Methods in Ecology and Evolution
British Ecological Society

Methods in Ecology and Evolution 2016 doi: 10.1111/2041-210X.12626

Bayesian model selection with BAMM: effects of the model prior on the inferred number of diversification shifts

Jonathan S. Mitchell* and Daniel L. Rabosky

Department of Ecology and Evolutionary Biology, Museum of Zoology, University of Michigan, Ann Arbor, MI 48109, USA

Summary

1. Understanding variation in rates of speciation and extinction – both among lineages and through time – is critical to the testing of many hypotheses about macroevolutionary processes. Bayesian Analysis of Macroevolutionary Mixtures (BAMM) is a flexible Bayesian framework for inferring the number and location of shifts in macroevolutionary rate across phylogenetic trees and has been widely used in empirical studies. BAMM requires that researchers specify a prior probability distribution on the number of diversification rate shifts before conducting an analysis. The consequences of this 'model prior' for inference are poorly known but could potentially influence both the probability of accepting models that are more (high error rate) or less (low power) complex than the generating model.

2. The hierarchical Poisson process prior in BAMM reduces to a simple geometric distribution on number of rate shifts, and we use this property to increase the efficiency of model selection with Bayes factors. Using BAMM v2.5, we analysed phylogenies simulated with and without diversification heterogeneity across a broad range of prior parameterizations. We also assessed the impact of the model prior on MCMC convergence times and on diversification rate estimates.

3. For all simulation scenarios, model evidence (Bayes factor support) for the number of shifts is not sensitive to the choice of model prior over the wide range examined here. The best-supported model found using BAMM rarely includes spurious shifts (<2% of all runs) when diversification models are selected using Bayes factors. BAMM was reliably able to infer the true number of diversification rate shifts across prior expectations that varied by three orders of magnitude. However, we find a strong effect of model prior on MCMC convergence properties: a flatter prior distribution (larger expected number of shifts) can dramatically increase the efficiency of the MCMC simulation.

4. Our results support the use of a liberal model prior in BAMM, as it reduces computation time without distorting the evidence for rate heterogeneity.

Key-words: Bayesian, Bayesian Analysis of Macroevolutionary Mixtures, birth–death model, macroevolution, rate variation

Introduction

Species richness is unequally partitioned across the tree of life, with some clades having far more species than their corresponding sister lineages. Understanding the root causes of this variation has long been a foundational research paradigm in macroevolution (Sloss 1950; Raup 1985; Jetz *et al.* 2012; Rabosky 2014). It is increasingly clear that much of the variation in species richness among clades involves differential rates of speciation and extinction (Jablonski 2008; Alfaro *et al.* 2009). Hypotheses to explain patterns of species richness range from the geographical complexity of regions in which different clades occur (e.g. Heaney 1986) to key innovations (e.g. Simpson 1953; Liem 1973; Coyne & Orr 2004; Jablonski 2008). However, the stochastic nature of the diversification process can lead to variation in species richness that is not associated with causal differences in macroevolutionary rates (Gould *et al.* 1977). Hence, robust tests of macroevolutionary hypotheses require methods that can identify differential rates of speciation and extinction across the tree of life (Slowinski & Guyer 1989; Phillimore & Price 2008; Rabosky 2014).

A number of methods have recently been developed that allow researchers to model heterogeneous rates of speciation and extinction across the branches of phylogenetic trees (Maddison, Midford & Otto 2007; Alfaro *et al.* 2009; FitzJohn, Maddison & Otto 2009; Morlon, Parsons & Plotkin 2011; Etienne & Haegeman 2012; Beaulieu & O'Meara 2016). Bayesian Analysis of Macroevolutionary Mixtures (BAMM; Rabosky 2014; Rabosky *et al.* 2014) is a method for automatically identifying heterogeneous mixtures of evolutionary rate regimes across time-calibrated phylogenetic trees of extant taxa that

*Correspondence author. E-mail: jonsmitc@umich.edu

has been widely applied to diverse empirical data sets. BAMM uses reversible-jump Markov chain Monte Carlo to approximate posterior distributions of diversification models, enabling researchers to reconstruct the number, magnitude and locations of rate shifts on phylogenetic trees. Shifts in evolutionary rates can occur along any branch of the phylogenetic tree, and the rates can vary through time within a rate regime (Rabosky 2014). This framework enables researchers to evaluate whether clades vary in their speciation or extinction rate without specifying particular clades to test *a priori* and can be used to assess the relationship between character states and lineage diversification rates (Rabosky & Huang 2015).

Perhaps, the most basic question that users seek to address with BAMM is whether a given data set contains evidence for variation in diversification rates among clades. BAMM simulates a posterior distribution of diversification models and can thus be used to compare the evidence favouring a simple model with no diversification heterogeneity to the evidence favouring models with more complex diversification dynamics. The complexity of diversification models sampled with BAMM is a function of the number of diversification rate shifts in the model (k). The prior distribution on k is the *model prior*, and we formally refer to a model with k shifts as model M_k. BAMM assumes that the number of rate shifts follows a Poisson distribution, where the rate parameter of the Poisson process is itself governed by an exponential hyperprior. This exponential hyperprior is specified *a priori* by users of BAMM (parameter 'poissonRatePrior').

In this article, we ask a simple question: Is model selection with BAMM sensitive to the prior on the number of diversification rate shifts? We have previously discussed the use of both posterior probabilities (Rabosky 2014) and Bayes factors (Rabosky *et al.* 2014) for inferring the number of diversification shifts. We explicitly compare these approaches as a function of the prior distribution on the number of rate shifts. We find that model posterior probabilities are only slightly influenced by the model prior. However, we demonstrate that Bayes factors are not sensitive to the model prior and we recommend their use for model selection with BAMM. We describe several practical scenarios where manipulation of the model prior can improve the statistical performance of BAMM.

Materials and methods

PRIOR PROBABILITY OF *K* SHIFTS IN BAMM

BAMM assumes that the number of rate shifts on the phylogeny is Poisson-distributed with a rate parameter Λ, but Λ is itself drawn from an exponential distribution with rate parameter θ. In the original implementation of BAMM, the program generated the prior distribution on the number of shifts using simulation. Here, we show that this distribution has a simple analytical form, enabling us to compute the exact prior probability of any model without recourse to simulation.

The probability of k shifts under the BAMM model is the product of Poisson and exponential densities. It is well documented in the probability literature (e.g. Grimmett & Stirzaker 2001) that a Poisson process with an exponentially distributed rate parameter simplifies to a geometric distribution with $p = 1/(\text{exponential mean})$. This basic result follows immediately from Greenwood & Yule (1920), who derived the negative binomial distribution as a mixture of Poisson random variables with gamma mixing weights. The exponential distribution is a special case of the gamma distribution, and the geometric is a special case of the negative binomial. Correspondingly, the geometric distribution is a special case of the gamma-Poisson mixture but where the gamma distribution is parameterized as a simple exponential (e.g. shape = 1, scale = 1/rate).

