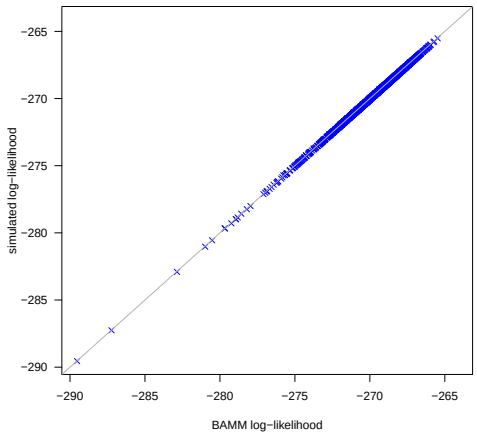

Figure S7: Log-likelihood of the whale tree under `BAMM` and our likelihood function, disallowing diversification-rate shifts on extinct lineages. We analyzed the whale tree using `BAMM`, assuming exactly two diversification-rate shifts and time-homogeneous speciation rates ($z = 0$). For each sample from the posterior distribution, we computed the likelihood of the data using our likelihood function for the sampled diversification-rate parameters (λ and μ) and assuming that $\eta = 0$. By assuming that $\eta = 0$, we effectively disallow diversification-rate shifts along extinct lineages (*i.e.,* as assumed in `BAMM`). Accordingly, in this case, the likelihoods computed by `BAMM` and our likelihood function should be nearly identical. Above, we plot the log-likelihood computed by `BAMM` (x-axis) against the log-likelihood computed using our likelihood function (y-axis). The close agreement between the log-likelihoods suggests our likelihood function is implemented correctly.

S3 Simulation study

S3.1 Simulation design

We explored aspects of the statistical behavior of BAMM by simulating trees both under a constant-rate birth-death process (*i.e.*, where the diversification rate *does not* vary across branches of the tree), and under the compound-Poisson birth-death process (*i.e.*, where the diversification rate *does* vary across branches of the tree). To ensure that our simulated trees are biologically realistic, we based our simulation on an empirical dataset—the whale tree presented in the original study [1].

S3.1.1 Constant-rate birth-death simulations

We estimated the posterior probability distribution of the speciation and extinction rates for the whale tree under a constant-rate birth-death model using RevBayes [14]. We then simulated 100 trees with 87 species (equal in size to the whale tree) under a constant-rate birth-death model using the R package TESS [15, 16]. Each tree was simulated under speciation and extinction rates that were independently sampled from the corresponding marginal posterior probability distributions that were previously estimated from the empirical tree.

S3.1.2 Variable-rate birth-death simulations

We simulated 100 variable-rate trees under a compound-Poisson birth-death process that conforms as closely as possible to the process assumed by BAMM. As in the constant-rate simulations, the variable-rate trees were based on the whale dataset presented in the original study [1]. We sampled speciation- and extinction-rate parameters from exponential distributions with means $\mu_\lambda = 0.15$ and $\mu_\mu = 0.05$, which are approximately centered on the means speciation- and extinction-rate estimates under the constant-rate birth-death model using RevBayes (as described in the previous section). Each tree was initialized with diversification-rate parameters sampled from these distributions; new diversification-rate parameters occurred with rate $\eta = \mu_\lambda/25 = 0.006$. We retained all simulated trees that: (1) experienced one or more diversification-rate shifts on an observed lineage, and; (2) had between 50 and 150 extant species. The mean number of diversification-rate shifts on observed lineages was 4.35 (the 95% quantile ranges of rate shifts on observed lineages was 1 to 10). The average range in net-diversification rate (*i.e.*, the difference between the branches with the lowest and highest net-diversification rate) for the rate-variable trees was 9.95.

We note that our variable-rate simulation differs substantially from that of Rabosky (2014). The rate-variable trees in that study were generated by first simulating a 'backbone' tree under a pure-birth (Yule) process with a fixed speciation rate. A branch was then arbitrarily selected, pruned, and replaced with a subclade that was simulated under a drastically different birth-death process. The simulated trees had a 43-fold average range in the net-diversification rate. Accordingly, our variable-rate trees explore a region of parameter space that is both far more relevant to the BAMM model (the process used to simulate trees corresponds more closely to the BAMM model) , and also to empirical datasets (the magnitude of net-diversification rate variation is unrealistic).

S3.2 Data analysis

Prior specification—For the constant-rate trees, we specified priors for the expected speciation rate, μ_λ, and extinction rate, μ_μ, that were centered on the true parameter values (the values used to simulate that tree). For the rate-variable trees, we used BAMMtools to select priors for μ_λ and μ_μ. We explored a range of values for the prior on the expected number of diversification-rate shifts, $\gamma = \{0.1, 0.5, 1, 2, 10\}$.

Posterior estimation—We inferred the joint posterior probability distribution for each simulated tree (under each of the five γ priors) using the Markov chain Monte Carlo (MCMC) algorithm implemented in the modified BAMM code, performing two independent, replicate MCMC simulations for 10^7 cycles, and thinned each chain by sampling every 1,000$^{\text{th}}$ state.

Prior estimation—We also inferred the joint prior probability distribution for each simulated dataset (under each of the five γ priors) using the two numerical options implemented in `BAMM` (described in Section S1.3.2): (1) we first inferred the prior using the `sampleFromPriorOnly` option by performing two independent, replicate MCMC simulations for 10^7 cycles, and thinned each chain by sampling every $1{,}000^{\text{th}}$ state. This approach uses the same MCMC algorithm used to estimate the joint posterior probability distribution, but forces the the likelihood function to return a log-likelihood value of 0. (2) We also inferred the prior using the `fastSimulatePrior` option, which provides a stand-alone algorithm for inferring the prior. Finally, we calculated the prior on the number of diversification-rate shifts using the analytical solution (Section S1.3.1).

MCMC diagnosis—We assessed the reliability of the MCMC simulations using `Tracer` [17] and `coda` [18]. Specifically, we assessed convergence of each MCMC simulation to the stationary (joint posterior or pior) distribution by plotting the time series for every parameter, and by calculating the effective sample size (ESS) [19] diagnostic for every parameter (ensuring ESS values $>>$ 500). We also assessed convergence by comparing the parameter estimates from the two replicate MCMC simulations for each dataset by calculating the potential scale reduction factor (PSRF) [20] diagnostic for all parameters (ensuring PSRF values ~ 1). Our inferences are based on the combined stationary samples from the two independent chains.

S3.3 Results summary

We summarized two aspects of the statistical behavior of `BAMM` on the simulated datasets. First, we explored the sensitivity of the inferred posterior number of diversification-rate shifts to the assumed γ prior. To this end, we combined samples from the joint posterior distributions of the 100 constant-rate birth-death trees inferred under each of the γ priors, and compared each of the estimated marginal posterior distributions on the number of diversification-rate shifts to the corresponding γ priors. Following Rabosky [1], we also plotted the MAP (=mode) of the marginal posterior distributions inferred under each value of the γ prior. These results are summarized in Figure 4.

Second, we assessed the accuracy of the estimated speciation and extinction rates (and also the derivative parameters, the net-diversification and relative-extinction rates) inferred using `BAMM`. To this end, for each simulated tree, we computed the posterior mean rates for each branch—marginalizing over all of the sampled diversification rate shifts—using the `BAMMtools` function `getMarginalBranchRateMatrix`. We compared the branch-specific estimates to their true rates using linear regression (as in Rabosky [1]). The linear regression model assumes that the estimated rate for a particular branch i, y_i, is predicted by:

$$y_i = \alpha + \beta y_{i,\text{true}} \tag{S14}$$

where $y_{i,\text{true}}$ is the true rate of that branch. Accordingly, values of α (the intercept) close to 0 and values of β (the slope) close to 1 indicate approximately unbiased estimates; conversely, values of β close to 0 indicate no relationship between true and estimated parameter, and negative values of β indicate a negative relationship between true and estimated parameters.

For the constant-rate simulations, we fit the linear regression model to all of the branch-specific estimates across the entire simulated dataset; these linear models therefore reflect the ability of `BAMM` to estimate parameters averaged over an entire simulated dataset when diversification rates are constant. Because individual constant-rate trees had a single set of diversification rate parameters, each estimated variable y_i had the same value of $y_{i,\text{true}}$, making it impossible to fit linear regression models to individual simulated trees. For the rate-variable simulations, we fit a linear regression model for the branch-specific rate estimates for *each individual* tree; these linear models thus reflect the accuracy of parameter estimates for individual `BAMM` analyses when rates vary (*i.e.,* when the likelihood is incorrect). We also computed the proportional error in branch-specific diversification-rate parameter estimates for the rate-variable trees as the posterior-mean estimate divided by the true parameter value. These results are summarized in Figures S8–S17.

S3.4 Impact of tree size

To explore whether the prior sensitivity apparent in Figure 4 was driven by small tree sizes (and a concomitantly low amount of data), we repeated the constant-rate simulation described above in Section S3.1.1, but increased the tree size to 500. We performed data analysis as described in Section S3.2. We then compared

the marginal posterior distribution of the number of rate shifts across all 100 trees. Results of this prior sensitivity analysis are summarized in Figure S18.

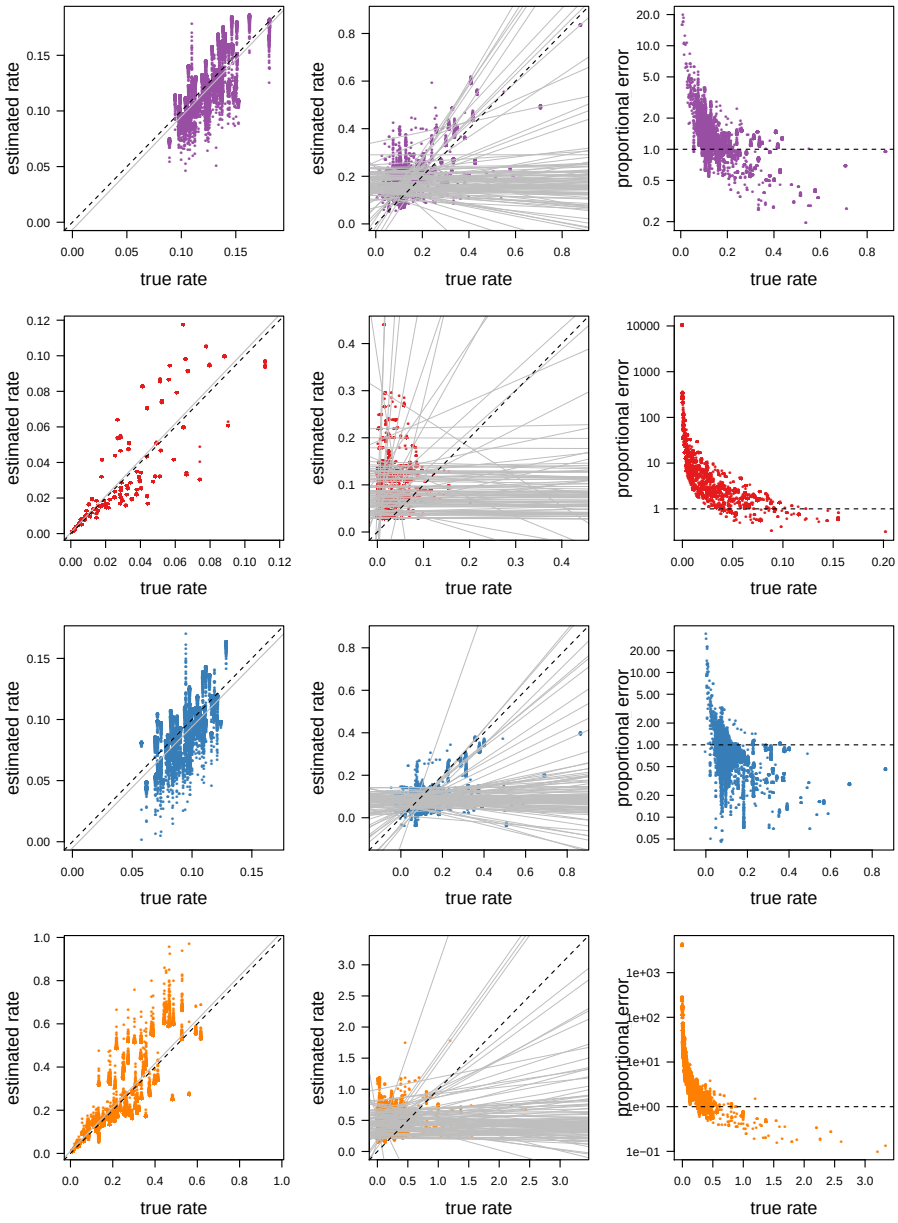

Figure S8: Branch-specific diversification-rate parameters estimated by BAMM are inaccurate when rates vary over the tree, $\gamma = 0.1$. We simulated 100 constant-rate trees (without diversification-rate shifts) and 100 rate-variable trees (with diversification-rate shifts) and analyzed each tree using BAMM. **Left column:** For the constant-rate trees, we plotted the posterior-mean estimates (the mean of the estimated marginal posterior distributions) of the speciation rate (purple, top row), extinction rate (red, second row), net-diversification (blue, third row), and relative-extn rate (orange, bottom row) for every branch of each tree (y-axis) against the corresponding true rates (x-axis). Following Rabosky [1], we used linear regression to estimate the relationship between the true rate and the estimated rates (solid grey line). For constant-rate trees, the branch-specific rate estimates are approximately unbiased. **Middle column:** For the rate-variable trees, we similarly plotted the posterior-mean estimates for the speciation, extinction, net-diversification, and relative-extinction rates for every branch of each tree (y-axis) against the corresponding true value (x-axis). For each simulated tree, we used linear regression to estimate the relationship between the true rates and the estimated rates (solid grey lines). **Right column:** For trees with diversification-rate shifts, we computed proportional error of the speciation-rate estimates (upper panel) and extinction-rate estimates (lower panel) for each branch in the tree, where the proportional error is simply the posterior-mean estimate of the rate divided by the true rate. (Note that the proportional error is plotted on a log scale.)

