

# Macroevolutionary analysis of discrete traits with rate heterogeneity

Michael C. Grundler<sup>1,\*</sup> and Daniel L. Rabosky<sup>1</sup>

<sup>1</sup> Museum of Zoology and Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI 48109-1079, USA

*\* Corresponding author: mgru@umich.edu*

# ABSTRACT

Organismal traits show dramatic variation in phylogenetic patterns of origin and loss across the Tree of Life. Understanding the causes and consequences of this variation depends critically on accounting for heterogeneity in rates of trait evolution among lineages. Here, we describe a method for modeling among-lineage evolutionary rate heterogeneity in a trait with two discrete states. The method assumes that the present-day distribution of a binary trait is shaped by a mixture of stochastic processes in which the rate of evolution varies among lineages in a phylogeny. The number and location of rate changes, which we refer to as rate-shift events, are inferred automatically from the data. Simulations reveal that the method accurately reconstructs rates of trait evolution and ancestral character states even when simulated data violate model assumptions. We apply the method to an empirical dataset of mimetic coloration in snakes and find elevated rates of trait evolution in two clades of harmless snakes that are broadly sympatric with dangerously venomous New World coral snakes, recapitulating an earlier analysis of the same dataset. Although the method performed well on many simulated data sets, we caution that overall power for inferring heterogeneous dynamics of single binary traits is low.

Keywords: macroevolution, discrete trait evolution, rate heterogeneity

## INTRODUCTION

Organismal traits evolve recurrently across the Tree of Life, and the frequency with which traits evolve varies widely among clades. Among extant amniotes, for example, viviparity has evolved at least 100 times but only a single origin of it (in mammals) occurs outside squamate reptiles (Blackburn 1982, 1985). Similarly, sugar-secreting glands known as extrafloral nectaries have arisen hundreds of times among plants, mainly within the legumes, but not once among gymnosperms (Weber and Keeler 2013). Identifying the causes and consequences of repeated convergence depends critically on accounting for heterogeneity in rates of trait evolution. Ancestral trait reconstructions reveal evidence of convergence and can identify other traits enabling the repeated evolution of a convergent trait (Maddison 1990; de Queiroz and Rodríguez-Robles 2006; Marazzi et al. 2012; Christin et al. 2013). However, inferences of ancestral states may be seriously misled by models that fail to account for rate heterogeneity (King and Lee 2015). Several evolutionary theories also link how quickly traits evolve to how quickly lineages diversify (Vermeij 1973; Stanley 1979), and the repeated evolution of a trait together with methods that model among-lineage variation in evolutionary rates allows for direct tests of this coupling (Rabosky et al. 2013; Igea et al. 2017).

Methods to account for evolutionary rate heterogeneity have grown in number and in sophistication over recent years (O'Meara 2006; Revell and Collar 2009; Eastman et al. 2011; Lloyd et al. 2012; Marazzi et al. 2012; Beaulieu et al. 2013; Landis et al. 2013; Rabosky et al. 2014; Uyeda and Harmon 2014), but much of this methodological progress has focused on continuous traits such as body mass or seed size. Phylogenetics deals extensively with rate heterogeneity in discrete character evolution because of the challenge it poses for inference of

evolutionary relationships (e.g. Heath et al. 2012), and some of these methods carry over to the study of non-molecular traits. The “hidden rates” (Beaulieu et al. 2013) and “precursor” (Marazzi et al. 2012) models, first introduced to study variation in growth form and extrafloral nectary production in plants, are closely related to covarion models of nucleotide substitution (Fitch and Markowitz 1970; Galtier 2001; Penny et al. 2001). The “random local clock” model (Drummond and Suchard 2010) used to study variation in the rate of nucleotide substitution has also been used to study evolutionary rate variation in mimetic color pattern (Davis Rabosky et al. 2016) and reproductive mode parity (King and Lee 2015) in squamate reptiles.

In this paper, we describe a method for modeling evolutionary rate heterogeneity in a trait with two discrete states and implement it using the Bayesian Analysis of Macroevolutionary Mixtures (BAMM) framework (Rabosky 2014; Rabosky et al. 2014). The model discussed here is closely related to several existing phylogeny inference methods that model among-lineage substitution rate variation (Huelsenbeck 2000; Drummond and Suchard 2010) but differs in details of likelihood calculation and implementation. The general approach assumes that the present-day distribution of a binary state character is shaped by a mixture of stochastic processes in which the rate of evolutionary transition between the two states experiences shifts under a Poisson process across the branches of a phylogeny. The number and location of rate changes, which we refer to as rate-shift events, are inferred automatically from the data. Simulations reveal that the method accurately infers rates of evolution, the number and location of rate-shift events, and ancestral character states even when simulated data violate model assumptions.

## MATERIALS & METHODS

### *Likelihood of a binary state character*

We assume that each branch of a phylogeny evolves independently of the others and that the probability of a lineage transitioning to a character state different from its current state does not depend on its prior history of trait evolution (Pagel 1994). For a discrete character with two states, state 0 and state 1, trait evolution is modeled by a “forward” transition rate (denoted  $q_{01}$ ), which governs how frequently a lineage in state 0 changes to state 1, and a “reverse” transition rate (denoted  $q_{10}$ ), which governs how frequently a lineage in state 1 changes to state 0. For example, over a sufficiently small interval time  $\Delta t$ , the probability of observing a transition from state 0 to state 1 is approximately  $q_{01}\Delta t$ .

To write down a likelihood function for estimating the transition rate parameters, we follow the approach of Maddison et al. (2007) and define  $D_{N0}(t)$  to be the probability that lineage  $N$  evolves the distribution of character states observed among its descendants given that it is in state 0 at time  $t$ . We define  $D_{N1}(t)$  analogously. Next, consider what can happen in the short interval of time between  $t$  and  $t+h$ , where  $t+h$  is closer to the root and where  $h$  is taken to be an interval of time small enough that the probability of more than one character state transition in the interval is negligible. There are only two possibilities. Either the lineage remains in the state that it was in at time  $t+h$  or it switches to the other state. We can therefore write  $D_{N0}(t+h)$  and  $D_{N1}(t+h)$  as functions of  $D_{N0}(t)$ ,  $D_{N1}(t)$ , and the transition rate parameters  $q_{01}$  and  $q_{10}$ ,

$$D_{N0}(t+h) = (1 - q_{01}h)D_{N0}(t) + q_{01}hD_{N1}(t)$$

$$D_{N1}(t+h) = (1 - q_{10}h)D_{N1}(t) + q_{10}hD_{N0}(t)$$

By rearranging and letting  $h$  approach zero we form two coupled differential equations that describe how these probabilities change through time,