To demonstrate this property, note that we can integrate over the Poisson parameter Λ to express the probability density of the number of shifts as a function of exponential hyperprior θ.

$$\Pr(k|\theta) = \int_0^\infty \Pr(k|\Lambda)\Pr(\Lambda|\theta)d\Lambda \qquad \text{eqn 1}$$

This can be expanded to

$$\Pr(k|\theta) = \int_0^\infty \frac{\Lambda^k e^{-\Lambda}}{k!}\theta e^{-\theta\Lambda}d\Lambda \qquad \text{eqn 2}$$

and reduced to

$$= \frac{\theta}{k!}\int_0^\infty \Lambda^k e^{-\Lambda(1+\theta)}d\Lambda. \qquad \text{eqn 3}$$

Equation (3) follows immediately from a hierarchical Poisson–exponential model and can be solved analytically using a gamma function identity, specifically:

$$\int_0^\infty \Lambda^x e^{-a\Lambda}d\Lambda \rightarrow \frac{\Gamma(x+1)}{a^{x+1}} \qquad \text{eqn 4}$$

and thus the full expression becomes

$$\Pr(k|\theta) = \frac{\Gamma(x+1)}{k!}\frac{\theta}{(1+\theta)^{k+1}} = \frac{\theta}{(\theta+1)^{k+1}}. \qquad \text{eqn 5}$$

Letting $\gamma = 1/\theta$, we have the following:

$$\begin{aligned}\Pr(k|\gamma) &= \frac{1/\gamma}{((1/\gamma)+1)^{k+1}}\frac{1/\gamma}{((1/\gamma)+1)((1/\gamma)+1)^k} \\ &= \left(\frac{1}{(\gamma+1)}\right)\left(\frac{\gamma}{(\gamma+1)}\right)^k\end{aligned} \qquad \text{eqn 6}$$

which is simply a geometric distribution with parameter $p = 1/(\gamma + 1)$. This well-known mathematical result facilitates more rapid and comprehensive evaluation of the prior. The mean of the distribution is $(1 - p)/p$, meaning that the expected number of shifts under a particular exponential hyperprior is γ. We now explicitly reference the model prior in terms of the expected number of shifts, γ.

This analytical formulation of the prior probability has been implemented in BAMM tools (Rabosky *et al.* 2014; MEE) to facilitate diversification model selection. Importantly, the original release of BAMM (BAMM versions < 2.3.1) contained an error in the acceptance probability for MCMC moves that updated the Poisson rate parameter Λ (first identified by C. Ané; see Fig. 1). Because this error amplified the effects of the model prior on the posterior density of rate shifts, we refer to it as the 'incorrect model prior'. We are grateful to C. Ané and B. Larget for discussions that led to resolution of this issue and for checking (May 2015) the analytical solution given above. Use of the incorrect model prior would potentially have impacted BAMM analyses conducted prior to June 2015. However, despite the severity of the incorrect model prior for some parameterizations (see below; Fig. 1), our previous assessments of BAMM's performance (Rabosky 2014; Rabosky *et al.* 2014) nonetheless revealed good statistical performance using BAMM's default model prior parameterization (poissonRatePrior = 1; Fig. S1, Supporting Information).

The use of a Poisson prior with an exponentially distributed hyperprior (resulting in a geometric distribution of k) in BAMM allows for consistent results across BAMM runs and is a conservative prior (as the zero-shift model is always the most likely outcome). There are many possible alternative priors, such as a gamma hyperprior on the mean of the Poisson prior resulting in a negative binomial distribution of k. A negative binomial prior would allow studies to directly compare model support after putting stronger priors on different values of k shifts (e.g. comparing the output of a run where $k = 2$ has the highest prior probability to a run where $k = 3$ does). The negative binomial in general could also allow for a fatter tail to the distribution, potentially making it easier to explore complex models. The open-source nature of the BAMM software platform allows other workers to incorporate any alternative prior they choose.

MODEL SELECTION WITH BAYES FACTORS

The analytical expression above makes it trivial to compute the prior probability of a diversification model with k shifts under the process modelled by BAMM. Model posterior probabilities can be taken directly from BAMM output. For a model of order k, this is simply the frequency of posterior samples that includes k shifts. The Bayes factor evidence favouring one model over another is the ratio of marginal likelihoods of the two models, which is identical to the posterior odds ratio for the models divided by the prior odds ratio. For a model with k shifts, $\Pr(M_k)$ and $\pi(M_k)$ denote the posterior and prior probabilities, respectively. For a pair of models with x and y shifts, the Bayes factor evidence in favour of model x is given by

$$\mathrm{BF}_{x,y} = \frac{\Pr(M_x)\ \pi(M_y)}{\Pr(M_y)\pi(M_x)}. \qquad \text{eqn 7}$$

Because the Bayes factor is a ratio of marginal likelihoods, it is expected to be invariant with respect to the prior odds ratio of the models. Rabosky (2014) used posterior probabilities for model selection, but it is clear on theoretical grounds that Bayes factors are a more robust framework for inference. Bayes factors are a metric of support for a particular model relative to an alternative that takes into account the prior probability of each model (Jeffreys 1935; Kass & Raftery 1995; Huelsenbeck, Larget & Alfaro 2004; Rabosky 2014). Larger Bayes factors indicate greater support for the numerator model, with a Bayes factor > 20 frequently interpreted as strong support, although some workers find lower values acceptable (see Kass & Raftery 1995).

There are at least two practical issues that we must address to use Bayes factors in the BAMM framework. First, we can only compute Bayes factors for sets of models where both the posterior and prior probabilities are known (or estimated). The analytical prior formulation above allows us to compute prior odds ratios for any pair of models, but we may be unable to approximate the posterior probability for models that are rarely (or never) sampled in the posterior. Secondly, for a set of N models, we obtain an $N \times N$ matrix of pairwise Bayes factors, and it is not immediately obvious how to select an overall 'best' model (Rabosky *et al.* 2014, MEE). We selected models in a stepwise fashion using Bayes factors. Beginning with the least complex sampled model (e.g. M_0), and we used Bayes factors to determine whether or not the next most complex model (e.g. M_1) was better supported. If the Bayes factor evidence supported the more complex model, then the procedure is continued up to the next level of complexity (e.g. comparing M_2 to M_1). The most complex model supported was then chosen as the 'best' model.

To increase the stringency of this test, a Bayes factor threshold can be chosen such that more complex models are only selected if they have a minimum level of support (e.g. Bayes factor > 5). Increasing the level of evidence needed to accept a more complex model will decrease the probability of detecting too many shifts, but increase the probability of detecting too few (a trade-off between type I and type II error rates). Here, to rigorously test whether or not users could 'stack the deck' with their selection of a model prior in BAMM, we selected a more complex model if the corresponding Bayes factor evidence relative to the less-complex model was greater than 1·0.