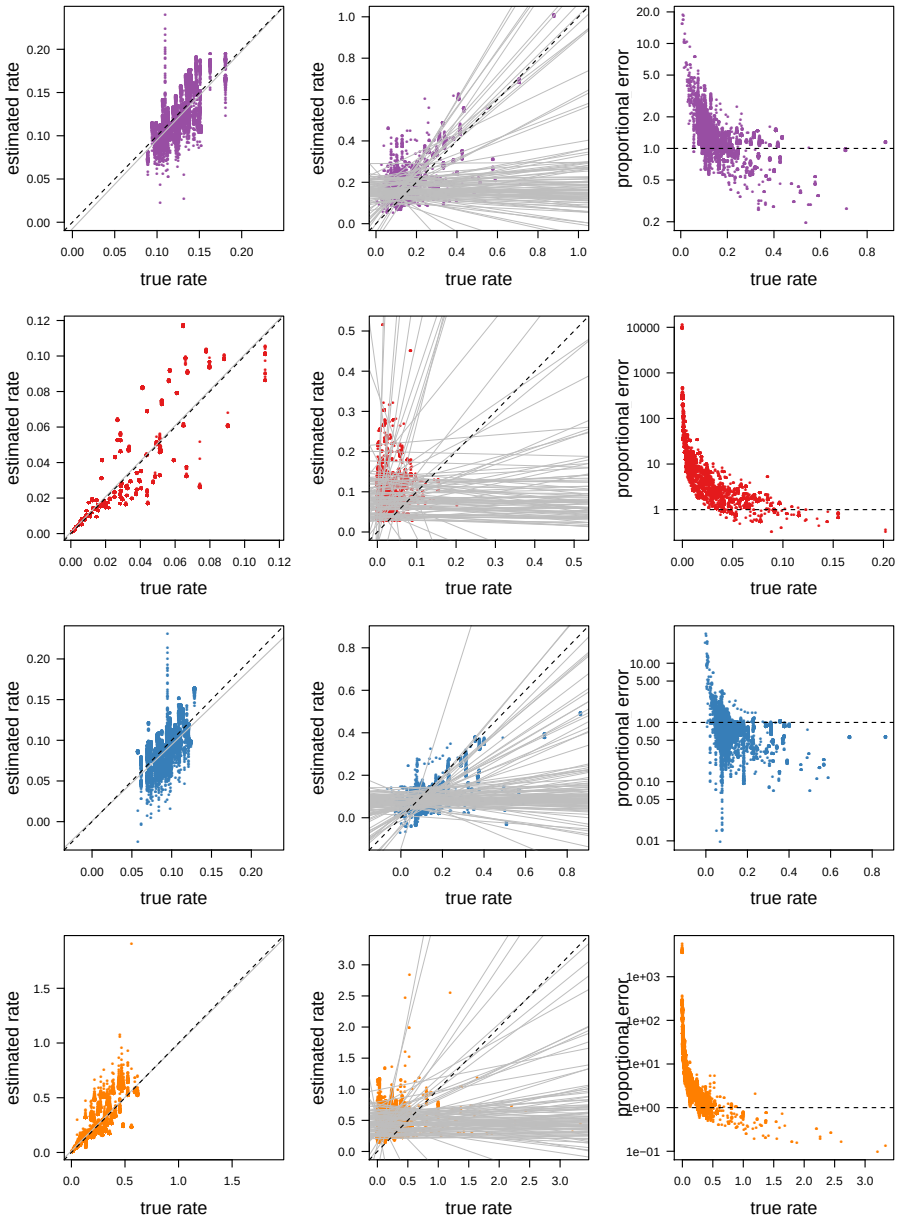

Figure S9: Branch-specific diversification-rate parameters estimated by BAMM are inaccurate when rates vary over the tree, $\gamma = 0.5$. We simulated 100 constant-rate trees (without diversification-rate shifts) and 100 rate-variable trees (with diversification-rate shifts) and analyzed each tree using BAMM. **Left column:** For the constant-rate trees, we plotted the posterior-mean estimates (the mean of the estimated marginal posterior distributions) of the speciation rate (purple, top row), extinction rate (red, second row), net-diversification (blue, third row), and relative-extn rate (orange, bottom row) for every branch of each tree (y-axis) against the corresponding true rates (x-axis). Following Rabosky [1], we used linear regression to estimate the relationship between the true rate and the estimated rates (solid grey line). For constant-rate trees, the branch-specific rate estimates are approximately unbiased. **Middle column:** For the rate-variable trees, we similarly plotted the posterior-mean estimates for the speciation, extinction, net-diversification, and relative-extinction rates for every branch of each tree (y-axis) against the corresponding true value (x-axis). For each simulated tree, we used linear regression to estimate the relationship between the true rates and the estimated rates (solid grey lines). **Right column:** For trees with diversification-rate shifts, we computed proportional error of the speciation-rate estimates (upper panel) and extinction-rate estimates (lower panel) for each branch in the tree, where the proportional error is simply the posterior-mean estimate of the rate divided by the true rate. (Note that the proportional error is plotted on a log scale.)

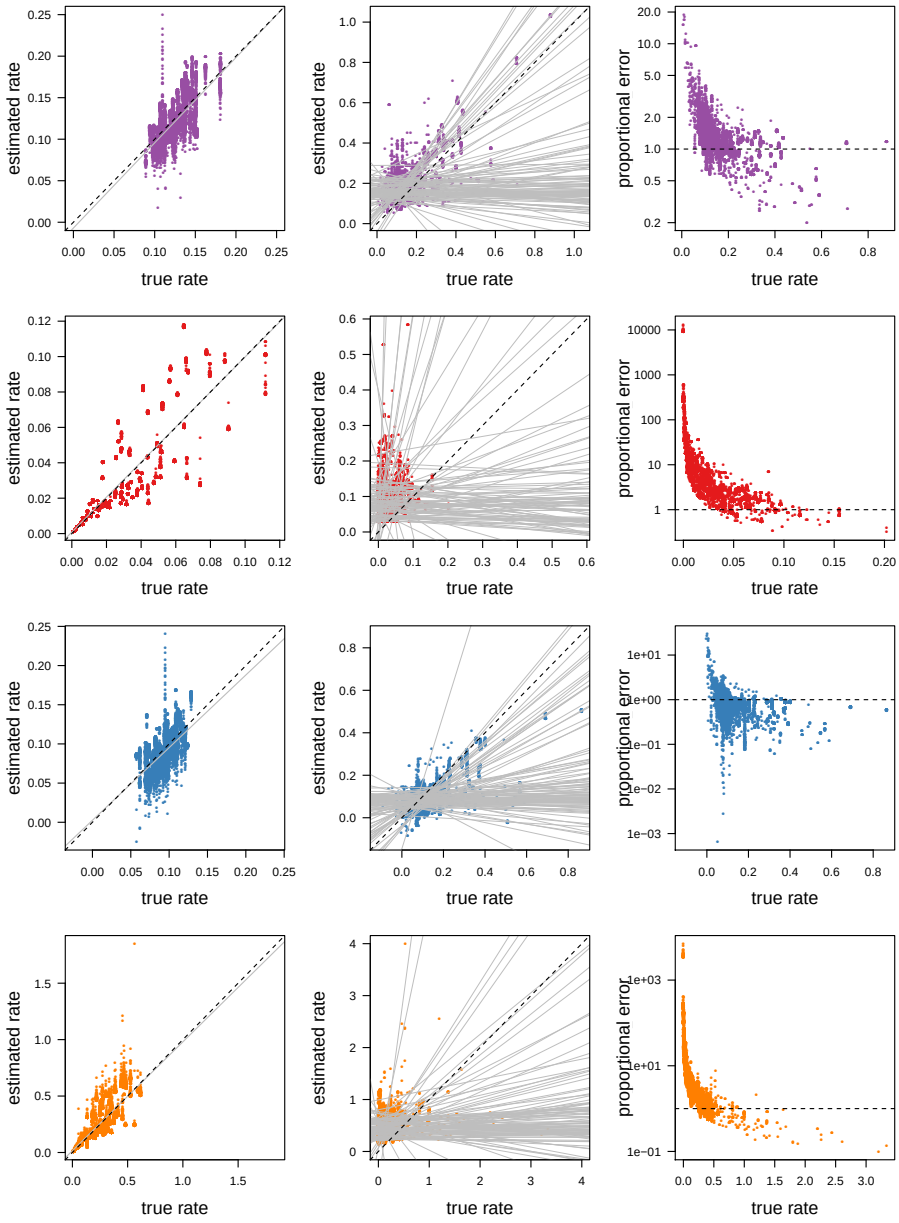

Figure S10: Branch-specific diversification-rate parameters estimated by BAMM are inaccurate when rates vary over the tree, $\gamma = 1.0$. We simulated 100 constant-rate trees (without diversification-rate shifts) and 100 rate-variable trees (with diversification-rate shifts) and analyzed each tree using BAMM. **Left column:** For the constant-rate trees, we plotted the posterior-mean estimates (the mean of the estimated marginal posterior distributions) of the speciation rate (purple, top row), extinction rate (red, second row), net-diversification (blue, third row), and relative-extn rate (orange, bottom row) for every branch of each tree (y-axis) against the corresponding true rates (x-axis). Following Rabosky [1], we used linear regression to estimate the relationship between the true rate and the estimated rates (solid grey line). For constant-rate trees, the branch-specific rate estimates are approximately unbiased. **Middle column:** For the rate-variable trees, we similarly plotted the posterior-mean estimates for the speciation, extinction, net-diversification, and relative-extinction rates for every branch of each tree (y-axis) against the corresponding true value (x-axis). For each simulated tree, we used linear regression to estimate the relationship between the true rates and the estimated rates (solid grey lines). **Right column:** For trees with diversification-rate shifts, we computed proportional error of the speciation-rate estimates (upper panel) and extinction-rate estimates (lower panel) for each branch in the tree, where the proportional error is simply the posterior-mean estimate of the rate divided by the true rate. (Note that the proportional error is plotted on a log scale.)