$$\frac{dD_{N0}(t)}{dt} = -q_{01}D_{N0}(t) + q_{01}D_{N1}(t)$$

$$\frac{dD_{N1}(t)}{dt} = -q_{10}D_{N1}(t) + q_{10}D_{N0}(t)$$

These can be solved (see appendix) to give the closed form solutions,

$$D_{N0}(t) = A + B_0 e^{-(q_{01}+q_{10})(t-t_0)} \quad (1)$$

$$D_{N1}(t) = A + B_1 e^{-(q_{01}+q_{10})(t-t_0)} \quad (2)$$

Where  $t_0$  is the initial time and,

$$A = \pi_0 D_{N0}(t_0) + \pi_1 D_{N1}(t_0)$$

$$B_0 = \pi_1 D_{N0}(t_0) - \pi_1 D_{N1}(t_0)$$

$$B_1 = \pi_0 D_{N1}(t_0) - \pi_1 D_{N0}(t_0)$$

Where  $\pi_i$  is the equilibrium frequency of state  $i$ . Equations (1) and (2) are evaluated for each branch of the phylogeny proceeding from the tips to the root in a post-order traversal. The quantities  $D_{N0}(t_0)$  and  $D_{N1}(t_0)$  are the initial conditions used to begin evaluation for each branch. If we have these values, we can compute the conditional likelihood of the data for a branch's stem group by simply setting  $t$  equal to the time at the base (rootward) of the branch and  $t_0$  equal to the time at the head (tipward) of the branch and evaluating equations (1) and (2). At each internal node, we form a new set of initial conditions by multiplying the  $D_{.0}$  and  $D_{.1}$  at the base of the node's left descendant branch with those of its right descendant. When we reach the root,  $R$ , of the tree  $D_{R0}(t_R)$  and  $D_{R1}(t_R)$  yield the probability of the data given the transition rate parameters and conditional on the root being in state 0 or state 1, respectively. To get the unconditional likelihood we must combine these two values, but doing so requires knowing the

probability of the root being in state 0 or in state 1. We assume that the root is in state 0 or state 1 with probabilities implied by their conditional likelihoods and compute the full likelihood as  $\frac{D_{R0}(t_R)}{D_{R0}(t_R)+D_{R1}(t_R)} D_{R0}(t_R) + \frac{D_{R1}(t_R)}{D_{R0}(t_R)+D_{R1}(t_R)} D_{R1}(t_R)$ . This is the same weighting scheme used in some models of trait-dependent speciation and extinction (FitzJohn et al. 2009). To begin computation, the initial conditions for each tip in the tree are set to  $D_{\cdot 0}(0) = 1$  and  $D_{\cdot 1}(0) = 0$  if the tip is in state 0 and vice versa if the tip is in state 1.

### *Rate-shift model for a binary state character*

At broad phylogenetic scales, it may be unreasonable to assume that the transition rates remain constant over all regions of phylogeny. We therefore allow regions of a phylogeny to belong to different macroevolutionary “rate-regimes”, which are independent sets of transition rate parameters that describe the evolution of a binary state character over the regions of phylogeny to which they pertain. Transitions between rate-regimes, which we refer to as “rate-shift events”, are assumed to occur along the branches of a phylogeny according to a compound Poisson process (Huelsenbeck 2000; Blanquart and Lartillot 2006; Rabosky 2014). To compute the likelihood of the data under a rate-shift model requires minimal modification of the process described in the previous section. Rather than traversing directly from a node to its ancestor when evaluating equations (1) and (2), we pause at each rate-shift event that occurs on the branch and use the values of (1) and (2) at that point as starting values for an additional evaluation of (1) and (2) under the new set of transition rates.

This model of rate variation is unsuitable when forward and reverse transition rates are asymmetric (Fig. 1). By proposing a new rate-shift event for every event of character state

change and simply making one rate arbitrarily large and the other rate arbitrarily small it is possible to fit any data set with probability 1 (when conditioned on the occurrence of the rate-shift events). This is because we can guarantee with probability 1 the origin and persistence of a derived character state by making the transition rate toward the derived character state arbitrarily large and the reverse rate arbitrarily small. In such a scenario, rate-shift events become decoupled from broad-scale among-lineage variation in the rate of trait evolution. Our implementation of an asymmetric version of this rate-shift model yielded results consistent with these expectations for several empirical datasets. For this reason, the default implementation in BAMM constrains forward and reverse transition rates to be identical. All analyses presented below use this symmetric implementation.

## *Implementation*

The rate-shift model described above is implemented in the Bayesian software program BAMM using reversible-jump Markov Chain Monte Carlo simulation (Rabosky 2014). Briefly, BAMM assumes that the number of rate-shift events on a phylogeny is drawn from a Poisson distribution with a rate parameter that is itself drawn from an exponential distribution. This formulation implies that the number of rate-shifts is drawn from a geometric distribution and that the expected number of rate-shifts, denoted by  $\Lambda$ , is simply the mean of the exponential hyperprior placed on the Poisson prior (Mitchell and Rabosky 2017). We extend the BAMM implementation for binary data by placing an exponential prior on the transition rate and use a proportional shrinking-expanding proposal mechanism to update its value. All other details remain the same and are described elsewhere (Mitchell and Rabosky 2017).



## *Simulation study*

We conducted a simulation study to assess how well the method estimates branch specific rates of evolution and infers the presence of rate-shift events. We also evaluated whether the estimated rates of trait evolution accurately reconstruct ancestral character states.

### (1) Choice of phylogeny

To carry out the simulations, we generated 100 phylogenies having between 50 and 1,000 tips by randomly sampling internal nodes from the 3,962-tip ultrametric squamate reptile phylogeny of Pyron and Burbrink (2014). Nodes were assigned weights such that all sizes (measured as the number of living descendants) of extracted clades had an equal probability of being selected. We chose to select subsets of a large empirical phylogeny, rather than simulated phylogenies, to introduce more realistic distributions of branch lengths than might be obtained using simple tree simulation models (e.g. Yule or constant-rate birth-death models).

### (2) Simulating trait evolution

For each phylogeny, we simulated evolution of a binary state trait 10 times using 2 different procedures. In the first case, we determined the number of rate-shift events to place on the tree by drawing a random integer from a Poisson distribution with a rate of 1. We determined the locations for these rate-shifts by selecting that number of internal nodes randomly without replacement, again using weights that gave all sizes of subtrees an equal probability of being chosen, and choosing a uniform random point along each chosen node's branch. We chose the transition rate for each rate-regime by drawing a random number from a log-normal distribution with a mean of  $\log 0.01$  and a standard deviation of  $-\frac{1}{3}\log 0.01$ . This corresponds to a log-

normal distribution with a spread that position rates of 0.0001 and 1 three standard deviations below and above the mean, respectively, and these values were chosen simply for their potential to generate datasets having a range of phylogenetic signals. The second case was identical to the first except that we allowed transition rates to be asymmetric. The degree of asymmetry was determined by drawing a random number from a log-normal distribution with a mean of  $\log 1$  and a standard deviation of  $\frac{1}{3} \log 100$ . This corresponds to a log-normal distribution centered on unbiased transition rates with a spread that positions a 100-fold bias in transition rates 3 standard deviations from the mean.