EFFECTS OF MODEL PRIOR: CONSTANT-RATE TREES

We first tested the effects of the model prior on the inferred number of rate shifts when phylogenies are simulated in the absence of diversification rate variation. We simulated 100 constant-rate phylogenetic trees with 100 tips using the function sim.bd.taxa from the TreeSim package (Stadler 2011). Values for the speciation rate (λ) were drawn from an exponential distribution with a rate parameter of 1, and the values for extinction rates (μ) were drawn such that the relative extinction rate (μ/λ) was uniformly distributed on the interval [0, 0·9]. We analysed these trees with the expected number of shifts (γ) set to 0·1, 0·5, 1, 2, 10 and 20 using BAMM v 2.5.0). We ran each analysis for 3 000 000 generations and discarded the first 10% of samples as burn-in. We tabulated the posterior probabilities of all classes of models sampled during the BAMM run, and – for each model of order $i > 0$ – we computed the pairwise Bayes factor between M_i and M_0, or $\mathrm{BF}_{i,0}$. For

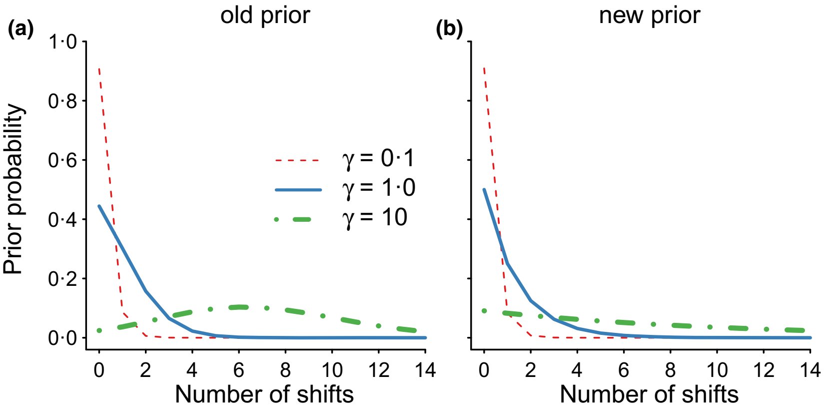

Fig. 1. Prior probability of k shifts as a function of the prior mean (γ) for the old (a) and new (b) model priors. Use of the model prior implemented in BAMM v2.3 and earlier results in greater prior probability of large shift numbers when γ is large. However, the difference between these implementations is relatively minor for the default parameterization of γ = 1 (Figs 2, 3 and S1).

comparison, we performed a parallel analysis using a outdated version of BAMM that included an error in the acceptance probability for MCMC moves that updated the Poisson rate parameter Λ, because this error was present in all released versions of BAMM < 2.4. We included this comparison since many published empirical studies have used BAMM v2.3 or lower. In addition to the model prior, BAMM also places priors on speciation and extinction rates. The present study was focused solely on assessing the impact of the model prior, and as such, these other priors were held constant at their default values (exponential distribution with mean of 1·0) across all simulations.

EFFECTS OF MODEL PRIOR: TREES WITH RATE SHIFTS

To assess the effects of the model prior on inference when diversification rate shifts are present, we re-analysed the original set of rate-variable phylogenies used in Rabosky's (2014) validation of BAMM's performance; this distribution of trees is available at Dryad (doi: 10.5061/dryad.hn1vn). The trees in this data set were simulated with one, two, three or four shifts in diversification rate regimes and range from 54 to 882 tips. Each 'shift regime' is a distinct linear diversity-dependent diversification process (speciation rate declines linearly with total clade richness; see Rabosky 2014 for more simulation details), and diversification rates thus vary among lineages and through time. A complete description of the simulation algorithm used to generate these phylogenies is found in Rabosky (2014). Due to computational resource availability, we analysed the first 300 trees for each number of simulated rate shifts (1, 2, 3 and 4) using priors on γ equal to 0·1, 1 and 100.

We note that, as in the original analysis of BAMM's performance (Rabosky 2014), the generating model is not identical to the inference model: phylogenies were simulated under a mixture of pure diversity-dependent processes, but speciation rates within BAMM rate regimes are restricted to a time-dependent exponential model. This functional relationship between speciation rate and time is expected to provide a good approximation to linear diversity-dependent dynamics (Quental & Marshall 2010; Rabosky 2014), but affords several computational advantages over formal diversity-dependent models (Etienne & Haegeman 2012).

Finally, we assessed the relationship between the model prior and the accuracy with which BAMM reconstructs both speciation rates and rate shift location. Speciation rate accuracy was measured as the ratio between the estimated and generating values of λ as per Rabosky (2014). To assess shift location, we created a pairwise cohort matrix (Rabosky *et al.* 2014) for each tree. A cohort matrix is, for a phylogeny of N taxa, an $N \times N$ matrix describing the pairwise probability that the ith and jth taxa are assigned to the same evolutionary rate regime. We graphically describe the use of cohort matrices for measuring shift accuracy in Appendix S1. For the 'true' cohort matrix, each value of the cohort matrix takes a value of 1 (if a given pair of taxa is in the same rate regime) and 0 (if the pair of taxa is in different rate regimes). We denote the true probability that two taxa are in the same regime with $C_{i,k}$. Each element $D_{i,k}$ of the 'observed' cohort matrix, derived from BAMM analysis, is computed as

$$D_{i,k} = \frac{1}{V}\sum_{z=1}^{V} I_{i,k,z} \qquad \text{eqn 8}$$

where V is the number of samples in the posterior and $I_{i,k,z}$ is an indicator variable taking a value of 1 if the ith and kth taxa from posterior sample z are assigned to the same rate regime and 0 otherwise. We used the average of the absolute value of the differences between the true cohort matrix and BAMM-reconstructed cohort matrix as an index of location accuracy, a quantity that we compute as

$$\frac{2}{N(N-1)}\sum_{k=2}^{N}\sum_{i=1}^{k-1} |C_{i,k} - D_{i,k}|. \qquad \text{eqn 9}$$

This statistic represents the average probability that BAMM has correctly determined the relationship between any two taxa ('same regime' or 'different regime') in the phylogeny. An overall value of 1·0 indicates that all pairs of species have been correctly assigned; this value can only be achieved if BAMM recovers the true locations of rate shifts with 100% accuracy in all samples from the posterior. Conversely, a value of 0 implies that all pairs of taxa are incorrectly assigned (e.g. species from different regimes are consistently assigned to the same evolutionary rate regime and species from the same regime are placed in different ones).

We computed this index for trees that were analysed with different model priors (γ = 0·1, 1 and 100). We then compared the accuracy of BAMM shift reconstructions to randomized shift placements. For a given BAMM analysis, a single such randomization involved sampling a shift configuration from the posterior and probabilistically assigning the observed number of shifts to branches based on the branch-specific prior probability of a shift; shifts were thus randomly and uniformly distributed across trees.