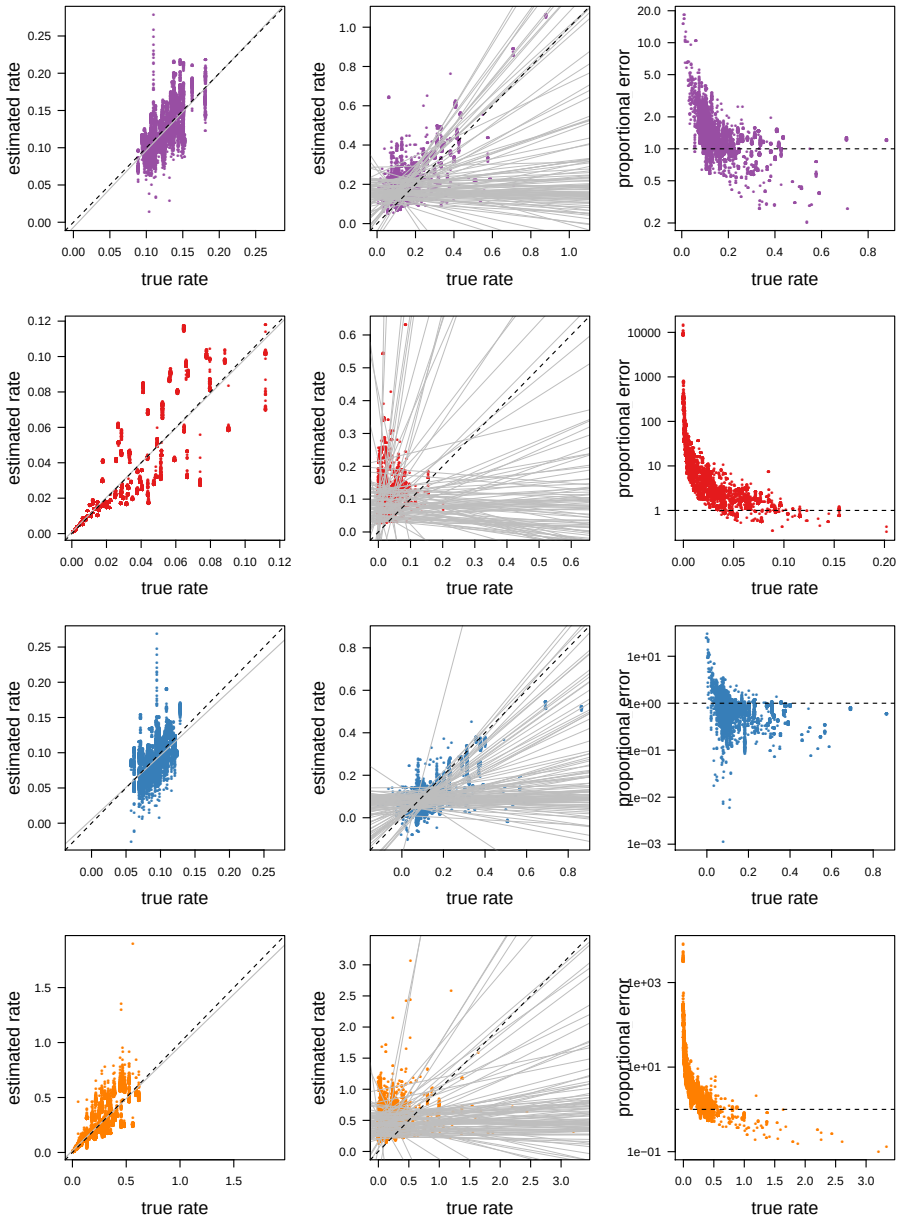

Figure S11: Branch-specific diversification-rate parameters estimated by BAMM are inaccurate when rates vary over the tree, $\gamma = 2.0$. We simulated 100 constant-rate trees (without diversification-rate shifts) and 100 rate-variable trees (with diversification-rate shifts) and analyzed each tree using BAMM. **Left column:** For the constant-rate trees, we plotted the posterior-mean estimates (the mean of the estimated marginal posterior distributions) of the speciation rate (purple, top row), extinction rate (red, second row), net-diversification (blue, third row), and relative-extn rate (orange, bottom row) for every branch of each tree (y-axis) against the corresponding true rates (x-axis). Following Rabosky [1], we used linear regression to estimate the relationship between the true rate and the estimated rates (solid grey line). For constant-rate trees, the branch-specific rate estimates are approximately unbiased. **Middle column:** For the rate-variable trees, we similarly plotted the posterior-mean estimates for the speciation, extinction, net-diversification, and relative-extinction rates for every branch of each tree (y-axis) against the corresponding true value (x-axis). For each simulated tree, we used linear regression to estimate the relationship between the true rates and the estimated rates (solid grey lines). **Right column:** For trees with diversification-rate shifts, we computed proportional error of the speciation-rate estimates (upper panel) and extinction-rate estimates (lower panel) for each branch in the tree, where the proportional error is simply the posterior-mean estimate of the rate divided by the true rate. (Note that the proportional error is plotted on a log scale.)

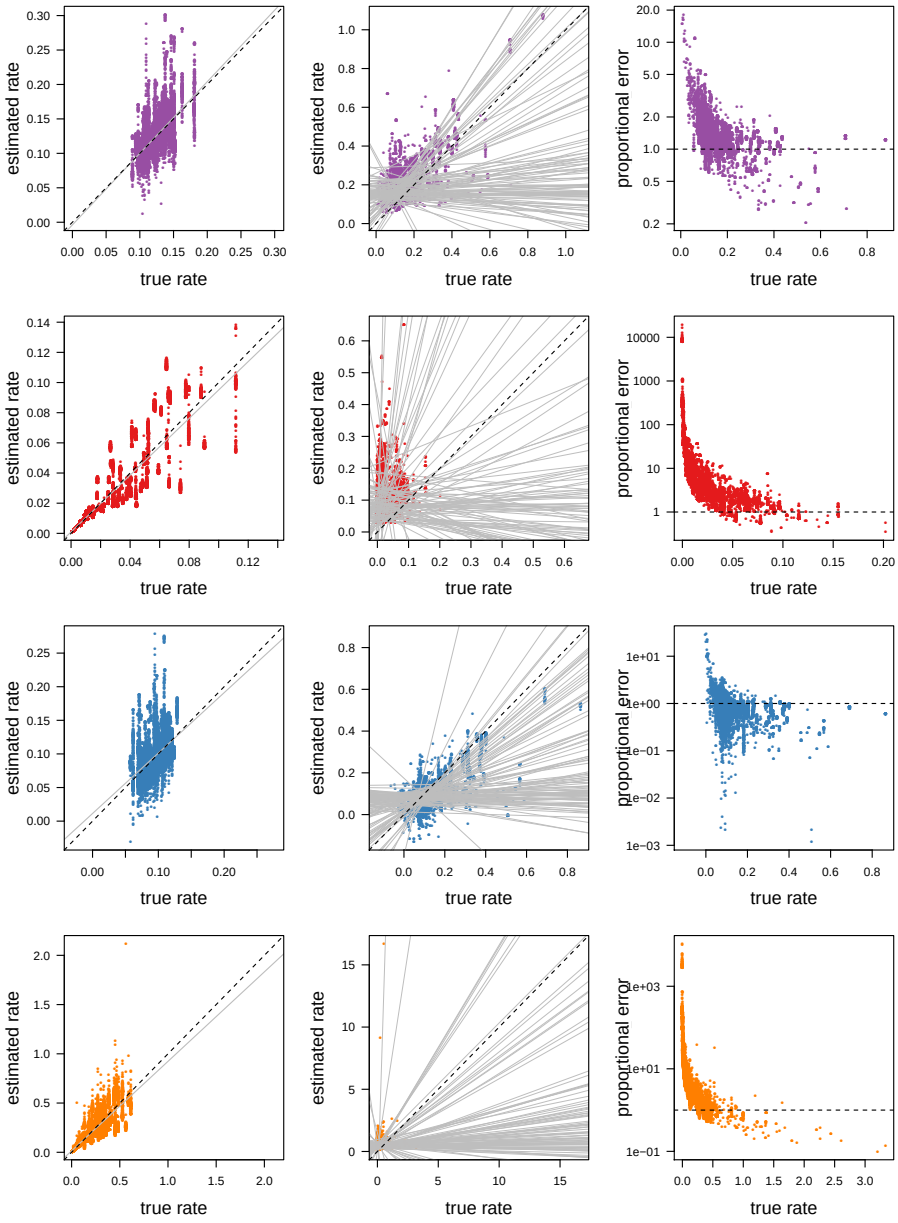

Figure S12: Branch-specific diversification-rate parameters estimated by BAMM are inaccurate when rates vary over the tree, $\gamma = 10.0$. We simulated 100 constant-rate trees (without diversification-rate shifts) and 100 rate-variable trees (with diversification-rate shifts) and analyzed each tree using BAMM. **Left column:** For the constant-rate trees, we plotted the posterior-mean estimates (the mean of the estimated marginal posterior distributions) of the speciation rate (purple, top row), extinction rate (red, second row), net-diversification (blue, third row), and relative-extn rate (orange, bottom row) for every branch of each tree (y-axis) against the corresponding true rates (x-axis). Following Rabosky [1], we used linear regression to estimate the relationship between the true rate and the estimated rates (solid grey line). For constant-rate trees, the branch-specific rate estimates are approximately unbiased. **Middle column:** For the rate-variable trees, we similarly plotted the posterior-mean estimates for the speciation, extinction, net-diversification, and relative-extinction rates for every branch of each tree (y-axis) against the corresponding true value (x-axis). For each simulated tree, we used linear regression to estimate the relationship between the true rates and the estimated rates (solid grey lines). **Right column:** For trees with diversification-rate shifts, we computed proportional error of the speciation-rate estimates (upper panel) and extinction-rate estimates (lower panel) for each branch in the tree, where the proportional error is simply the posterior-mean estimate of the rate divided by the true rate. (Note that the proportional error is plotted on a log scale.)

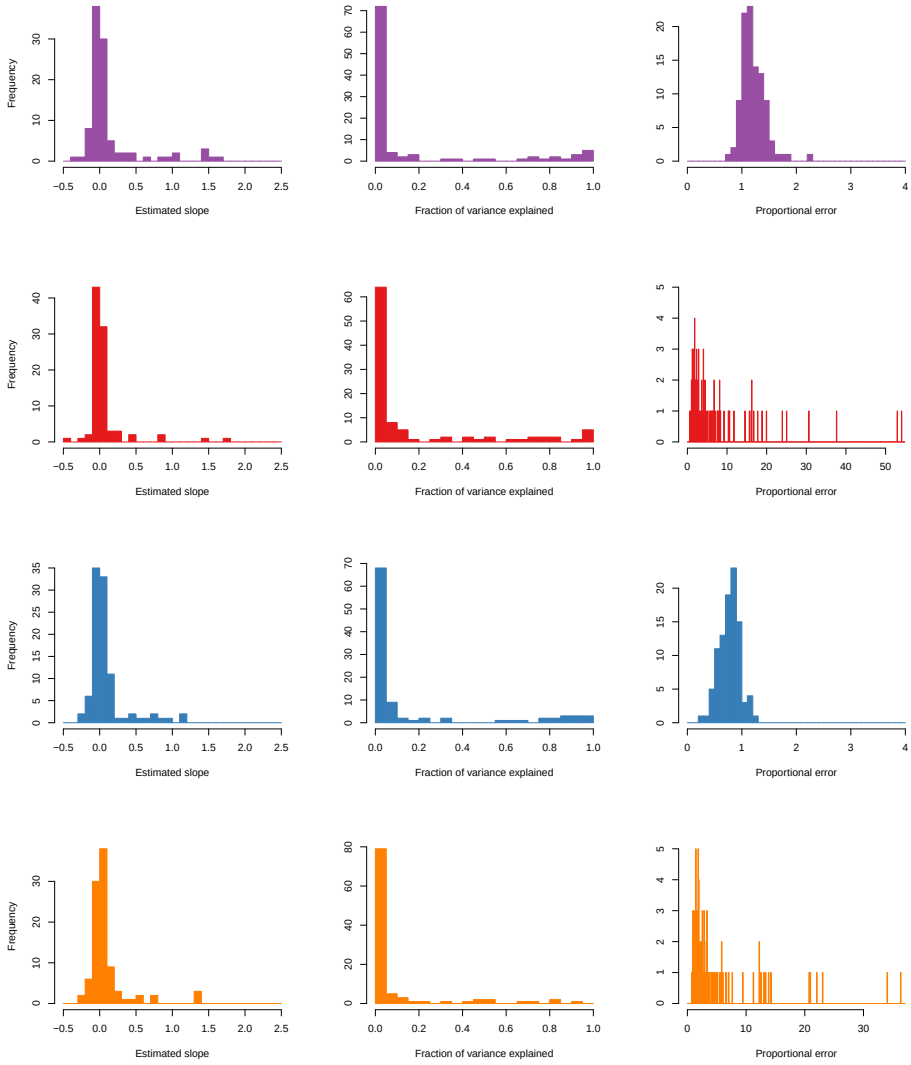

Figure S13: Linear model summaries of branch-specific diversification-rate parameters estimated by `BAMM`, $\gamma = 0.1$. We simulated 100 rate-variable trees (with diversification-rate shifts) and analyzed each tree using `BAMM` with priors selected using `BAMMtools`. For each simulated tree, we used linear regression to estimate the relationship between the true and estimated rates, and computed the proportional error of the rate estimates (*i.e.*, the posterior-mean estimate divided by the true rate) for each branch in the tree. For each diversification-rate parameter, speciation rate (purple, top row), extinction rate (red, second row), net-diversification (blue, third row), and relative extinction (orange, bottom row), we show summaries of the linear models. **Left column:** We plotted a frequency histogram of the slope parameter, β. **Center column:** We plotted a frequency histogram of the fraction of the variance explained by each linear model. **Right column:** We plotted the proportional error in the parameter estimates across branches in the simulation study.