### (3) Information content of rate-shift events

Due to the stochastic nature of the simulations we expected rate-shift events to vary in their degree of detectability. To quantify this, we followed Rabosky et al. (2017) and calculated the “information content” of each rate-shift, which is a measure of how strongly the data support a model with rate variation over a model with no rate variation. For each simulated trait distribution, we optimized the value of the transition rate parameter to maximize the likelihood of the full data under a model with no rate-shifts. We denote this maximum-likelihood parameter estimate by  $\theta_{CR}$ . Next, we denote by  $D_i$  the trait data contained in the subtree formed by the set of all nodes and branch segments belonging to the  $i$ -th rate-regime and by  $\theta_i$  the transition rate that generated those data. We calculated the information content of the  $i$ -th rate regime as the difference in log likelihoods of  $D_i$  under the two parameter sets,

$$\Delta \text{Log } L_i = \log L(D_i | \theta_i) - \log L(D_i | \theta_{CR})$$

Where  $L(D_i | \theta_i)$  denotes the likelihood of the data in the  $i$ -th rate regime given the generating parameter, and  $\theta_{CR}$  indicates the transition rate obtained by maximizing the likelihood for the

full data under a model with no rate-shifts. If the data support a model with rate-shift events, the  $\Delta\text{Log } L$  statistic must be greater than 0. In general, we expect that this statistic must be substantially greater than 0 for BAMM to detect a rate-shift event. If  $\Delta\text{Log } L$  is expressed using the Akaike Information Criterion it can be rewritten as  $\Delta\text{Log } L = S/2 + k$ , where  $S$  is the difference in AIC scores needed to accept a model with an additional rate-shift and  $k$  is the number of extra parameters required to fit an additional rate-shift (in our implementation  $k = 2$ , corresponding to the location of the rate-shift and its rate parameter). If  $S \geq 0$  is interpreted as support for a model with an extra rate-shift, the minimum  $\Delta\text{Log } L$  needed to detect an event is 2.

#### *BAMM analysis*

We analyzed each simulation with BAMM using a single Markov chain that ran for ten million generations with  $\Lambda$  set equal to 1. The transition rates governing trait evolution need starting values before BAMM can proceed. For each simulation, we divided the total number of observed character state transitions by the summed branch length of the simulation's phylogeny and used this value for the initial transition rate and for the median of the exponential prior placed on the transition rate. This number is unavailable in empirical datasets but will be close to the number implied by parsimony when rates of character evolution are low.

#### *Performance assessment*

For each dataset simulated under a symmetric or asymmetric rate-shift model we assessed BAMM's ability to estimate rates of trait evolution and to detect rate-shift events. We performed each assessment using the estimated posterior distribution after discarding the first ten-percent of samples.

## (1) Estimating the rate of trait evolution

To determine how accurately BAMM infers rates of trait evolution we scaled each branch length to correspond to its average estimated rate of trait evolution. We performed this branch-by-branch calculation for each sample in the simulated posterior distribution using the BAMM-estimated transition rate parameters and assigned each branch an overall length equal to,

$$v_i = \frac{\sum_s v_{i,s}}{N}$$

Where  $v_{i,s}$  denotes the rate of evolution of branch  $i$  in the  $s$ -th posterior sample, and  $N$  indicates the number of samples in the posterior distribution. Using these values, we computed the tree-wide proportional branch rate error for a simulation as,

$$R = \sum_i \frac{t_i}{T} (\log v_i - \log V_i)$$

Where  $V_i$  denotes the true rate of evolution of branch  $i$ , and the summation is taken over all branches in the phylogeny. The error of the  $i$ -th branch estimate was weighted by the proportional contribution of its branch length,  $t_i$ , to the sum of all branch lengths,  $T$ , in the phylogeny. For unbiased transition rate estimates this equation is equal to 0.

## (2) Detection of rate-shift events

For each simulated rate-shift event, we computed the mean detection accuracy as  $\frac{\sum_s J_s}{N}$ , where  $J_s$  indicates the detection accuracy in the  $s$ -th posterior sample and  $N$  indicates the number of samples in the posterior distribution. Letting  $C_s$  denote the set of rate-shift events detected by BAMM in the  $s$ -th posterior sample, we measured the detection accuracy of a generating rate-shift event as,

$$J_s = \max \left( \left\{ \frac{|N \cap N_i|}{|N \cup N_i|}, i \in C_s \right\} \right)$$

Where  $N$  is the set of tips descended from the generating rate-shift event and  $N_i$  is the set of tips descended from the  $i$ -th rate-shift detected by BAMM in the  $s$ -th posterior sample. A value of 1 occurs when BAMM identifies the precise node that corresponds to the generating rate-shift event, and a value of 0 occurs when BAMM fails to identify the generating event at all. Intermediate values occur when BAMM identifies rate-shift events that correspond to nodes above or below the node where the generating rate-shift event occurred.

### (3) Empirical application

Finally, we analyzed an empirical dataset of mimetic coloration in snakes previously analyzed with several related methods (Davis Rabosky et al. 2016). Red-black banded coloration arises repeatedly among harmless colubrid snakes that occur in broad sympatry with dangerously venomous red-black banded coral snakes in the Neotropics and parts of the North temperate zone. The incidence of red-black banded coloration is particularly high among dipsadine snakes, a highly diverse clade of colubrid snakes that occur in local and regional sympatry with coral snakes across the Neotropics. Using the random local clock model (Drummond and Suchard 2010) and a Medusa-like (Alfaro et al. 2009) model of discrete trait evolution, Davis Rabosky et al. (2016) inferred elevated rates of mimetic color evolution within dipsadine snakes resulting from repeated independent origins of red-black banding coincident with the diversification of coral snakes in the Neotropics. We repeated their macroevolutionary rate analysis of this dataset using the method developed in this paper. We divided the number of parsimony-implied character state changes by the total branch length of the phylogeny to obtain a median rate for the transition rate prior and ran BAMM for 10 million generations with  $\Lambda$  set equal to 10. We

repeated this analysis 3 times under 3 different transition rate prior specifications corresponding to a 2-, 5-, and 10-fold speedup over the median rate implied by parsimony.

## RESULTS

### (1) Estimating the rate of trait evolution

Tree-wide proportional branch rate errors were low regardless of whether the data were simulated under a symmetric (mean = 0.09, median = 0.08) or asymmetric (mean = -0.27, median = -0.17) rate-shift model (Fig. 2). The correlations between estimated and true branch rates were high when data were simulated using a symmetric model (Pearson's  $\rho = 0.64$ , Spearman's  $\rho = 0.90$ ) but were substantially noisier when data were simulated using an asymmetric model (Pearson's  $\rho = 0.26$ , Spearman's  $\rho = 0.60$ ). When branch rates were multiplied by the temporal duration of each branch to convert them into the expected number of character state changes these correlations improved, particularly for the asymmetric model (*symmetric*: Pearson's  $\rho = 0.66$ , Spearman's  $\rho = 0.93$ ; *asymmetric*: Pearson's  $\rho = 0.37$ , Spearman's  $\rho = 0.75$ ). Spearman rank correlations were substantially higher than Pearson product-moment correlations for both models, indicating that relative branch rates are generally better estimated than absolute branch rates.