Results

For constant-rate (zero-shift) simulations, when we compared the Bayes factor evidence for model M_1 to model M_0, we found no effect of the model prior (Fig. 2a). However, there is a relatively modest effect of the model prior on the posterior probability of model M_0 which approaches an asymptote of approximately 0·5 for $\gamma > 5$, which also did not lead to the rejection of the constant-rate model (Fig. 2b). Model inference is thus not sensitive the prior across a broad range of expected shift numbers ($\gamma = 0{\cdot}1$ to $\gamma = 100$). We did not observe positive evidence (Bayes factors > 1) for one or more shifts in any of the 100 simulated constant-rate phylogenies, thus indicating a very low type I error rate for BAMM on constant-rate phylogenies. In contrast, model selection under the incorrect prior (BAMM v2.3 and earlier) is substantially influenced by the prior parameterization, regardless of whether model selection is performed using Bayes factors (Fig. 3a) or posterior probabilities (Fig. 3b). However, even with the incorrect prior, we found no evidence of bias towards (spurious) rate heterogeneity under BAMM's default prior ($\gamma = 1$; Fig. S1).

The incorrect model prior (BAMM v. 2.3 and earlier) is now dropped from further consideration; all results below reflect only the correct implementation of the model prior in BAMM (see Appendix S2 for a comparison of a previous data set analysed using both the old, incorrect model prior and the new; Figs S2 and S3).

For each simulation scenario (e.g. constant rate; 1-shift), we found the average posterior probability for each value of k across all simulated phylogenies under three prior parameterizations (Fig. 4). The best-fitting model was chosen using Bayes factor comparisons, and this best model was most often equal to the generating model (Fig. 5). The stepwise Bayes factor procedure selected models that were more complex than the true (generating) model in fewer than 2% of all trees with shifts (the highest rate was 3·1% in the 4-shift trees with $\gamma = 0{\cdot}1$; Fig. 5). As the number of shifts increased, the probability of selecting a less-complex model increased, suggesting that even under very liberal priors BAMM is conservative and more prone to low power than to the inference of spurious rate regimes. Critically, we find no evidence that use of flatter prior values (e.g. high number of expected shifts) can lead to biased inference with BAMM when Bayes factors are used for model selection.

In general, the very liberal prior ($\gamma = 100$) produced better convergence performance with higher effective sample sizes (Fig. 6). However, the model prior does not appear to impact the accuracy with which BAMM reconstructs shift locations (Fig. 7). For all three prior parameterizations, BAMM consistently identified the correct pairwise relationship between taxa ('same regime' or 'different regime') for the overwhelming majority of such comparisons. Likewise, the mean proportional error in speciation rate (Table 1) did not vary substantially with the model prior. The error associated with speciation rates is impacted by the number, location and magnitude of the shifts as well as the assumptions of the model, so low error in the rate estimate means that BAMM is performing well. To test consistency among runs, we compared the estimates of the tip rates (λ and μ) for the 4-shift trees between model priors of $\gamma = 0{\cdot}1$ and $\gamma = 100$ and found that separate runs produced highly correlated results for a clear majority of trees (Fig. 8). These results on accuracy and consistency between runs with different values of the model prior suggest that γ has little impact on estimates of tip rates. It is still possible that researchers could bias their estimate of tip rates by placing extremely strong priors on the number of shifts (e.g. $\gamma = 0{\cdot}00001$), but we find no evidence that rates are biased

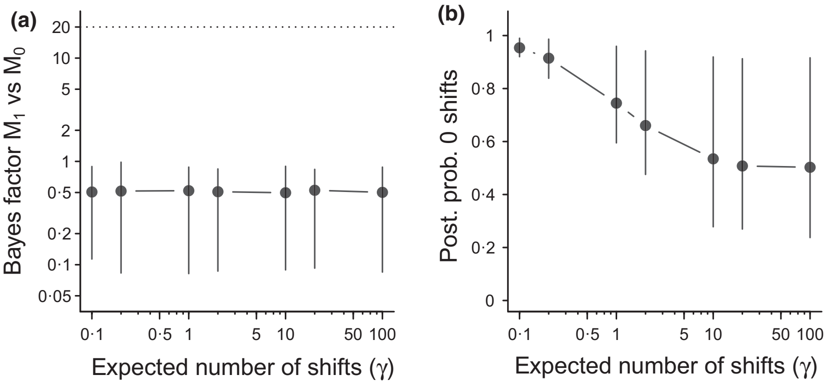

Fig. 2. Distribution (median and 5–95% quantiles) of model support values across constant-rate trees as a function of the model prior (expected number of shifts; γ) in recent versions (v 2.4+) of BAMM. (a) Bayes factor evidence favouring a model with rate variation (one-shift) relative to the true (zero-shift) model. Bayes factors greater than one indicate support for a model with rate variation; horizontal dashed line corresponds to strong or 'significant' Bayes factor support (BF = 20) in favour of rate variation. (b) Posterior probabilities of the zero-shift model as a function of γ. Bayes factor evidence for rate variation is not sensitive with respect to the prior (a), and even liberal prior distributions (e.g. $\gamma = 100$) yield no evidence for rate variation for constant-rate phylogenies. Posterior probabilities are influenced by the model prior (b) but did not achieve conventional thresholds ($P = 0{\cdot}05$) for rejecting the true (zero-shift) model, even with liberal model priors.

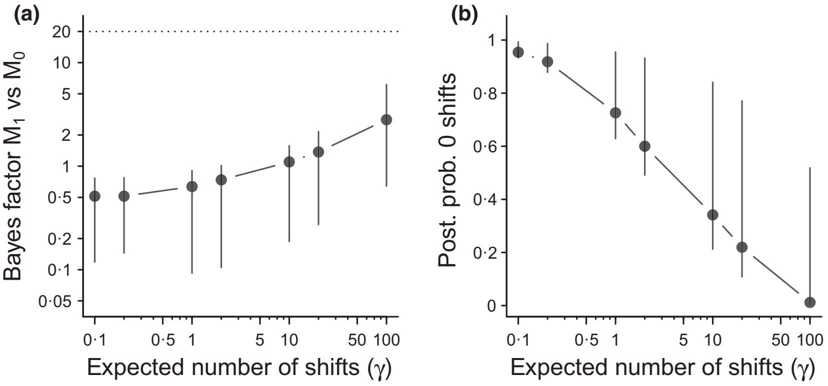

Fig. 3. Distribution (median and 5–95% quantiles) of model support values across constant-rate trees as a function of the model prior for old (v 2.3 and earlier) versions of BAMM, which contained an error in the Hastings ratio calculation for MCMC moves that updated the Poisson rate parameter Λ. (a) Bayes factor evidence favouring a model with rate variation (one-shift) relative to the true (zero-shift) model. (b) Posterior probabilities of the zero-shift model as a function of γ. The incorrect implementation magnified the effects of the prior on the posterior relative to the correct MCMC implementation. However, across the range of model priors ($\gamma = 0{\cdot}1$ to $\gamma = 100$), Bayes factors did not result in strong evidence (BF = 20; horizontal dashed line) for models with rate variation, despite increasing support for overly complex models with increasing γ. The default model prior in BAMM ($\gamma = 1$ for all versions) does not typically support models with too many shifts, even when posterior probabilities alone are used for model selection (b), although increasing γ did increase support for overly complex models in the old version of BAMM.