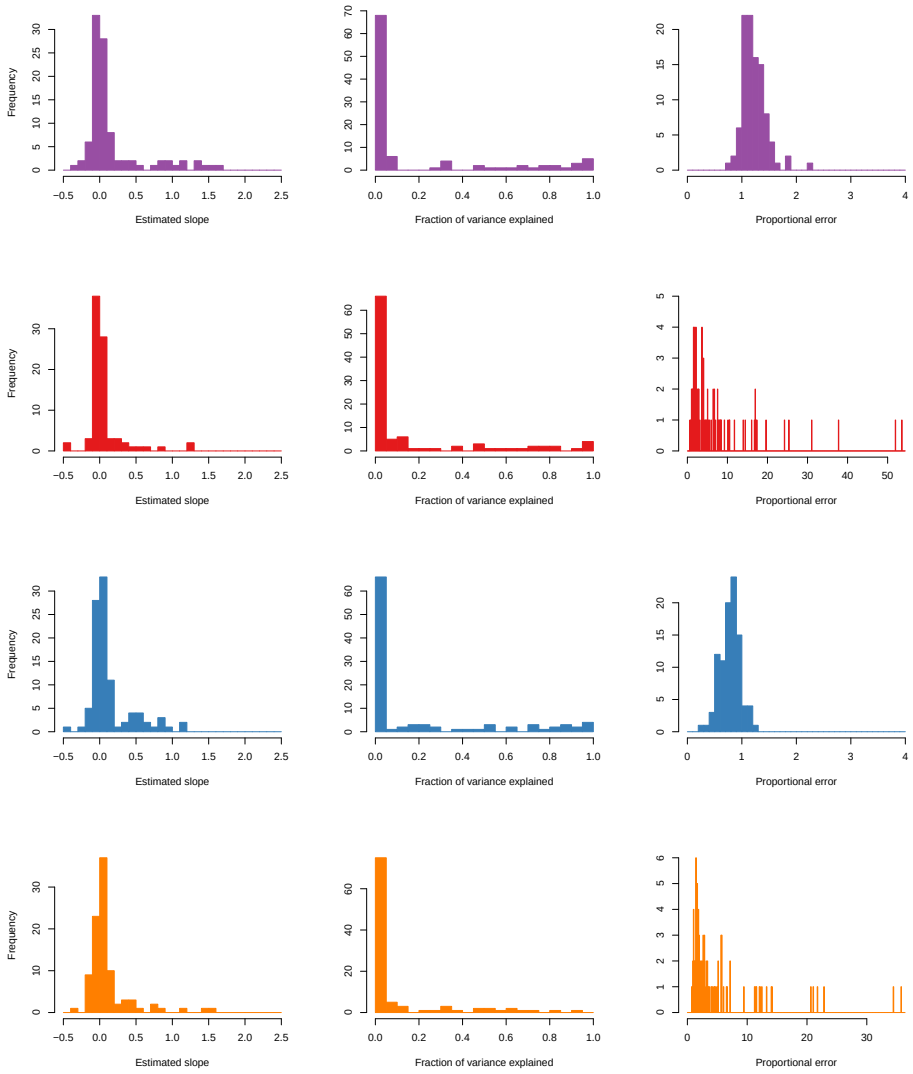

Figure S14: Linear model summaries of branch-specific diversification-rate parameters estimated by `BAMM`, $\gamma = 0.5$. We simulated 100 rate-variable trees (with diversification-rate shifts) and analyzed each tree using `BAMM` with priors selected using `BAMMtools`. For each simulated tree, we used linear regression to estimate the relationship between the true and estimated rates, and computed the proportional error of the rate estimates (*i.e.*, the posterior-mean estimate divided by the true rate) for each branch in the tree. For each diversification-rate parameter, speciation (top row, purple), extinction (second row, red), net-diversification (third row, blue), and relative extinction (bottom row, orange), we show summaries of the linear models. **Left column:** We plotted a frequency histogram of the slope parameter, β. **Center column:** We plotted a frequency histogram of the fraction of the variance explained by each linear model. **Right column:** We plotted the proportional error in the parameter estimates across branches in the simulation study.

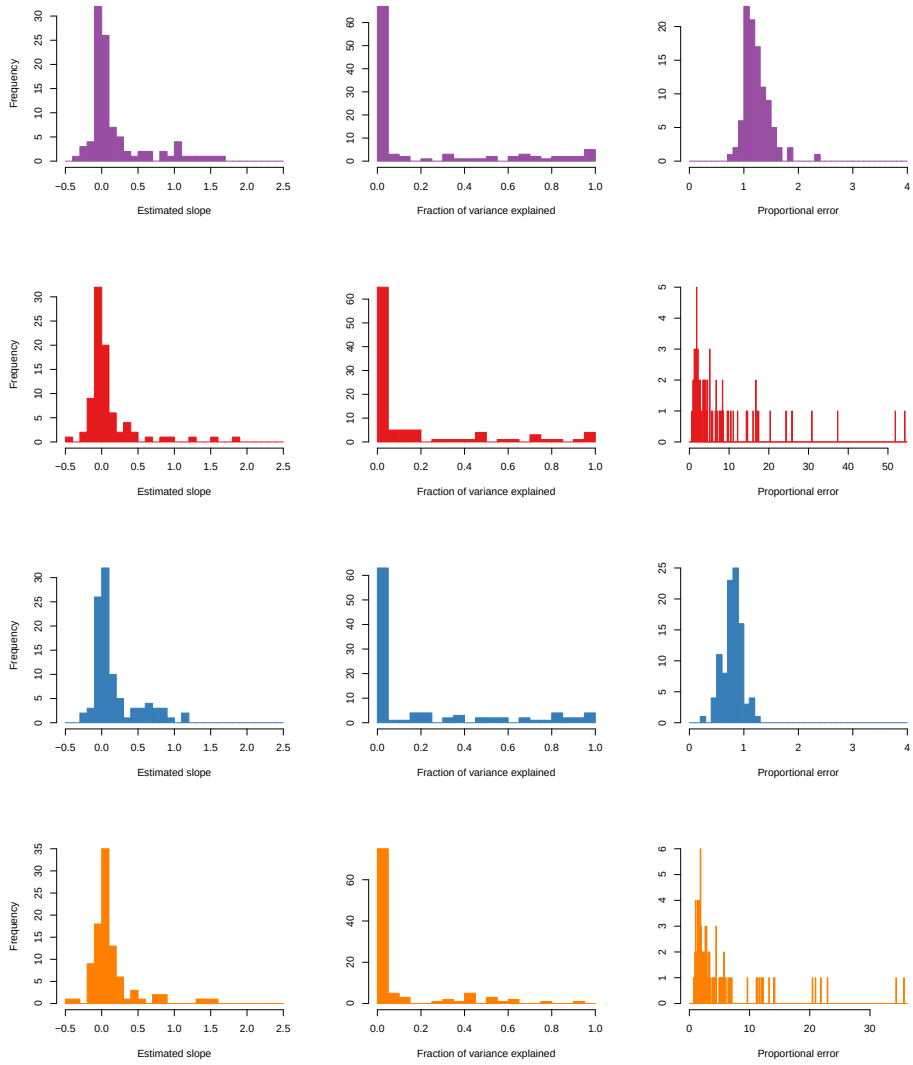

Figure S15: Linear model summaries of branch-specific diversification-rate parameters estimated by `BAMM`, $\gamma = 1.0$. We simulated 100 rate-variable trees (with diversification-rate shifts) and analyzed each tree using `BAMM` with priors selected using `BAMMtools`. For each simulated tree, we used linear regression to estimate the relationship between the true and estimated rates, and computed the proportional error of the rate estimates (*i.e.*, the posterior-mean estimate divided by the true rate) for each branch in the tree. For each diversification-rate parameter, speciation (top row, purple), extinction (second row, red), net-diversification (third row, blue), and relative extinction (bottom row, orange), we show summaries of the linear models. **Left column:** We plotted a frequency histogram of the slope parameter, β. **Center column:** We plotted a frequency histogram of the fraction of the variance explained by each linear model. **Right column:** We plotted the proportional error in the parameter estimates across branches in the simulation study.

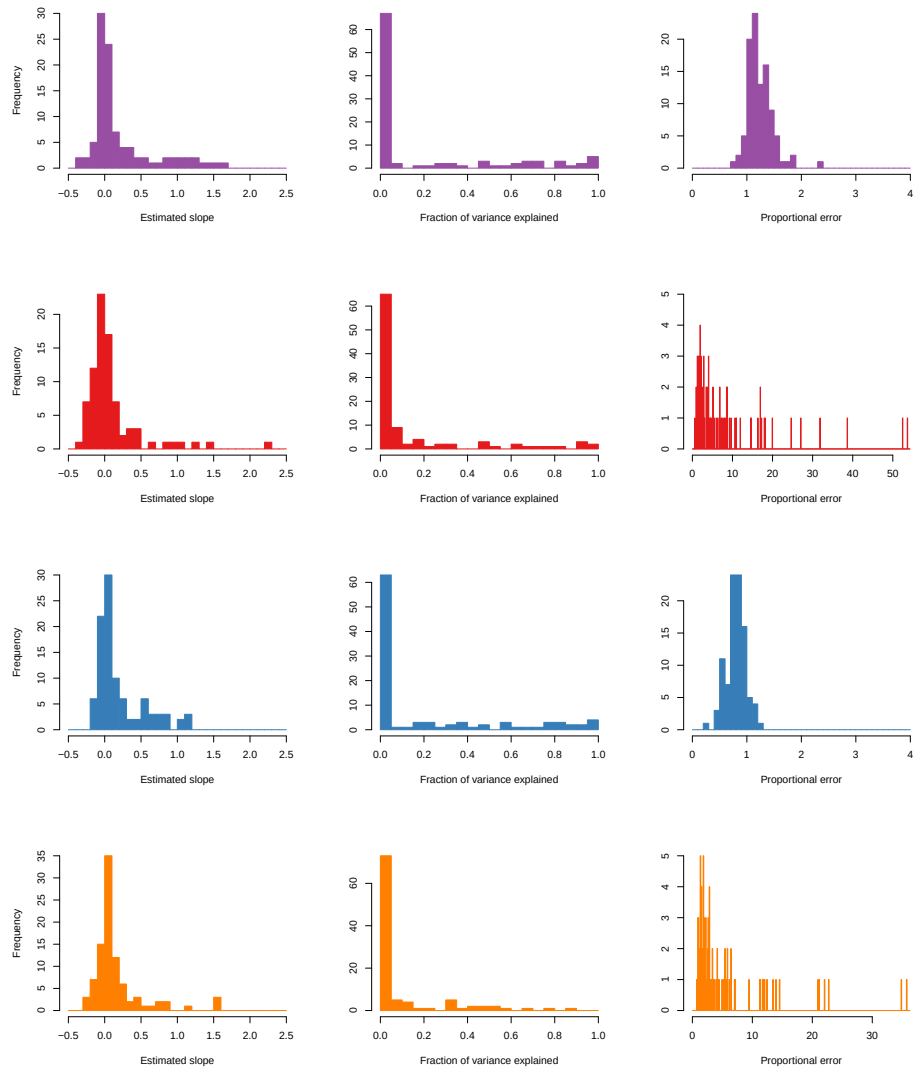

Figure S16: Linear model summaries of branch-specific diversification-rate parameters estimated by `BAMM`, $\gamma = 2.0$. We simulated 100 rate-variable trees (with diversification-rate shifts) and analyzed each tree using `BAMM` with priors selected using `BAMMtools`. For each simulated tree, we used linear regression to estimate the relationship between the true and estimated rates, and computed the proportional error of the rate estimates (*i.e.*, the posterior-mean estimate divided by the true rate) for each branch in the tree. For each diversification-rate parameter, speciation (top row, purple), extinction (second row, red), net-diversification (third row, blue), and relative extinction (bottom row, orange), we show summaries of the linear models. **Left column:** We plotted a frequency histogram of the slope parameter, β. **Center column:** We plotted a frequency histogram of the fraction of the variance explained by each linear model. **Right column:** We plotted the proportional error in the parameter estimates across branches in the simulation study.