### (2) Detection of rate-shift events

BAMM's ability to detect rate-shift events was generally low due to the limited information content of rate-shift events in the simulated data (Fig. 4). Despite relative rate differences between ancestral and derived rate-shift events that varied over 6 orders of magnitude, only 400 of 2045 simulated rate-shift events had an information content above 2, the theoretical minimum

above which BAMM is expected to have power to detect them (see Discussion). This is due in part to the small average number of tips in simulated rate-regimes but also to the legitimate difficulty of simulating binary character data that reveal strong evidence for rate heterogeneity (cf. Fig. 4 rightmost panel). When BAMM has information to detect rate-shift events, however, it does reasonably well, and its performance improves monotonically as the information content of rate-shifts increases (cf. Fig. 4 leftmost panel). For example, BAMM detected the locations of 36 rate-shifts with an information content of at least 10 with a mean accuracy of 88%.

### (3) Empirical application

Analysis of the empirical dataset of red-black banded coloration in snakes largely recapitulated previous results showing an increased rate of trait evolution in Neotropical dipsadine snakes (but also revealed high rates in a clade of North temperate colubrine snakes) (Fig. 5). When analyzed under a strong, well-informed transition rate prior, rates of trait evolution ranged from a low of 0.00065 My<sup>-1</sup> in basal snake lineages to a high of 0.0093 My<sup>-1</sup> in dipsadine snakes, with an overall mean of 0.0024 My<sup>-1</sup> that closely matched the overall rate of 0.002 My<sup>-1</sup> obtained by dividing the number of parsimony-inferred state changes by the total branch length (cf. Fig. 5 leftmost panel). More liberal priors did not change this general picture, however rates of trait evolution in the upper tail of the estimated branch rate distribution tended to creep upward as the transition rate prior flattened. Flatter transition rate priors were associated with a higher number of posterior rate-shifts caused by the partitioning of single rate-shift events inferred under steeper transition rate priors into multiple smaller rate-shift events (i.e. lineages left out of these new rate-shift events fell back into the ancestral root regime). These new rate-shift events tended to

have estimated rates of evolution that were higher than estimates made using steeper transition rate priors.

## DISCUSSION

In this paper, we describe a Bayesian method, implemented in the BAMM software program, for studying among-lineage variation in the rate of evolution of a binary character. Overall, our results show that the method accurately infers rates of trait evolution and the presence and location of among-lineage evolutionary rate variation, even when simulated data violate model assumptions. Although the method performed well on many simulated data sets, we caution that overall power for inferring heterogeneous dynamics of single binary traits may be low.

The ability of the method to detect rate-shift events depends on the size of the clade belonging to a rate-regime but also on how much information the data contain with respect to the parameters of the rate-regime. In our simulations, we estimated this information content using a log-likelihood ratio that measures the likelihood of a given rate-shift event under the true parameters relative to the corresponding likelihood under a simple model where the rate is set to the whole-tree average. This information content can be surprisingly low even for large clades. In retrospect, this is not necessarily surprising because with only two character states there is a limit to how different data generated by two different rates can become. A 10- or even 100-fold rate increase will not be detectable if the ancestral rate is already high enough as to leave no phylogenetic signal. Similarly, a 10- or 100-fold rate decrease will not be detectable if the ancestral rate is already so low as to make character change highly improbable. In general, we expect detectability of rate-shifts to depend strongly on how distinct one rate's phylogenetic



signal is from another. This will have a strong stochastic component to it, and for binary data will likely have a low signal-to-noise ratio making detection of rate-shift events difficult.

While we did not explore prior sensitivity exhaustively in this study, the empirical results indicate that branch rate estimates (and ancestral state reconstructions by implication) are sensitive to the transition rate prior. In the empirical example, disagreements among the different priors occur in regions of phylogeny having elevated rates of evolution. This is not surprising given that the method works with only a single binary character and that the fast-evolving clades in the empirical phylogeny have relatively few taxa, but it does call for vigilance. A sensible rule of thumb is to treat with caution any result where the overall mean rate estimate disagrees strongly with the rate implied by parsimony. Encouragingly, the different prior specifications are in broad agreement on where relative rate differences occur in the empirical phylogeny despite disagreements over absolute rate estimates. The results also indicate that the transition rate prior interacts with the prior on the number of rate-shifts, suggesting that the method's ability to estimate the precise location of a rate-shift and its ability to estimate the associated rate of evolution may trade-off. Finally, the combination of a prior with a high number of expected rate-shift events and a relatively flat transition rate prior can lead to an apparent abundance of single tips with derived character states having elevated rates of evolution. This is because dropping a high rate of evolution on such a branch entails no penalty. It makes the derived state more probable and with only a single lineage does not suffer from the likelihood penalty that a larger clade fixed for a derived state would suffer from if given an elevated rate. Users should remain alert to this scenario and treat its presence as an indication that the priors are exerting an undue influence. For all these reasons, we recommend the use of a strong, well-informed prior on the

transition rate and setting the median rate of the prior equal to the parsimony-implied rate seems like a sensible choice.

A fundamental difficulty facing macroevolutionary models of discrete character evolution is that the data contain low information content with respect to rates of evolution since actual events of character state change are not directly observed. This challenge is exacerbated further in asymmetric Markov models which must estimate the rate of evolution while simultaneously inferring the equilibrium frequencies of the character states. From this fact alone we should expect asymmetric Markov models to have lower information content with respect to transition rates than symmetric Markov models, which a priori assume equilibrium frequencies of character states are equal. In a discrete-time version of the Markov model used by this and many other studies, Sanderson (1993) has shown that maximum likelihood parameter estimates are intrinsically biased upwards for an asymmetric model with two character states but are unbiased for a symmetric model. The extent to which these conclusions apply in continuous-time or generalize to more than two states is not known to us but warrants further study.

One consequence of the low information content of binary data are ancestral state reconstructions that may appear nonsensical when performed with asymmetric models. For example, Pagel (1999) presented a comb phylogeny in which every tip in the clade was fixed for one of two possible character states and showed that an asymmetric Markov model reconstructed the root as belonging to either state with equal probability. From an asymmetric model's perspective, a clade that is almost entirely fixed for a single character state is likely to have been generated by a process with a very high transition rate toward the majority state and a very low

transition rate away from it, which makes the tip states mostly independent of the ancestral states because the asymmetry in rates will yield the same outcome at the tips regardless of the assignment of states to internal nodes. By contrast, from a symmetric model's perspective the clade simply has a very low rate of evolution and the tip states resemble their ancestors as a result. In fact, transition rates are nearly always biased in the direction of the state appearing most frequently among the tips of the tree (Nosil and Mooers 2005; Maddison 2006). This means that in an analysis that infers a high rate of transition from state 0 to state 1 and low rate of transition from state 1 to state 0, it will sometimes be the case that a majority of the evolutionary transitions were from state 1 to state 0. This runs counter to intuition and indicates the need for caution when using transition rates to infer directionality in the history of trait evolution (Nosil and Mooers 2005; Goldberg and Igic 2008).