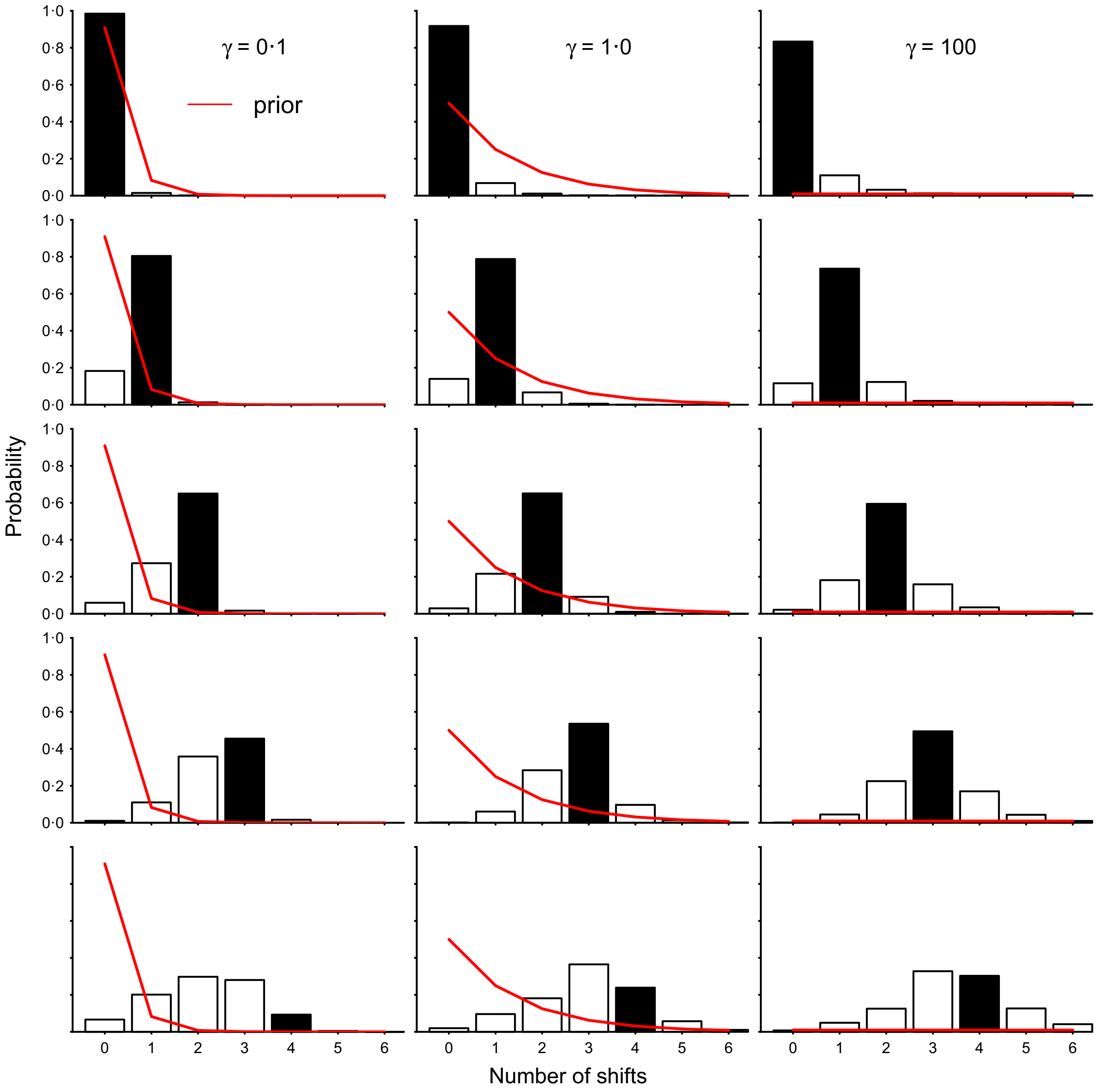

Fig. 4. Marginal posterior probability distributions on the number of shifts for phylogenies simulated with and without rate heterogeneity, under three prior parameterizations (γ = 0·1, 1 and 100). Histograms represent the mean of the corresponding marginal posterior distributions across all simulated phylogenies with a specified level of rate heterogeneity (rows). Top row consists of 100 constant-rate trees; rows 2–5 correspond to distributions of phylogenies with 1, 2, 3 and 4 shifts, respectively. Phylogenies with rate variation are taken from Rabosky (2014). The prior distribution on the number of shifts (red line) is illustrated in each panel; filled histogram bars correspond to the true number of shifts for each simulation scenario.

across prior parameterizations that varied by three orders of magnitude.

Discussion

We demonstrate that diversification model selection with BAMM is largely robust to choice of model prior (Fig. 2). BAMM successfully detected the correct number of shifts in the majority of simulated data sets and rarely selected overly complex models (Figs 2 and 3). However, as the number of shifts increased, the probability of selecting a less-complex model increased (Figs 4 and 5). This result implies that BAMM is slightly conservative, even under very liberal priors. Critically, using a flatter prior value (high number of expected shifts) did not ‘stack the deck’ in favour of selecting excessively complex models when using Bayes factors for model selection. When constant-rate phylogenies were analysed with BAMM, we found a striking invariance of Bayes factors to the model prior (Fig. 2).

BAMM underestimated the number of rate shifts in a substantial fraction of the simulated trees (low power to detect some shifts). However, the simulation algorithm