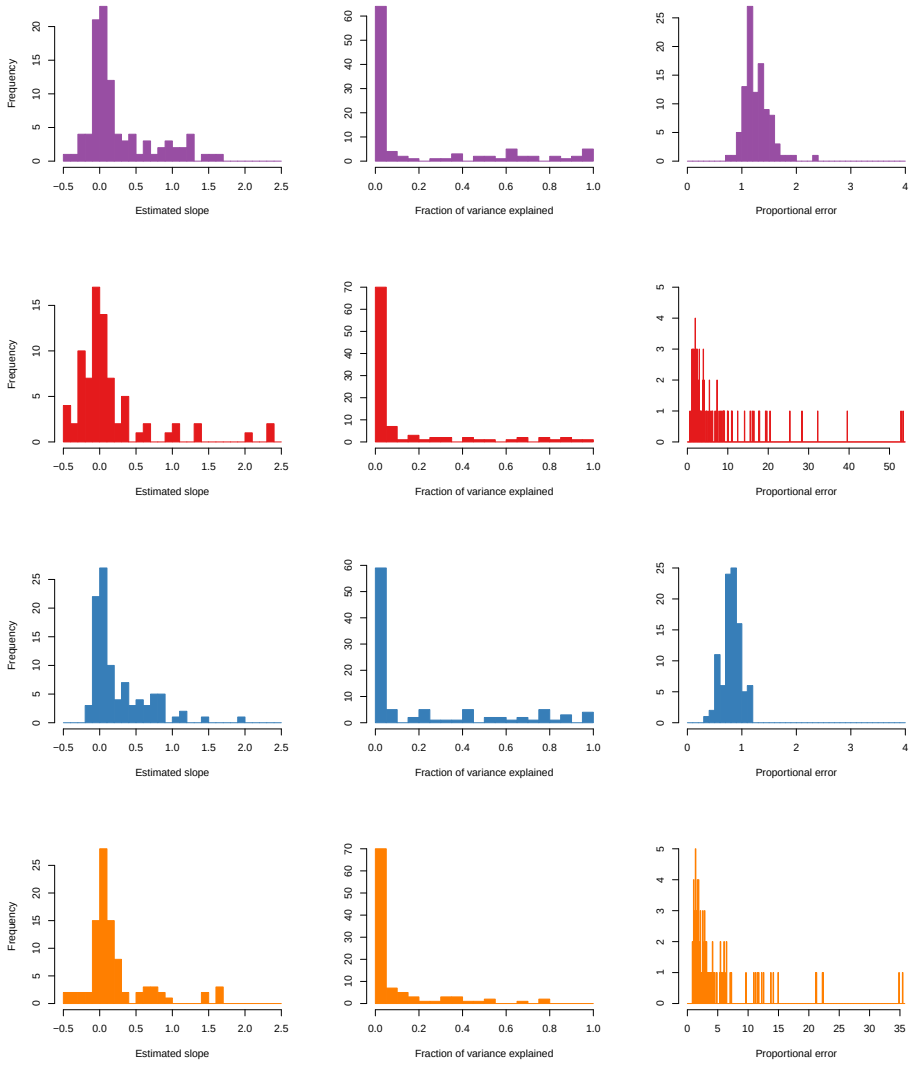

Figure S17: Linear model summaries of branch-specific diversification-rate parameters estimated by `BAMM`, $\gamma = 10.0$. We simulated 100 rate-variable trees (with diversification-rate shifts) and analyzed each tree using `BAMM` with priors selected using `BAMMtools`. For each simulated tree, we used linear regression to estimate the relationship between the true and estimated rates, and computed the proportional error of the rate estimates (*i.e.*, the posterior-mean estimate divided by the true rate) for each branch in the tree. For each diversification-rate parameter, speciation (top row, purple), extinction (second row, red), net-diversification (third row, blue), and relative extinction (bottom row, orange), we show summaries of the linear models. **Left column:** We plotted a frequency histogram of the slope parameter, β. **Center column:** We plotted a frequency histogram of the fraction of the variance explained by each linear model. **Right column:** We plotted the proportional error in the parameter estimates across branches in the simulation study.

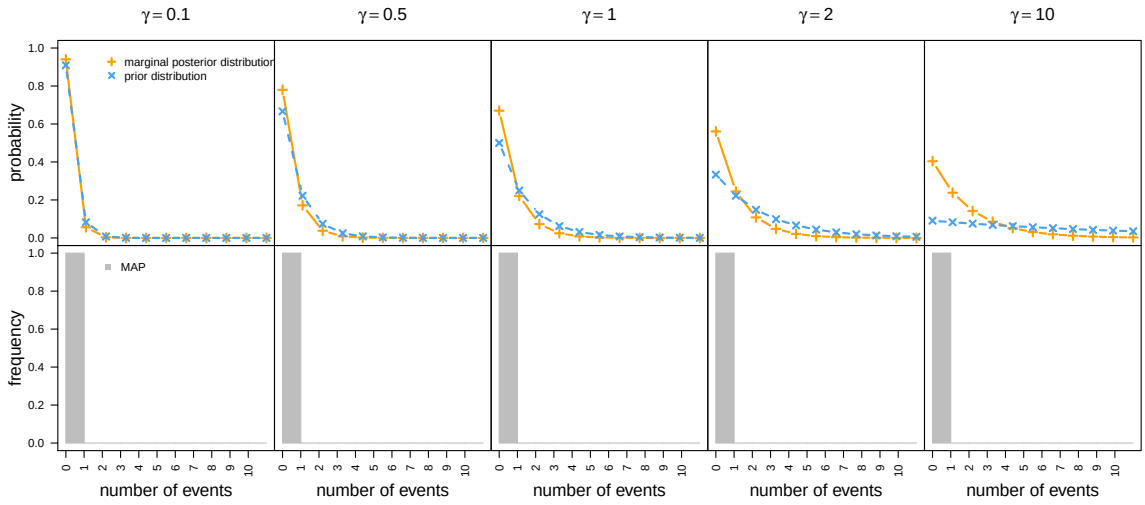

Figure S18: The posterior number of diversification-rate shifts is highly sensitive to the assumed prior. We simulated 100 constant-rate trees with 500 species and analyzed each using `BAMM` under a variety of priors for the expected number of diversification-rate shifts, γ (columns). **Top row:** For each value of the γ prior, we combined the MCMC samples from analyses of the 100 trees and plotted the marginal posterior probability density of the number of rate shifts estimated by `BAMM` (dashed orange lines) and the corresponding prior density (dashed blue lines). For all values of the γ prior, the estimated posterior is virtually indistinguishable from the assumed prior, and the *mode* of the prior densities is zero (*i.e.*, zero diversification-rate shifts). **Bottom row:** We then summarized the results of our simulation following Rabosky [1]. For each value of the γ prior, we recorded the mode of the posterior density (MAP) for each of the 100 trees, and then summarized these 100 MAP values as a histogram. For all values of the γ prior, the most frequent MAP has a value of zero (simply because the mode of the geometric prior is always zero, and the estimated posterior density closely mirrors the assumed prior).

S4 Empirical analyses

At the time of our study, `BAMM` had been used to infer the number and location of diversification-rate shifts in a total of 44 datasets within 35 published studies (S3). We were able to obtain the necessary information—the study phylogeny and control files—to reanalyze 15 of these datasets.

S4.1 Data analysis

Prior specification—For each empirical dataset, we specified priors for the speciation rate, λ, extinction rate, μ, and time-dependence parameter, z, using `BAMMtools` [55], and explored a range of values for the prior on the expected number of diversification-rate shifts, $\gamma = \{0.1, 0.5, 1, 2, 10\}$.

Posterior estimation—We inferred the joint posterior probability distribution for each empirical dataset (under each of the five γ priors) using the Markov chain Monte Carlo (MCMC) algorithm implemented in the modified `BAMM` code, performing four independent, replicate MCMC simulations for 10^7 cycles, and thinned each chain by sampling every $1{,}000^{\text{th}}$ state.

Prior estimation—We also inferred the joint prior probability distribution for each simulated dataset (under each of the five γ priors) using the two numerical options implemented in `BAMM` (described in Section S1.3.2): (1) we first inferred the prior using the `sampleFromPriorOnly` option by performing four independent, replicate MCMC simulations for 10^7 cycles, and thinned each chain by sampling every $1{,}000^{\text{th}}$ state. This approach uses the same MCMC algorithm used to estimate the joint posterior probability distribution, but forces the the likelihood function to return a log-likelihood value of 0. (2) We also inferred the prior using the `fastSimulatePrior` option, which provides a stand-alone algorithm for inferring the prior. Finally, we calculated the prior on the number of diversification-rate shifts using the analytical solution (Section S1.3.1).

MCMC diagnosis—We assessed the reliability of the MCMC simulations using the `Tracer` [17] and `coda` [18] packages. Specifically, we assessed convergence of each MCMC simulation to the stationary (joint posterior) distribution by plotting the time series for every parameter, and by calculating the effective sample size (ESS) [19] diagnostic for every parameter (ensuring ESS values $>>$ 500). We also assessed convergence by comparing the parameter estimates from the replicate MCMC simulations by calculating the potential scale reduction factor (PSRF) [20] diagnostic for all parameters (ensuring PSRF values ~ 1). Our inferences are based on the combined stationary samples from each of the four independent chains.

S4.2 Results summary

For each of the 15 empirical datasets, we present the marginal posterior distribution on the number of diversification-rate shifts inferred under each of the γ priors, comparing these posterior estimates to the corresponding estimated (simulated) and analytical priors.