## LITERATURE CITED

- Alfaro M.E., Santini F., Brock C., Alamillo H., Dornburg A., Rabosky D.L., Carnevale G., Harmon L.J. 2009. Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. *PNAS*. 106:13410–13414.
- Beaulieu J.M., O'Meara B.C., Donoghue M.J. 2013. Identifying Hidden Rate Changes in the Evolution of a Binary Morphological Character: The Evolution of Plant Habit in Campanulid Angiosperms. *Syst Biol*. 62:725–737.
- Blackburn D.G. 1982. Evolutionary Origins of Viviparity in the Reptilia. I. Sauria. *Amphibia-Reptilia*. 3:185–205.
- Blackburn D.G. 1985. Evolutionary Origins of Viviparity in the Reptilia. II. Serpentes, Amphisbaenia, and Ichthyosauria. *Amphibia-Reptilia*. 6:259–291.
- Blanquart S., Lartillot N. 2006. A Bayesian Compound Stochastic Process for Modeling Nonstationary and Nonhomogeneous Sequence Evolution. *Mol Biol Evol*. 23:2058–2071.
- Christin P.-A., Osborne C.P., Chatelet D.S., Columbus J.T., Besnard G., Hodkinson T.R., Garrison L.M., Vorontsova M.S., Edwards E.J. 2013. Anatomical enablers and the evolution of C4 photosynthesis in grasses. *PNAS*. 110:1381–1386.

438 Davis Rabosky A.R., Cox C.L., Rabosky D.L., Title P.O., Holmes I.A., Feldman A., McGuire  
439 J.A. 2016. Coral snakes predict the evolution of mimicry across New World snakes.  
440 Nature Communications. 7:1–9.

441 Drummond A.J., Suchard M.A. 2010. Bayesian random local clocks, or one rate to rule them all.  
442 BMC Biology. 8:114.

443 Eastman J.M., Alfaro M.E., Joyce P., Hipp A.L., Harmon L.J. 2011. A Novel Comparative  
444 Method for Identifying Shifts in the Rate of Character Evolution on Trees. *Evolution*.  
445 65:3578–3589.

446 Fitch W.M., Markowitz E. 1970. An improved method for determining codon variability in a  
447 gene and its application to the rate of fixation of mutations in evolution. *Biochem Genet*.  
448 4:579–593.

449 FitzJohn R.G., Maddison W.P., Otto S.P. 2009. Estimating Trait-Dependent Speciation and  
450 Extinction Rates from Incompletely Resolved Phylogenies. *Syst Biol*. 58:595–611.

451 Galtier N. 2001. Maximum-Likelihood Phylogenetic Analysis Under a Covarion-like Model.  
452 *Mol Biol Evol*. 18:866–873.

453 Goldberg E.E., Igić B. 2008. On Phylogenetic Tests of Irreversible Evolution. *Evolution*.  
454 62:2727–2741.

455 Heath T.A., Holder M.T., Huelsenbeck J.P. 2012. A Dirichlet Process Prior for Estimating  
456 Lineage-Specific Substitution Rates. *Mol Biol Evol*. 29:939–955.

457 Huelsenbeck J.P., Larget B., Swofford D. 2000. A Compound Poisson Process for Relaxing the  
458 Molecular Clock. *Genetics*. 154:1879–1892.

459 Igea J., Miller E.F., Papadopoulos A.S.T., Tanentzap A.J. 2017. Seed size and its rate of evolution  
460 correlate with species diversification across angiosperms. *PLOS Biology*. 15:e2002792.

461 King B., Lee M.S.Y. 2015. Ancestral State Reconstruction, Rate Heterogeneity, and the  
462 Evolution of Reptile Viviparity. *Syst Biol*. 64:532–544.

463 Landis M.J., Schraiber J.G., Liang M. 2013. Phylogenetic Analysis Using Lévy Processes:  
464 Finding Jumps in the Evolution of Continuous Traits. *Syst Biol*. 62:193–204.

465 Lloyd G.T., Wang S.C., Brusatte S.L. 2012. Identifying Heterogeneity in Rates of Morphological  
466 Evolution: Discrete Character Change in the Evolution of Lungfish (sarcopterygii;  
467 Dipnoi). *Evolution*. 66:330–348.

468 Maddison W.P. 1990. A Method for Testing the Correlated Evolution of Two Binary Characters:  
469 Are Gains or Losses Concentrated on Certain Branches of a Phylogenetic Tree?  
470 *Evolution*. 44:539–557.

471 Maddison W.P. 2006. Confounding Asymmetries in Evolutionary Diversification and Character  
472 Change. *Evolution*. 60:1743–1746.

473 Maddison W.P., Midford P.E., Otto S.P. 2007. Estimating a Binary Character’s Effect on  
474 Speciation and Extinction. *Syst Biol*. 56:701–710.

475 Marazzi B., Ané C., Simon M.F., Delgado-Salinas A., Luckow M., Sanderson M.J. 2012.  
476 Locating Evolutionary Precursors on a Phylogenetic Tree. *Evolution*. 66:3918–3930.

477 Mitchell J.S., Rabosky D.L. 2017. Bayesian model selection with BAMM: effects of the model  
478 prior on the inferred number of diversification shifts. *Methods in Ecology and Evolution*.  
479 8:37–46.

480 Nosil P., Mooers A.Ø. 2005. Testing Hypotheses About Ecological Specialization Using  
481 Phylogenetic Trees. *Evolution*. 59:2256–2263.

482 O’Meara B.C., Ané C., Sanderson M.J., Wainwright P.C. 2006. Testing for Different Rates of  
483 Continuous Trait Evolution Using Likelihood. *Evolution*. 60:922–933.

484 Pagel M. 1994. Detecting correlated evolution on phylogenies: a general method for the  
485 comparative analysis of discrete characters. *Proceedings of the Royal Society of London. Series B: Biological Sciences*. 255:37–45.

487 Pagel M. 1999. The Maximum Likelihood Approach to Reconstructing Ancestral Character  
488 States of Discrete Characters on Phylogenies. *Systematic Biology*. 48:612–622.

489 Penny D., McComish B.J., Charleston M.A., Hendy M.D. 2001. Mathematical Elegance with  
490 Biochemical Realism: The Covarion Model of Molecular Evolution. *J Mol Evol*. 53:711–  
491 723.

492 Pyron R.A., Burbrink F.T. 2014. Early origin of viviparity and multiple reversions to oviparity in  
493 squamate reptiles. *Ecology Letters*. 17:13–21.

494 de Queiroz A., Rodríguez-Robles J.A. 2006. Historical Contingency and Animal Diets: The  
495 Origins of Egg Eating in Snakes. *The American Naturalist*. 167:684–694.

496 Rabosky D.L. 2014. Automatic Detection of Key Innovations, Rate Shifts, and Diversity-  
497 Dependence on Phylogenetic Trees. *PLOS ONE*. 9:e89543.

498 Rabosky D.L., Grudler M., Anderson C., Title P., Shi J.J., Brown J.W., Huang H., Larson J.G.  
499 2014. BAMMtools: an R package for the analysis of evolutionary dynamics on  
500 phylogenetic trees. *Methods in Ecology and Evolution*. 5:701–707.