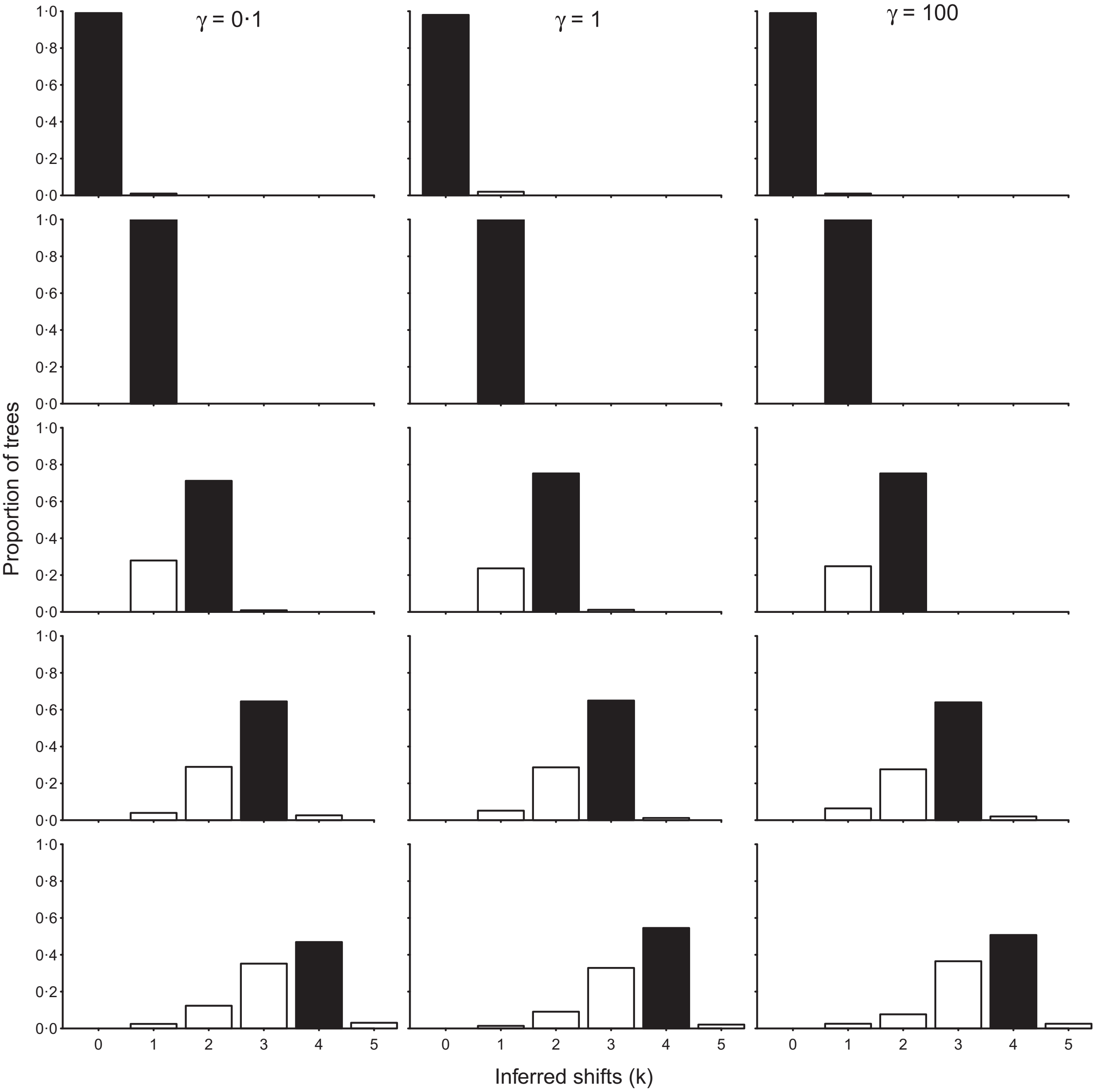

Fig. 5. Frequency distribution of the ‘best model’ across sets of phylogenies simulated under five diversification scenarios (rows), selected using the stepwise Bayes factor procedure described in the text. Each column represents analyses done with a different model prior (left column $\gamma = 0{\cdot}1$, middle column $\gamma = 1$, right column $\gamma = 100$); rows (top to bottom) denote sets of phylogenies with 0, 1, 2, 3 and 4 shifts, respectively. Black bars in each panel indicate the proportion of analyses where BAMM recovered the true number of rate shifts in the simulated data sets. The best-supported model contained an excessive number of shifts (e.g. type I error) in 2% of trees for each set of analyses. Panels only show trees that had reached convergence (effective sample size > 200) with a minimum of 150 trees in each panel (other than the constant-rate panels where all trees converged). Convergence problems arose for some analyses with $\gamma = 0{\cdot}1$ (see Fig. 6).

allowed multiple shifts to occur in close temporal and topological proximity on the tree. When rate shifts are very close temporally, our ability to estimate them should be greatly reduced, as there is less time for new lineages and branch length (e.g. data) to accumulate between the rate shifts. Furthermore, all simulated shifts entailed sampling parameters from an identical distribution, such that speciation and extinction rates themselves may not have varied substantially between some shifts. Similarity in rate parameters for adjacent shift regimes would have further reduced our ability to detect rate heterogeneity. Rabosky (2014) observed a similar reduction in statistical power with increasing numbers of shifts, but BAMM was nonetheless able to reliably infer branch-specific variation in the rate of speciation even when the number of shifts was underestimated.

The effects of the model prior on speciation and extinction rates appear to be limited, as the two most extreme model priors ($\gamma = 0{\cdot}1$ vs. $\gamma = 100$) produced highly correlated diversification rate estimates for trees with four rate shifts. These results pertain specifically to the model prior (i.e. the number