Table S3: Empirical studies using BAMM to infer diversification-rate shifts

Study	Study Group
Armstrong et al. [21]	Sapotaceae
Bouchenak-Khelladi et al. [22]	Ericaceae
Bouchenak-Khelladi et al. [22]	Fagales
Bouchenak-Khelladi et al. [22]	Poales
Burns et al. [23]	Thraupidae
Condamine et al. [24]	Cycads
Cook et al. [25]	*Daviesia*
Cook et al. [25]	Bossiaea
Couvreur et al. [26]	Arecaceae
Ebel et al. [27]	*Adelpha*
Estep et al. [28]	Andropogoneae
Guo et al. [29]	*Paphiopedilum*
Hamm and Fordyce [30]	Nymphalidae
Hammer et al. [31]	*Ptilotus*
Hedges et al. [32]	Life
Huang and Rabosky [33]	*Aves*
Hughes and Atchison [34]	*Lupinus*
Koenen et al. [35]	Meliaceae
Kozak et al. [36]	*Heliconius*
Kraichak et al. [37]	Graphidaceae
Kraichak et al. [37]	Parmeliaceae
Linder and Bouchenak-Khelladi [38]	Danthonioideae
McGuire et al. [39]	Hummingbirds
Meudt et al. [40]	*Veronica*
Nürk et al. [41]	*Hypericum*
Peña et al. [42]	*Erebia*
Rabosky [1]	Cetaceans
Rabosky et al. [43]	Australian scincids
Rabosky et al. [44]	Aves
Salzman et al. [45]	*Costus*
Schilling et al. [46]	*Brickellia*
Schwery et al. [47]	Ericaceae
Spriggs et al. [48]	*Viburnum*
Verde Arregoitia et al. [49]	Lagomorpha
Villarreal et al. [50]	Hornworts
Weber and Agrawal [51]	*Byttneria*
Weber and Agrawal [51]	*Pleopeltis*
Weber and Agrawal [51]	*Polygoneae*
Weber and Agrawal [51]	*Senna*
Weber and Agrawal [51]	*Turnera*
Weber and Agrawal [51]	*Viburnum*
Weeks et al. [52]	Terebinthaceae
Willis et al. [53]	Malpighiaceae
Zelditch et al. [54]	Sciuridae

S4.2.1 *Adelpha*

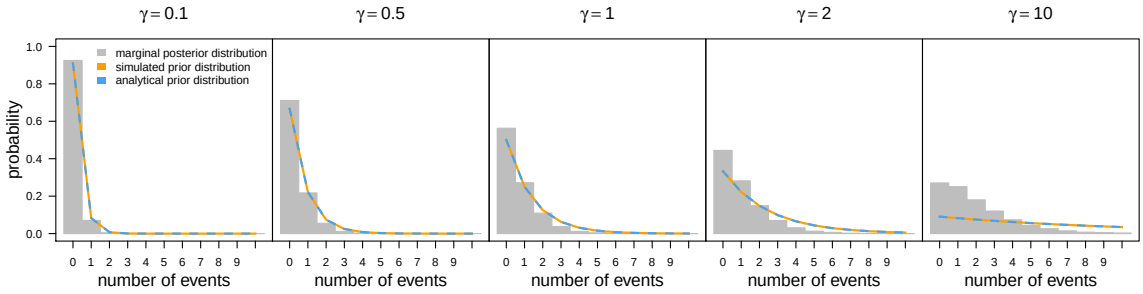

Figure S19: The marginal distribution on the number of shifts, k, as a function of the prior, γ, for *Adelpha* [27]. For each prior on the expected number of diversification-rate shifts, $\gamma = \{0.1, 0.5, 1, 2, 10\}$, we used the MCMC algoritm in `BAMM` to estimate the marginal posterior distribution on the number of diversification-rate shifts, k (grey bars). We compared these marginal posterior distributions to both the estimated prior distributions (inferred using `fastSimulatePrior` algorithm; orange curve), and also the analytical prior distributions (blue curve).

S4.2.2 *Byttneria*

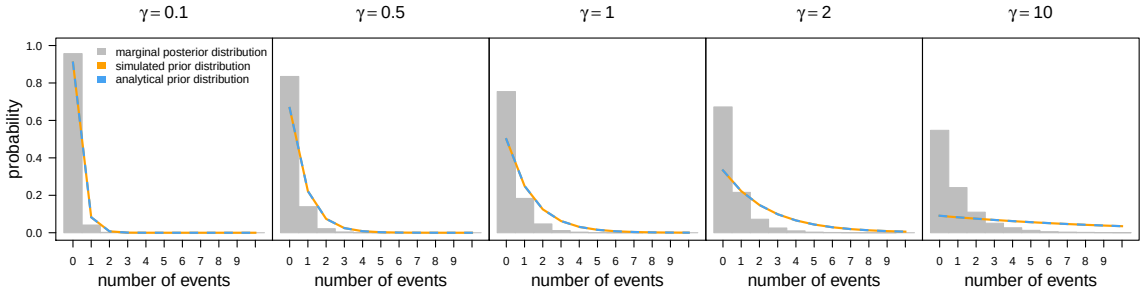

Figure S20: The marginal distribution on the number of shifts, k, as a function of the prior, γ, for *Byttneria* [51]. For each prior on the expected number of diversification-rate shifts, $\gamma = \{0.1, 0.5, 1, 2, 10\}$, we used the MCMC algoritm in `BAMM` to estimate the marginal posterior distribution on the number of diversification-rate shifts, k (grey bars). We compared these marginal posterior distributions to both the estimated prior distributions (inferred using `fastSimulatePrior` algorithm; orange curve), and also the analytical prior distributions (blue curve).

S4.2.3 Cetaceans

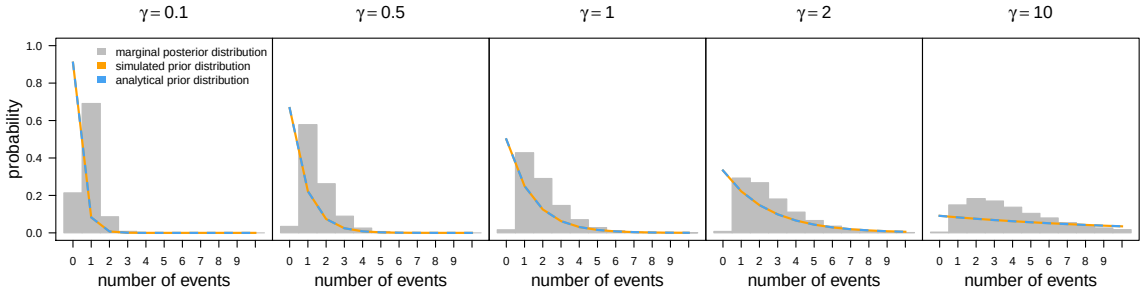

Figure S21: The marginal distribution on the number of shifts, k, as a function of the prior, γ, for whales [1]. For each prior on the expected number of diversification-rate shifts, $\gamma = \{0.1, 0.5, 1, 2, 10\}$, we used the MCMC algoritm in `BAMM` to estimate the marginal posterior distribution on the number of diversification-rate shifts, k (grey bars). We compared these marginal posterior distributions to both the estimated prior distributions (inferred using `fastSimulatePrior` algorithm; orange curve), and also the analytical prior distributions (blue curve).

S4.2.4 Ericaceae

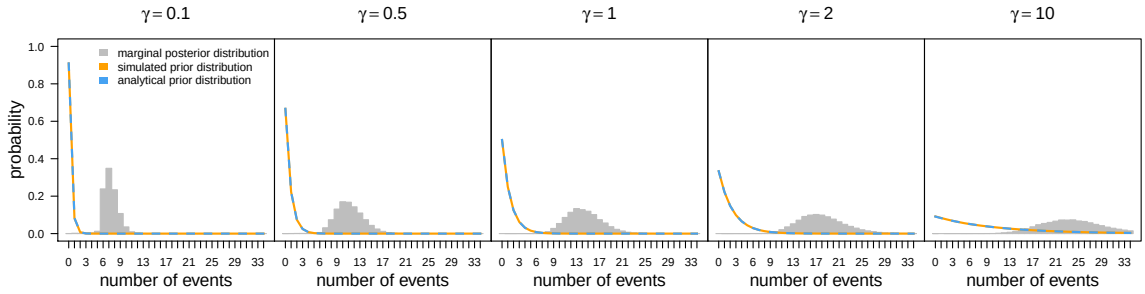

Figure S22: The marginal distribution on the number of shifts, k, as a function of the prior, γ, for Ericaceae [47]. For each prior on the expected number of diversification-rate shifts, $\gamma = \{0.1, 0.5, 1, 2, 10\}$, we used the MCMC algoritm in `BAMM` to estimate the marginal posterior distribution on the number of diversification-rate shifts, k (grey bars). We compared these marginal posterior distributions to both the estimated prior distributions (inferred using `fastSimulatePrior` algorithm; orange curve), and also the analytical prior distributions (blue curve).

S4.2.5 Graphidaceae

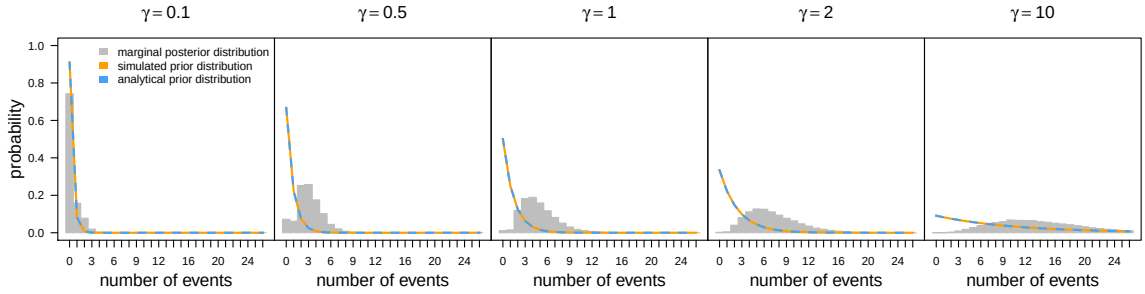

Figure S23: The marginal distribution on the number of shifts, k, as a function of the prior, γ, for Graphidaceae [37]. For each prior on the expected number of diversification-rate shifts, $\gamma = \{0.1, 0.5, 1, 2, 10\}$, we used the MCMC algoritm in `BAMM` to estimate the marginal posterior distribution on the number of diversification-rate shifts, k (grey bars). We compared these marginal posterior distributions to both the estimated prior distributions (inferred using `fastSimulatePrior` algorithm; orange curve), and also the analytical prior distributions (blue curve).

S4.2.6 Nymphalidae

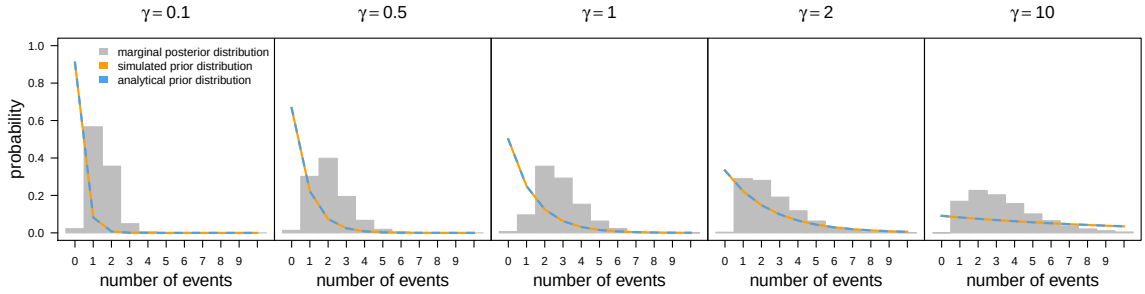

Figure S24: The marginal distribution on the number of shifts, k, as a function of the prior, γ, for Nymphalidae [30]. For each prior on the expected number of diversification-rate shifts, $\gamma = \{0.1, 0.5, 1, 2, 10\}$, we used the MCMC algoritm in `BAMM` to estimate the marginal posterior distribution on the number of diversification-rate shifts, k (grey bars). We compared these marginal posterior distributions to both the estimated prior distributions (inferred using `fastSimulatePrior` algorithm; orange curve), and also the analytical prior distributions (blue curve).

S4.2.7 *Paphiopedilum*

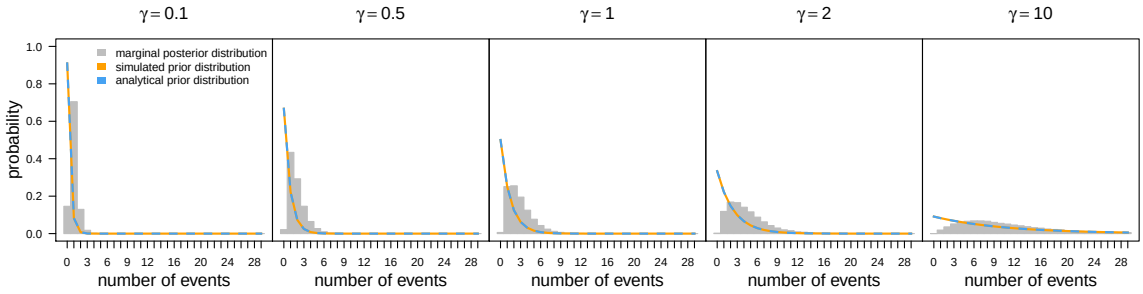

Figure S25: The marginal distribution on the number of shifts, k, as a function of the prior, γ, for *Paphiopedilum* [29]. For each prior on the expected number of diversification-rate shifts, $\gamma = \{0.1, 0.5, 1, 2, 10\}$, we used the MCMC algoritm in `BAMM` to estimate the marginal posterior distribution on the number of diversification-rate shifts, k (grey bars). We compared these marginal posterior distributions to both the estimated prior distributions (inferred using `fastSimulatePrior` algorithm; orange curve), and also the analytical prior distributions (blue curve).