501 Rabosky D.L., Mitchell J.S., Chang J. 2017. Is BAMM Flawed? Theoretical and Practical  
502 Concerns in the Analysis of Multi-Rate Diversification Models. *Syst Biol*. 66:477–498.

- Rabosky D.L., Santini F., Eastman J., Smith S.A., Sidlauskas B., Chang J., Alfaro M.E. 2013. Rates of speciation and morphological evolution are correlated across the largest vertebrate radiation. *Nature Communications*. 4:1–8.
- Revell L.J., Collar D.C. 2009. Phylogenetic Analysis of the Evolutionary Correlation Using Likelihood. *Evolution*. 63:1090–1100.
- Sanderson M.J. 1993. Reversibility in Evolution: A Maximum Likelihood Approach to Character Gain/Loss Bias in Phylogenies. *Evolution*. 47:236–252.
- Stanley S.M. 1979. *Macroevolution, pattern and process*. San Francisco: W. H. Freeman.
- Uyeda J.C., Harmon L.J. 2014. A Novel Bayesian Method for Inferring and Interpreting the Dynamics of Adaptive Landscapes from Phylogenetic Comparative Data. *Syst Biol*. 63:902–918.
- Vermeij G.J. 1973. Adaptation, Versatility, and Evolution. *Syst Biol*. 22:466–477.
- Weber M.G., Keeler K.H. 2013. The phylogenetic distribution of extrafloral nectaries in plants. *Ann Bot*. 111:1251–1261.

## APPENDIX

### *Derivation of likelihood equations (1) and (2)*

Begin with the differential equations,

$$\frac{dD_{N0}(t)}{dt} = -q_{01}D_{N0}(t) + q_{01}D_{N1}(t) \quad (A1)$$

$$\frac{dD_{N1}(t)}{dt} = -q_{10}D_{N1}(t) + q_{10}D_{N0}(t) \quad (A2)$$

By solving (A1) for  $D_{N1}(t)$  and equating its derivative to (A2) we can form the second-order differential equation,

$$\frac{d^2D_{N0}(t)}{dt^2} + (q_{01} + q_{10})\frac{dD_{N0}(t)}{dt} = 0$$

Which, after finding the roots of its auxiliary equation, has the general solution,

$$D_{N0}(t) = c_1 + c_2 e^{-(q_{01}+q_{10})(t-t_0)} \quad (A3)$$

Where  $t_0$  is a starting time on node  $N$ 's branch,  $c_1$  and  $c_2$  are constants, and  $t$  is larger than  $t_0$  but smaller than the time at the base of node  $N$ 's branch.

530

To solve for  $c_1$  note that at our initial condition when  $t = t_0$ ,

$$c_2 = D_{N0}(t_0) - c_1 \quad (A4)$$

Furthermore, differentiating equation (A3) gives,

$$\frac{dD_{N0}(t)}{dt} = -(q_{01} + q_{10})c_2 e^{-(q_{01}+q_{10})(t-t_0)} \quad (A5)$$

And equating (A1) with (A5) and setting  $t = t_0$  yields, after some algebra,

$$q_{01}c_1 - q_{10}c_2 = q_{01}D_{N1}(t_0) \quad (A6)$$

By substituting the right-hand side of (A4) for  $c_2$  in (A6) and solving for  $c_1$  we find,

$$c_1 = \frac{q_{01}D_{N1}(t_0) + q_{10}D_{N0}(t_0)}{q_{01} + q_{10}} \quad (A7)$$

Finally, by substituting the right-hand side of (A7) for  $c_1$  in (A6) and solving for  $c_2$  we find,

$$c_2 = \frac{q_{01}D_{N0}(t_0) - q_{01}D_{N1}(t_0)}{q_{01} + q_{10}} \quad (A8)$$

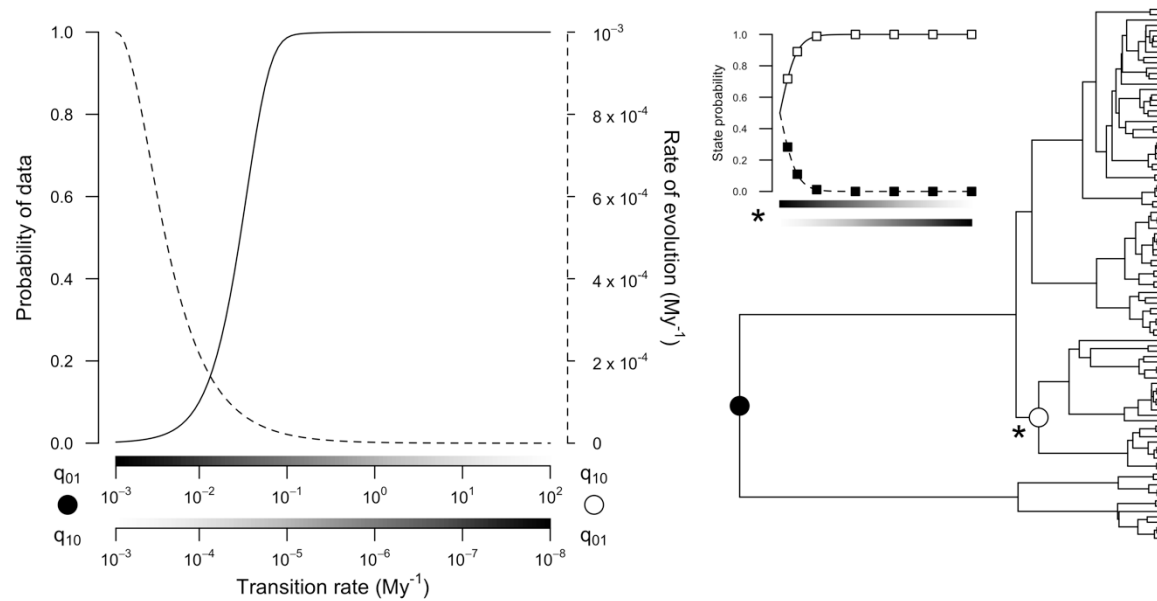
Thus, equation (A3) is equivalent to equation (1) after substitution of (A7) and (A8), noting that

$$\frac{q_{01}}{q_{01} + q_{10}} = \pi_1 \text{ and that } \frac{q_{10}}{q_{01} + q_{10}} = \pi_0. \text{ An equivalent derivation of equation (2) is performed by}$$

starting with  $D_{N1}(t) = c_1 + c_2 e^{-(q_{01}+q_{10})(t-t_0)}$ . When  $\pi_0 = \pi_1 = \frac{1}{2}$  as in the BAMM