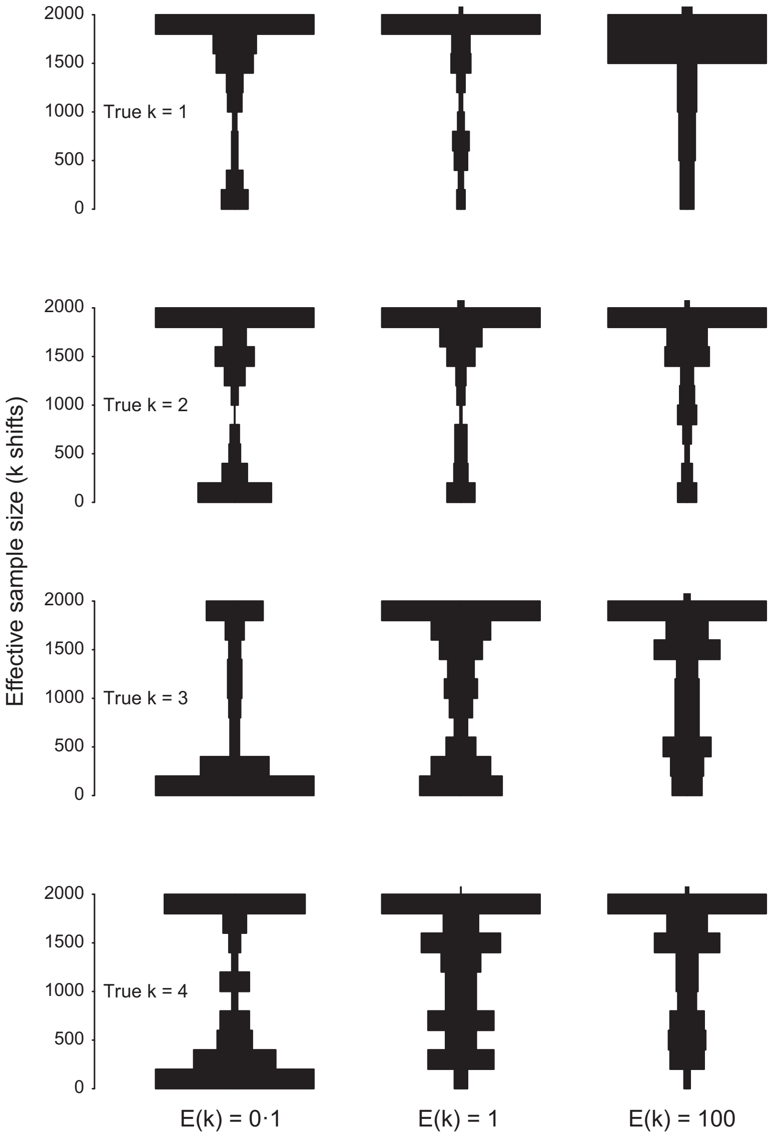

Fig. 6. Spindle plots illustrating effects of model prior on convergence properties of the Markov chain Monte Carlo simulation in BAMM. Each panel shows the distribution of effective sample sizes in the number of rate shifts for trees simulated with k = 1, 2, 3 or 4 rate shifts (described by Rabosky 2014 and in the text) and analysed with three different model prior values (γ = 0·1, 1 and 100). Width of bars is proportional to the number of trees with effective sample sizes that fall into each bin. A set of analyses with good convergence properties would appear 'top-heavy'; conversely, a set of analyses with poor convergence properties would appear 'bottom-heavy', reflecting a high proportion of analyses with low effective sample sizes. Analyses that specified larger values of γ (expected numbers of shifts) result in larger effective sample sizes (i.e. chains run with liberal priors were more likely to converge and converged more quickly), relative to analyses with small values of γ.

of expected shifts, γ), as we did not explore the impact of the rate parameter priors. The effects of the rate parameter priors on posterior estimates of speciation and extinction rates in BAMM remain largely unexplored (but see Callahan & McPeek 2016 for an empirical example).

Choice of model prior has a substantial effect on the efficiency of the MCMC simulation in BAMM. Restrictive prior distributions led to poor MCMC convergence properties in our analyses. We speculate that this result is attributable to the flattening of the posterior probability landscape that occurs with increasingly liberal priors in BAMM's compound Poisson process model of rate variation. A flatter model prior allows the MCMC algorithm to explore a larger amount of parameter space and converge more quickly by flattening the posterior probability surface with respect to the number of rate shifts.

The simple analytical form of the prior (geometric; eqn 6) allows us to calculate the prior probability of any number of shifts precisely. This is a substantial advance relative to earlier versions of BAMM, which relied on explicit simulation of the prior distribution on the number of rate shifts. For very large trees, it may be the case that samples from the posterior never include the no-shift model (e.g. 6000+ tip trees for birds and fish; Rabosky *et al.* 2013; Rabosky & Huang 2016), leading to difficulties in computing Bayes factors where the posterior probability of one model is poorly estimated. The model prior in BAMM decreases monotonically from zero shifts, which means that model M_0 (zero shifts) always has the highest prior probability regardless of γ. This simple property of the prior distribution implies that failure to sample model M_0 in the posterior is evidence for rate heterogeneity when γ is low. However, if M_0 is unsampled, it is difficult to estimate the corresponding posterior probability of the model with any degree of accuracy, and estimates of model posterior probabilities are essential for computing Bayes factors.

Our stepwise procedure for selecting the best-fitted model using Bayes factors always selected the better supported model, even when the difference in support was small relative to the increase in complexity (i.e. a threshold of 1·0). Although we observed good statistical

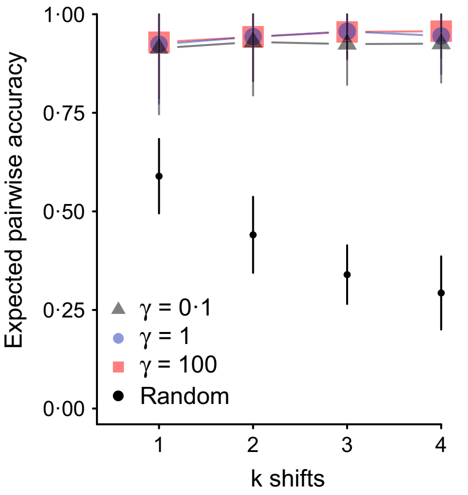

Fig. 7. Shift location accuracy is independent of model prior. For each tree, we computed the mean cohort assignment accuracy, a measure of the extent to which BAMM correctly assigns taxa to the same (or different) rate regime. On average, cohort accuracy under each prior exceeded 0·95. Values of 1·0 can only be obtained when BAMM correctly infers the correct location of all rate shifts for each sample from the posterior. For comparison, the distribution of mean cohort assignment accuracies is shown after randomizing shift locations across the focal phylogenies.

Table 1. Proportionality ratios for speciation rates (estimated vs. true) under each of the differing priors. For each tree, we used the mean value of the estimated λ for each branch divided by the true λ value for that branch. A value of 1 indicates that across all of the trees, the average estimated value of λ was identical to the value used to generate the trees. These values are consistent with the results shown in Rabosky (2014) using these same trees.

Model	$\gamma = 0{\cdot}1$	$\gamma = 1$	$\gamma = 100$
$k = 1$	1·02	0·99	0·99
$k = 2$	0·94	0·88	0·88
$k = 3$	0·88	0·85	0·84
$k = 4$	0·86	0·81	0·81

performance for constant- and variable-rate trees, researchers may want to choose a higher Bayes factor threshold to be more conservative in some cases. Also, this procedure assumes that unsampled models are so poorly supported that they can be ignored. It is possible to relax this assumption by approximating the 'maximum' posterior probability of the unsampled model: $P_{\mathrm{MAX}} = 1/(Z + 1)$, where Z is the number of samples in the posterior simulated with MCMC. In general, we expect that the true posterior probability of a model that is never sampled is less than this value, making this a reasonably conservative and quick approximation.

Alternatively, as a more rigorous but computationally demanding method, Huelsenbeck, Larget & Alfaro (2004) provide a framework for approximating the posterior probability of an unsampled model. In their approach, a second posterior distribution is generated using a seeded prior, in which the unsampled model of interest is very strongly favoured. Huelsenbeck, Larget & Alfaro (2004) provide an equation, reproduced below, for calculating the posterior probability of the unsampled model given the prior and posterior from the seeded and unseeded priors:

$$\mathrm{Pr}'(M|X) = \frac{\mathrm{Pr}(M_i|X)\frac{\mathrm{Pr}'(M_i)}{\mathrm{Pr}(M_i)}}{\sum_j \mathrm{Pr}(M_j|X)\frac{\mathrm{Pr}'(M_j)}{\mathrm{Pr}(M_j)}} \qquad \text{eqn 10}$$

where X is the observed data, $\mathrm{Pr}'(M)$ is the posterior distribution of a model under the unseeded prior, and $\mathrm{Pr}(M)$ is the posterior of a model under the seeded prior. This allows a researcher interested in examining the probability of a zero-shift model to simulate a posterior distribution of shift configurations under a seeded prior that will maximize the probability that a no-shift model is sampled in the posterior. Using the prior and posterior from the seeded run, it is possible to use eqn 10 to compute the posterior of the unsampled model for the unseeded run. The downside to this approach is that, for