S4.2.8 Parmeliaceae

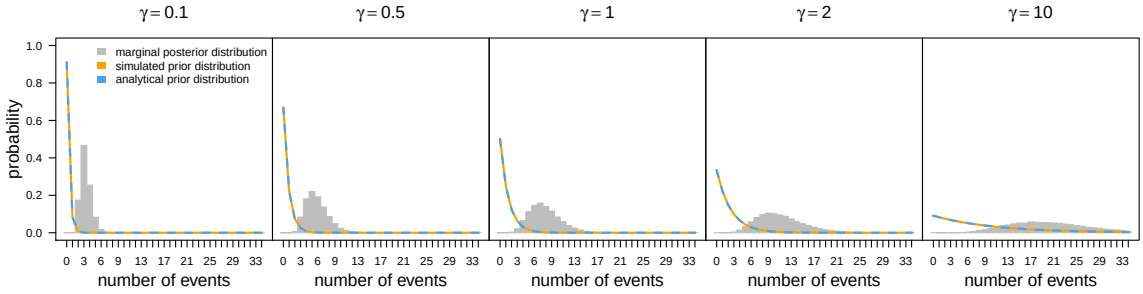

Figure S26: The marginal distribution on the number of shifts, k, as a function of the prior, γ, for Parmeliaceae [37]. For each prior on the expected number of diversification-rate shifts, $\gamma = \{0.1, 0.5, 1, 2, 10\}$, we used the MCMC algoritm in `BAMM` to estimate the marginal posterior distribution on the number of diversification-rate shifts, k (grey bars). We compared these marginal posterior distributions to both the estimated prior distributions (inferred using `fastSimulatePrior` algorithm; orange curve), and also the analytical prior distributions (blue curve).

S4.2.9 *Pleopeltis*

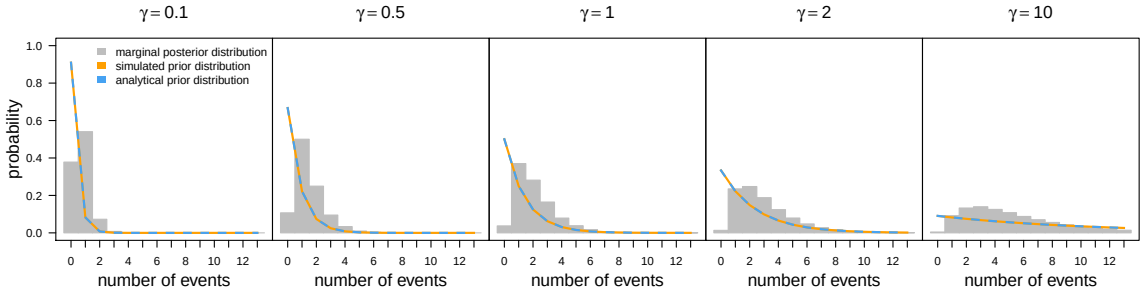

Figure S27: The marginal distribution on the number of shifts, k, as a function of the prior, γ, for *Pleopeltis* [51]. For each prior on the expected number of diversification-rate shifts, $\gamma = \{0.1, 0.5, 1, 2, 10\}$, we used the MCMC algoritm in `BAMM` to estimate the marginal posterior distribution on the number of diversification-rate shifts, k (grey bars). We compared these marginal posterior distributions to both the estimated prior distributions (inferred using `fastSimulatePrior` algorithm; orange curve), and also the analytical prior distributions (blue curve).

S4.2.10 Polygonea

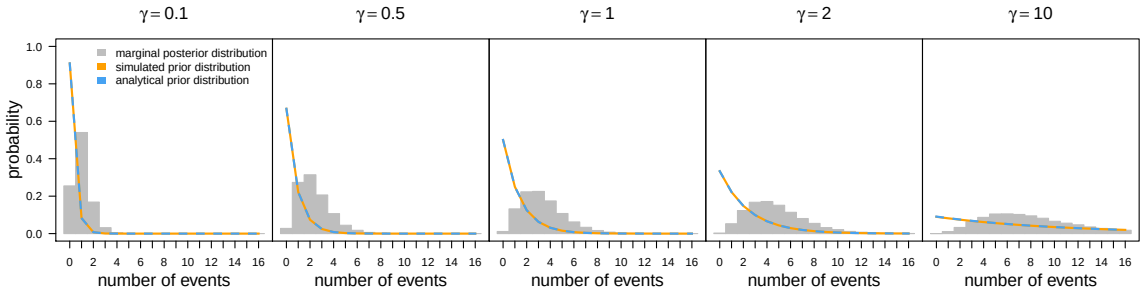

Figure S28: The marginal distribution on the number of shifts, k, as a function of the prior, γ, for Polygonea [51]. For each prior on the expected number of diversification-rate shifts, $\gamma = \{0.1, 0.5, 1, 2, 10\}$, we used the MCMC algoritm in `BAMM` to estimate the marginal posterior distribution on the number of diversification-rate shifts, k (grey bars). We compared these marginal posterior distributions to both the estimated prior distributions (inferred using `fastSimulatePrior` algorithm; orange curve), and also the analytical prior distributions (blue curve).

S4.2.11 *Senna*

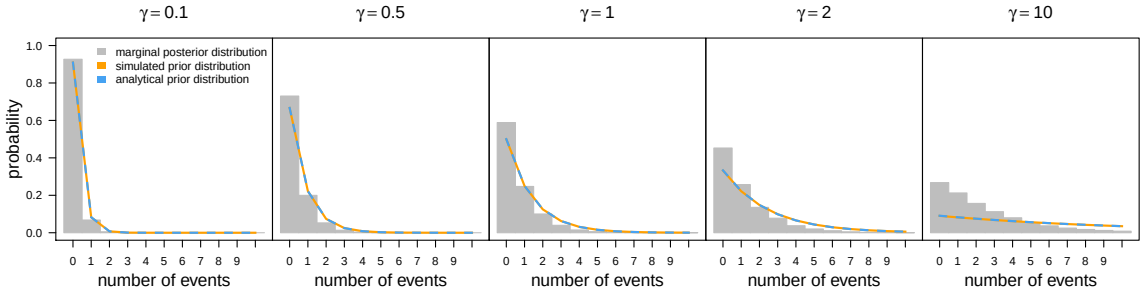

Figure S29: The marginal distribution on the number of shifts, k, as a function of the prior, γ, for *Senna* [51]. For each prior on the expected number of diversification-rate shifts, $\gamma = \{0.1, 0.5, 1, 2, 10\}$, we used the MCMC algoritm in `BAMM` to estimate the marginal posterior distribution on the number of diversification-rate shifts, k (grey bars). We compared these marginal posterior distributions to both the estimated prior distributions (inferred using `fastSimulatePrior` algorithm; orange curve), and also the analytical prior distributions (blue curve).

S4.2.12 Terebinthaceae

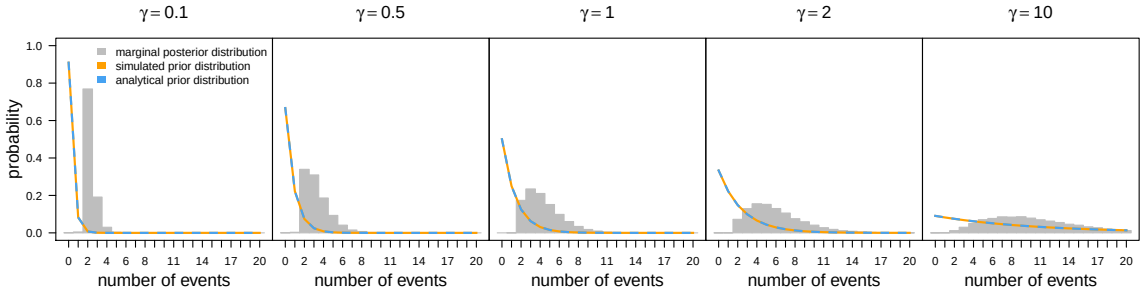

Figure S30: The marginal distribution on the number of shifts, k, as a function of the prior, γ, for Terebinthaceae [52]. For each prior on the expected number of diversification-rate shifts, $\gamma = \{0.1, 0.5, 1, 2, 10\}$, we used the MCMC algoritm in `BAMM` to estimate the marginal posterior distribution on the number of diversification-rate shifts, k (grey bars). We compared these marginal posterior distributions to both the estimated prior distributions (inferred using `fastSimulatePrior` algorithm; orange curve), and also the analytical prior distributions (blue curve).

S4.2.13 *Turnera*

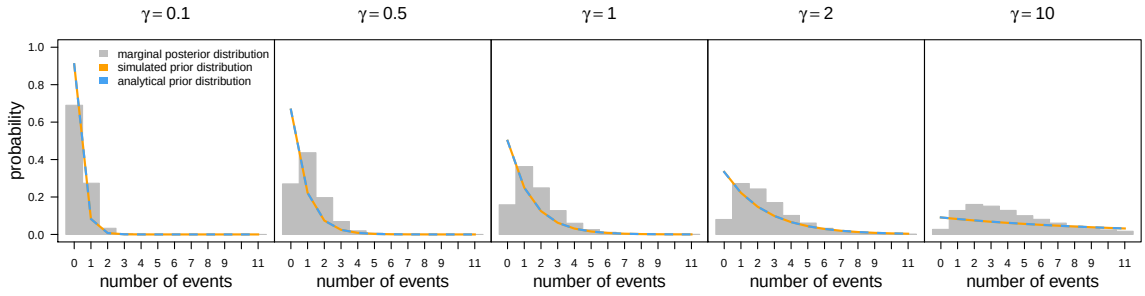

Figure S31: The marginal distribution on the number of shifts, k, as a function of the prior, γ, for *Turnera* [51]. For each prior on the expected number of diversification-rate shifts, $\gamma = \{0.1, 0.5, 1, 2, 10\}$, we used the MCMC algoritm in `BAMM` to estimate the marginal posterior distribution on the number of diversification-rate shifts, k (grey bars). We compared these marginal posterior distributions to both the estimated prior distributions (inferred using `fastSimulatePrior` algorithm; orange curve), and also the analytical prior distributions (blue curve).

S4.2.14 *Viburnum*

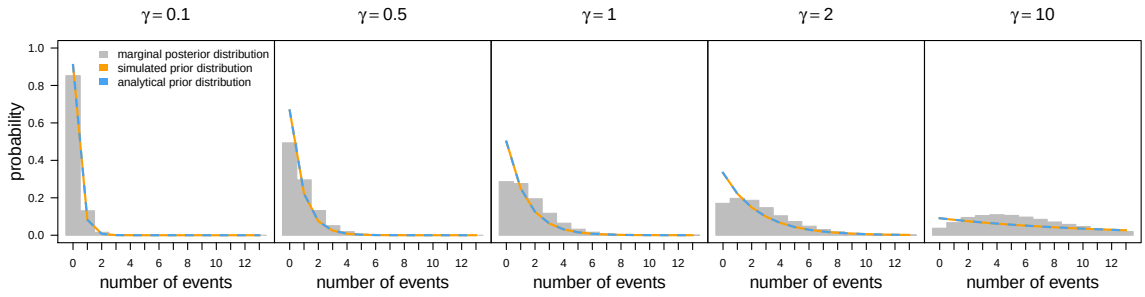

Figure S32: The marginal distribution on the number of shifts, k, as a function of the prior, γ, for *Viburnum* [51]. For each prior on the expected number of diversification-rate shifts, $\gamma = \{0.1, 0.5, 1, 2, 10\}$, we used the MCMC algoritm in `BAMM` to estimate the marginal posterior distribution on the number of diversification-rate shifts, k (grey bars). We compared these marginal posterior distributions to both the estimated prior distributions (inferred using `fastSimulatePrior` algorithm; orange curve), and also the analytical prior distributions (blue curve).

References

[1] Daniel L Rabosky. Automatic detection of key innovations, rate shifts, and diversity-dependence on phylogenetic trees. *PLoS One*, 9(2):e89543, 2014.

[2] Sean Nee, Robert M. May, and Paul H. Harvey. The reconstructed evolutionary process. *Philosophical Transactions: Biological Sciences*, 344(1309):305–311, 1994.

[3] Samantha R Cook, Andrew Gelman, and Donald B Rubin. Validation of software for Bayesian models using posterior quantiles. *Journal of Computational and Graphical Statistics*, 15(3), 2006.