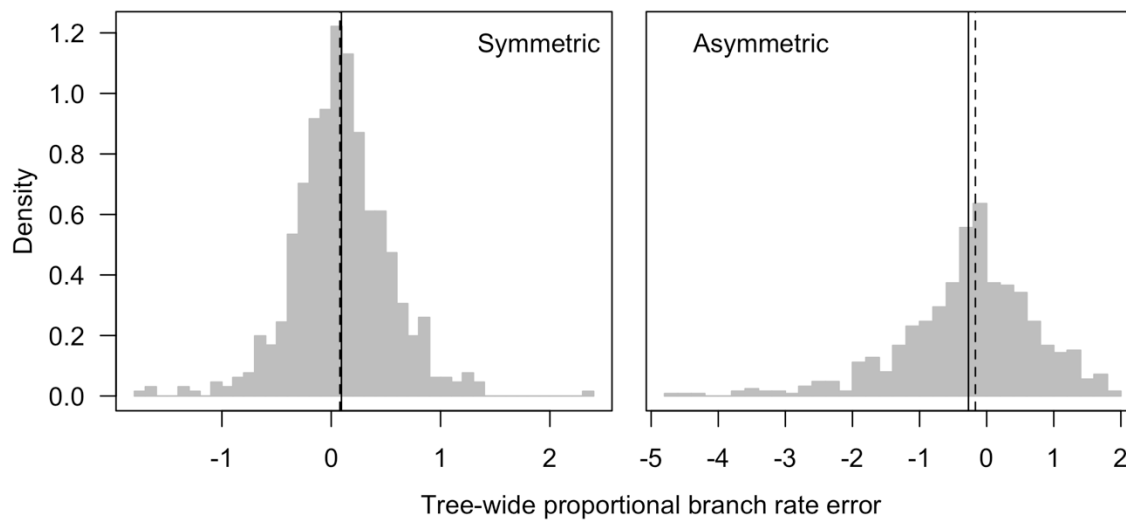
implementation this is simply a special case of the JC69 model (Jukes and Cantor 1969) with two states.

546  
547  
548  
549  
550

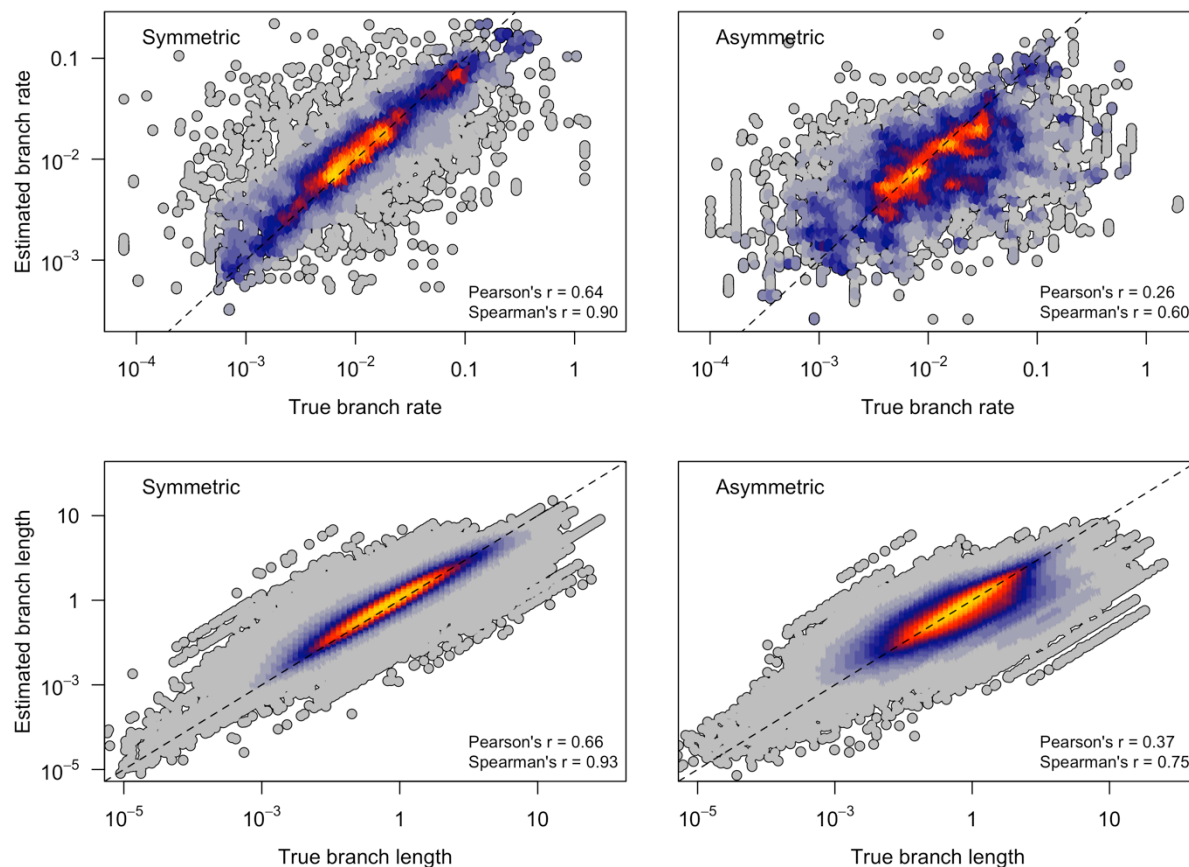


*Figure 1.* A worked example showing why asymmetric Markov models are unsuitable for BAMM-type rate-shift models. A single rate-shift event (denoted by a white circle) is placed along the branch subtending the clade fixed for character state 0. The  $x$ -axes in the left plot depict transition rates for this rate-shift and for the root event (denoted by a black circle). Note that they run in opposite directions, e.g. the upper  $x$ -axis is  $q_{01}$  for the root event but is  $q_{10}$  for the rate-shift event. As the asymmetry in transition rates is increased the probability of the data (denoted by the solid line) rises to 1 while the overall rate of character evolution (denoted by the dashed line) falls to 0. Simultaneously, as the rate of character evolution falls to 0 the asymmetry in transition rates causes the prior probability of being in one or the other character state to rise to 1, e.g. the inset plot denoted by an asterisk depicts the probability of being in state 0 or state 1 as a function of the asymmetric transition rates of the rate-shift event. Thus, at the extreme a “rate-shift” event simply introduces a second way to observe a stochastic event of character state change and does not correspond to among-lineage heterogeneity in rates of character evolution.

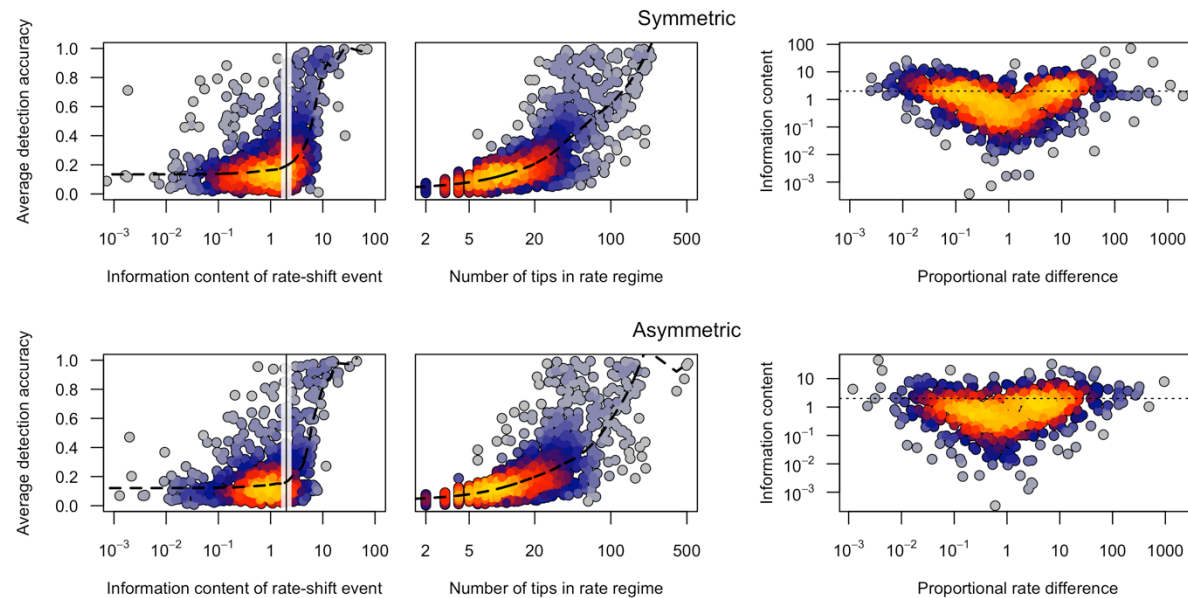




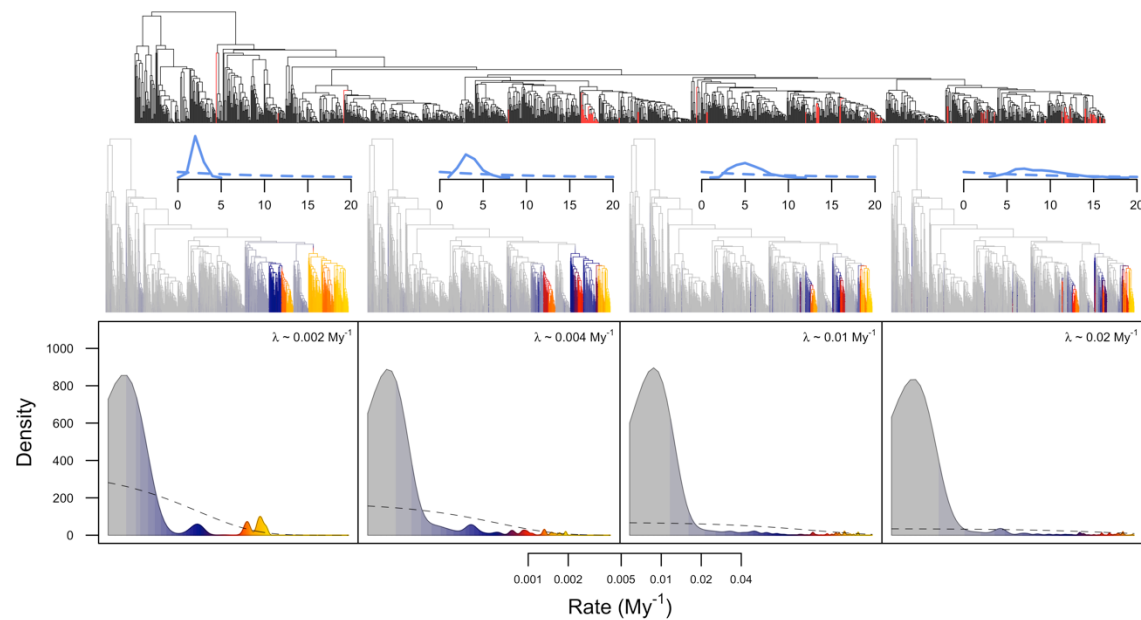
*Figure 2.* Tree-wide proportional branch rate error in BAMM-estimated branch rates when data are simulated under a symmetric rate-shift model (*left*) or an asymmetric rate-shift model (*right*). Branch rates are measured as the expected number of character transitions per million years. Tree-wide proportional branch rate error is a weighted sum of the logarithmic difference between estimated branch rates and true branch rates over all branches in a phylogeny. Errors on short branches are down-weighted relative to errors on long branches. Unbiased estimates have an error of 0, negative and positive values correspond to under- and over-estimation errors, respectively. Each histogram depicts the distribution of tree-wide proportional branch rate error over simulations with at least one rate-shift event. The vertical lines show the mean (solid) and median (dashed) of each distribution.



580  
581 *Figure 3.* Correlation between true branch rates/lengths and BAMM-estimated branch  
582 rates/lengths when data are simulated under a symmetric rate-shift model (*left panels*) or an  
583 asymmetric rate-shift model (*right panels*). Branch rates (*top panels*) are measured as the  
584 expected number of character transitions per million years. Branch lengths (*bottom panels*) are  
585 measured as the expected number of character transitions occurring along the length of a branch.  
586 Each point depicts the average posterior rate estimate for a single branch of a phylogeny in one  
587 simulation, and the color of each point corresponds to the density of neighboring points (warm  
588 colors indicate high densities). Only branches from phylogenies with at least one simulated rate-  
589 shift are represented. The one-to-one line is dashed.



**Figure 4.** Rate-shift detection accuracy with BAMM when data are simulated under a symmetric rate-shift model (*top panels*) or an asymmetric rate-shift model (*bottom panels*). The average detection accuracy measures how accurately BAMM inferred a rate-shift's location. The information content of a rate-shift event is a log-likelihood ratio that measures the likelihood of a given rate-shift event under the true parameters relative to the corresponding likelihood under a simple model where the rate is set to the whole-tree average. In the leftmost panels, the vertical line is the theoretical minimum log-likelihood ratio above which BAMM is expected to have power to detect a rate shift event, and it closely coincides with the upward inflection of the LOWESS regression lines (dashed). The rightmost panels plot an event's information content against its proportional rate difference. The proportional rate difference of an event is the ratio of its rate of evolution to the ancestral rate preceding it. In all panels, each point represents a simulated rate-shift event, and the color of each point corresponds to the density of neighboring points (warm colors indicate high densities).



*Figure 5.* Evolution of mimetic coloration in snakes under four different transition rate priors in BAMM. A high rate of evolution of red-black banded mimetic coloration in snakes is inferred in Neotropical dipsadine snakes and North American colubrids, although precise locations of rate-shifts and rate estimates differ across prior specifications. In the lower panels, the  $x$ -axis is the average posterior rate of trait evolution (expected number of character state changes per million years) and the  $y$ -axis is a kernel density estimate of the branch-specific rate distribution over 4 different prior specifications. The dashed line shows the prior distribution of the rate of evolution and the prior median is indicated by the inset expression. Average posterior branch rates are mapped onto the phylogeny in the center panels. The line graphs above each phylogeny show the posterior distribution of the number of rate-shift events (solid line) in the credible shift set compared to the prior distribution (dashed line). Branches colored red in the topmost panel show the distribution of red-black banded mimetic coloration among tips in the empirical phylogeny. Phylogeny from Pyron and Burbrink (2014).