[4] John P Huelsenbeck and Bruce Rannala. Frequentist properties of Bayesian posterior probabilities of phylogenetic trees under simple and complex substitution models. *Systematic Biology*, 53(6):904–913, 2004.

[5] W.P. Maddison, P.E. Midford, and S.P. Otto. Estimating a binary character's effect on speciation and extinction. *Systematic Biology*, 56(5):701, 2007.

[6] R.G. FitzJohn, W.P. Maddison, and S.P. Otto. Estimating trait-dependent speciation and extinction rates from incompletely resolved phylogenies. *Systematic Biology*, 58(6):595–611, 2009.

[7] Richard G FitzJohn. Diversitree: comparative phylogenetic analyses of diversification in r. *Methods in Ecology and Evolution*, 3(6):1084–1092, 2012.

[8] J. Felsenstein. *Inferring Phylogenies*. Sinauer Associates, Inc., Sunderland, Massachusetts, 2004.

[9] Major Greenwood and G Udny Yule. An inquiry into the nature of frequency distributions representative of multiple happenings with particular reference to the occurrence of multiple attacks of disease or of repeated accidents. *Journal of the Royal statistical society*, pages 255–279, 1920.

[10] G. U. Yule. A mathematical theory of evolution, based on the conclusions of Dr. J. C. Wills, F. R. S. *Philosophical Transactions of the Royal Society of London, Biology*, 213:21–87, 1924.

[11] Dirk Eddelbuettel and Romain François. Rcpp: Seamless R and C++ integration. *Journal of Statistical Software*, 40(8):1–18, 2011.

[12] Dirk Eddelbuettel. *Seamless R and C++ Integration with Rcpp*. Springer, New York, 2013.

[13] David G. Kendall. On the generalized "birth-and-death" process. *The Annals of Mathematical Statistics*, 19(1):1–15, 1948.

[14] Sebastian Höhna, Michael J. Landis, Bastien Boussau, Brian R. Moore, Nicolas Lartillot, Tracy A. Heath, John P. Huelsenbeck, and Fredrik Ronquist. RevBayes: A flexible framework for Bayesian inference of phylogeny. *Systematic Biology*, 65:(in press), 2016.

[15] Sebastian Höhna. Fast simulation of reconstructed phylogenies under global time-dependent birth-death processes. *Bioinformatics*, 29(11):1367–1374, 2013.

[16] Sebastian Höhna, Michael R May, and Brian R Moore. TESS: Bayesian inference of lineage diversification rates from (incompletely sampled) molecular phylogenies in R. *Bioinformatics*, 32:789–791, 2016.

[17] A J Drummond, M A Suchard, D Xie, and A Rambaut. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*, 29:1969–1973, 2012.

[18] Martyn Plummer, Nicky Best, Kate Cowles, and Karen Vines. CODA: Convergence diagnosis and output analysis for MCMC. *R News*, 6(1):7–11, 2006. URL `http://CRAN.R-project.org/doc/Rnews/`.

[19] Stephen P. Brooks and Andrew Gelman. General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, 7:434–455, 1997.

[20] Andrew Gelman and D.B. Rubin. Inference from iterative simulation using multiple sequences (with discussion). *Statistical Science*, 7:457–511, 1992.

[21] Kate E Armstrong, Graham N Stone, James A Nicholls, Eugenio Valderrama, Arne A Anderberg, Jenny Smedmark, Laurent Gautier, Yamama Naciri, Richard Milne, and James E Richardson. Patterns of diversification amongst tropical regions compared: a case study in sapotaceae. *Frontiers in genetics*, 5, 2014.

[22] Yanis Bouchenak-Khelladi, Renske E Onstein, Yaowu Xing, Orlando Schwery, and H Peter Linder. On the complexity of triggering evolutionary radiations. *New Phytologist*, 2015.

[23] Kevin J Burns, Allison J Shultz, Pascal O Title, Nicholas A Mason, F Keith Barker, John Klicka, Scott M Lanyon, and Irby J Lovette. Phylogenetics and diversification of tanagers (passeriformes: Thraupidae), the largest radiation of neotropical songbirds. *Molecular phylogenetics and evolution*, 75: 41–77, 2014.

[24] Fabien L Condamine, Nathalie S Nagalingum, Charles R Marshall, and Hélène Morlon. Origin and diversification of living cycads: a cautionary tale on the impact of the branching process prior in bayesian molecular dating. *BMC evolutionary biology*, 15(1):65, 2015.

[25] Lyn G Cook, Nate B Hardy, and Michael D Crisp. Three explanations for biodiversity hotspots: small range size, geographical overlap and time for species accumulation. an australian case study. *New Phytologist*, 2014.

[26] Thomas LP Couvreur, W Daniel Kissling, Fabien L Condamine, Jens-Christian Svenning, Nick P Rowe, and William J Baker. Global diversification of a tropical plant growth form: environmental correlates and historical contingencies in climbing palms. *Frontiers in genetics*, 5, 2014.

[27] Emily R Ebel, Jeffrey M DaCosta, Michael D Sorenson, Ryan I Hill, Adriana D Briscoe, Keith R Willmott, and Sean P Mullen. Rapid diversification associated with ecological specialization in Neotropical *Adelpha* butterflies. *Molecular ecology*, 24(10):2392–2405, 2015.

[28] Matt C Estep, Michael R McKain, Dilys Vela Diaz, Jinshun Zhong, John G Hodge, Trevor R Hodkinson, Daniel J Layton, Simon T Malcomber, Rémy Pasquet, and Elizabeth A Kellogg. Allopolyploidy, diversification, and the miocene grassland expansion. *Proceedings of the National Academy of Sciences*, 111(42):15149–15154, 2014.

[29] Yan-Yan Guo, Yi-Bo Luo, Zhong-Jian Liu, and Xiao-Quan Wang. Reticulate evolution and sea-level fluctuations together drove species diversification of slipper orchids (*Paphiopedilum*) in South-East Asia. *Molecular Ecology*, 2015.

[30] Christopher A Hamm and James A Fordyce. Patterns of host plant utilization and diversification in the brush-footed butterflies. *Evolution*, 69(3):589–601, 2015.

[31] Timothy Hammer, Robert Davis, and Kevin Thiele. A molecular framework phylogeny for ptilotus (amaranthaceae): Evidence for the rapid diversification of an arid australian genus. *Taxon*, 64(2): 272–285, 2015.

[32] S Blair Hedges, Julie Marin, Michael Suleski, Madeline Paymer, and Sudhir Kumar. Tree of life reveals clock-like speciation and diversification. *Molecular biology and evolution*, page msv037, 2015.

[33] Huateng Huang and Daniel L Rabosky. Sexual selection and diversification: reexamining the correlation between dichromatism and speciation rate in birds. *The American Naturalist*, 184(5):E101–E114, 2014.

[34] Colin E Hughes and Guy W Atchison. The ubiquity of alpine plant radiations: from the andes to the hengduan mountains. *New Phytologist*, 2015.

[35] Erik JM Koenen, James J Clarkson, Terence D Pennington, and Lars W Chatrou. Recently evolved diversity and convergent radiations of rainforest mahoganies (meliaceae) shed new light on the origins of rainforest hyperdiversity. *New Phytologist*, 207(2):327–339, 2015.

[36] Krzysztof M Kozak, Niklas Wahlberg, Andrew FE Neild, Kanchon K Dasmahapatra, James Mallet, and Chris D Jiggins. Multilocus species trees show the recent adaptive radiation of the mimetic heliconius butterflies. *Systematic biology*, page syv007, 2015.

[37] Ekaphan Kraichak, Robert Lücking, and H Thorsten Lumbsch. A unique trait associated with increased diversification in a hyper-diverse family of tropical lichen-forming fungi. *Int. J. Plant Sci*, 2015.

[38] H Peter Linder and Yanis Bouchenak-Khelladi. The causes of southern african spatial patterns in species richness: speciation, extinction and dispersal in the danthonioideae (poaceae). *Journal of Biogeography*, 42(5):914–924, 2015.

[39] JimmyA. McGuire, ChristopherC. Witt, J.V. Remsen Jr., Ammon Corl, DanielL. Rabosky, DouglasL. Altshuler, and Robert Dudley.

[40] Heidi M Meudt, Blanca M Rojas-Andrés, Jessica M Prebble, Evonne Low, Phil J Garnock-Jones, and Dirk C Albach. Is genome downsizing associated with diversification in polyploid lineages of veronica? *Botanical Journal of the Linnean Society*, 178(2):243–266, 2015.

[41] Nicolai M Nürk, Simon Uribe-Convers, Berit Gehrke, David C Tank, and Frank R Blattner. Oligocene niche shift, miocene diversification–cold tolerance and accelerated speciation rates in the st. johns worts (hypericum, hypericaceae). *BMC evolutionary biology*, 15(1):80, 2015.

[42] Carlos Peña, Heike Witthauer, Irena Klečková, Zdeněk Fric, and Niklas Wahlberg. Adaptive radiations in butterflies: evolutionary history of the genus erebia (nymphalidae: Satyrinae). *Biological Journal of the Linnean Society*, 2015.

[43] Daniel L Rabosky, Stephen C Donnellan, Michael Grundler, and Irby J Lovette. Analysis and visualization of complex macroevolutionary dynamics: an example from Australian scincid lizards. *Systematic Biology*, 63(4):610–627, 2014.

[44] Daniel L Rabosky, Huateng Huang, et al. Minimal effects of latitude on present-day speciation rates in new world birds. *Proceedings of the Royal Society of London B: Biological Sciences*, 282(1809):20142889, 2015.

[45] Shayla Salzman, Heather E Driscoll, Tanya Renner, Thiago André, Stacy Shen, and Chelsea D Specht. Spiraling into history: A molecular phylogeny and investigation of biogeographic origins and floral evolution for the genus costus. *Systematic Botany*, 40(1):104–115, 2015.

[46] Edward E Schilling, Jose L Panero, Bonnie S Crozier, Randall W Scott, and Patricia Dávila. Brickle-bush (brickellia) phylogeny reveals dimensions of the great asteraceae radiation in mexico. *Molecular phylogenetics and evolution*, 85:161–170, 2015.

[47] Orlando Schwery, Renske E Onstein, Yanis Bouchenak-Khelladi, Yaowu Xing, Richard J Carter, and Hans Peter Linder. As old as the mountains: the radiations of the Ericaceae. *New Phytologist*, 2014.

[48] Elizabeth L Spriggs, Wendy L Clement, Patrick W Sweeney, Santiago Madriñán, Erika J Edwards, and Michael J Donoghue. Temperate radiations and dying embers of a tropical past: the diversification of viburnum. *New Phytologist*, 2015.

[49] Luis D Verde Arregoitia, Katie Leach, Neil Reid, and Diana O Fisher. Diversity, extinction, and threat status in lagomorphs. *Ecography*, 2015.

[50] Juan Carlos Villarreal, Natalie Cusimano, and Susanne S Renner. Biogeography and diversification rates in hornworts: The limitations of diversification modeling. *Taxon*, 64(2):229–238, 2015.

[51] Marjorie G Weber and Anurag A Agrawal. Defense mutualisms enhance plant diversification. *Proceedings of the National Academy of Sciences*, 111(46):16442–16447, 2014.

[52] Andrea Weeks, Felipe Zapata, Susan K Pell, Douglas C Daly, John Mitchell, and Paul V A Fine. To move or to evolve: contrasting patterns of intercontinental connectivity and climatic niche evolution in "Terebinthaceae"(Anacardiaceae and Burseraceae). *Frontiers in Genetics*, 5:409, 2014.

[53] Charles G Willis, Brian F Franzone, Zhenxiang Xi, and Charles C Davis. The establishment of central american migratory corridors and the biogeographic origins of seasonally dry tropical forests in mexico. *Frontiers in genetics*, 5, 2014.

[54] Miriam L Zelditch, Jingchun Li, Lucy AP Tran, and Donald L Swiderski. Relationships of diversity, disparity, and their evolutionary rates in squirrels (sciuridae). *Evolution*, 69(5):1284–1300, 2015.

[55] Daniel L Rabosky, Michael Grundler, Carlos Anderson, Jeff J Shi, Joseph W Brown, Huateng Huang, Joanna G Larson, et al. BAMMtools: an R package for the analysis of evolutionary dynamics on phylogenetic trees. *Methods in Ecology and Evolution*, 5(7):701–707, 2014.