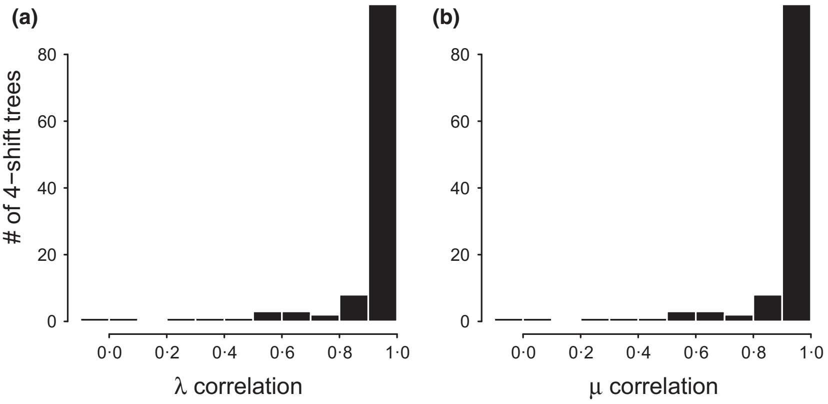

Fig. 8. Estimates of speciation and extinction rates are highly correlated across different model priors. Each phylogeny from the 4-shift data set was analysed with BAMM under model priors of $\gamma = 0{\cdot}1$ and $\gamma = 100$. For each tree, we computed the Pearson correlation between tip-specific estimates of the rate of speciation (a) or extinction (b) under the two priors; highly correlated estimates indicate that BAMM runs with these very different model priors resulted in concordant estimates of evolutionary rates. A small number of trees showed low correlations (~0) in tip rate estimates; these analyses generally involved runs where most of the posterior shift distribution for the conservative model prior ($\gamma = 0{\cdot}1$) was centred on 0 or 1 shift.

very large empirical trees with high levels of rate variation, an extraordinarily restrictive prior must be used, and – as we have shown – use of more restrictive priors can decrease MCMC efficiency and lead to convergence problems.

In summary, we have demonstrated that inference of the number of diversification rate shifts on a phylogeny in BAMM is robust to the choice of model prior when Bayes factors are used as a criterion for model selection. Regardless of model prior, BAMM analyses rarely found support for overly complex models. Rates of speciation and extinction at the tips of the phylogeny appear to be relatively insensitive to the model prior, although further research is needed on the sensitivity of BAMM analyses to variation in speciation and extinction rate priors. Even when a model of interest (such as the zero-shift model) is not sampled in the posterior, there are several possible methods for computing a meaningful Bayes factor to assess support. Because model selection using Bayes factors is robust to the choice of model prior, and because MCMC efficiency appears to be positively correlated with the mean of the prior distribution on the number of shifts, we recommend the use of a liberal model prior in studies using BAMM.

Acknowledgements

We thank Cécile Ané and Bret Larget for checking (May 2015) our analytical solution to the model prior as implemented in BAMM. We thank P. Title for discussions and assistance running these analyses. This research was supported in part by NSF-DEB-1256330 and by the David and Lucile Packard Foundation.

Data accessibility

All input and output files, as well as R scripts used to analyse these data, are available on Dryad Digital Repository at http://dx.doi.org/10.5061/dryad.0564q (Mitchell & Rabosky 2016).

References

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. *Proceedings of the National Academy of Sciences*, **106**, 13410–13414.

Beaulieu, J.M. & O'Meara, B.C. (2016) Detecting hidden diversification shifts in models of trait-dependent speciation and extinction. *Systematic Biology*, **65**, 583–601.

Callahan, M.S. & McPeek, M.A. (2016) Multi-locus phylogeny and divergence time estimates of *Enallagma damselflies* (Odonata: Coenagrionidae). *Molecular Phylogenetics and Evolution*, **94**, 182–195.

Coyne, J.A. & Orr, H.A. (2004) *Speciation*. Sinauer Associates, Sunderland, MA, USA, 545 pp.

Etienne, R.S. & Haegeman, B. (2012) A conceptual and statistical framework for adaptive radiations with a key role for diversity dependence. *The American Naturalist*, **180**, E75–E89.

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

Gould, S.J., Raup, D.M., Sepkoski, J.J., Schopf, T.J.M. & Simberloff, D.S. (1977) The shape of evolution: a comparison of real and random clades. *Paleobiology*, **3**, 22–40.

Greenwood, M. & Yule, G.U. (1920) 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*, **83**, 255–279.

Grimmett, G.R. & Stirzaker, D.R. (2001) *Probability and Random Processes*, third edn. Oxford University Press Inc., New York, NY, USA.

Heaney, L.R. (1986) Biogeography of mammals in SE Asia: estimates of rates of colonization, extinction and speciation. *Biological Journal of the Linnean Society*, **28**, 127–165.

Huelsenbeck, J.P., Larget, B. & Alfaro, M.E. (2004) Bayesian phylogenetic model selection using reversible jump Markov Chain Monte Carlo. *Molecular Biology and Evolution*, **21**, 1123–1133.

Jablonski, D. (2008) Species selection: theory and data. *Annual Reviews in Ecology, Evolution and Systematics*, **39**, 501–524.

Jeffreys, H. (1935) Some tests of significance, treated by the theory of probability. *Proceedings of the Cambridge Philosophy Society*, **31**, 203–222.

Jetz, W., Thomas, G.H., Joy, J.B., Hartmann, K. & Mooers, A.O. (2012) The global diversity of birds in space and time. *Nature*, **491**, 444–448.

Kass, R.E. & Raftery, A.E. (1995) Bayes factors. *Journal of the American Statistical Association*, **90**, 773–795.

Liem, K.F. (1973) Evolutionary strategies and morphological innovations: cichlid pharyngeal jaws. *Systematic Zoology*, **22**, 425–441.

Maddison, W.P., Midford, P.E. & Otto, S.P. (2007) Estimating a binary character's effect on speciation and extinction. *Systematic Zoology*, **56**, 701–710.

Mitchell, J.S. & Rabosky, D.L. (2016) Data from: Bayesian model selection with Bayesian Analysis of Macroevolutionary Mixtures: effects of the model prior on the inferred number of diversification shifts. *Dryad Digital Repository*, http://dx.doi.org/10.5061/dryad.0564q

Morlon, H., Parsons, T.L. & Plotkin, J.B. (2011) Reconciling molecular phylogenies with the fossil record. *Proceedings of the National Academy of Sciences*, **108**, 16327–16332.

Phillimore, A.B. & Price, T.D. (2008) Density-dependent cladogenesis in birds. *PLOS Biology*, **6**, e71.

Quental, T.B. & Marshall, C.R. (2010) Diversity dynamics: molecular phylogenies need the fossil record. *Trends in Ecology & Evolution*, **25**, 434–441.

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

Rabosky, D.L. & Huang, H. (2015) Minimal effects of latitude on present-day speciation rates in New World birds. *Proceedings of the Royal Society B*, **282**, 20142889.

Rabosky, D.L. & Huang, H. (2016) A robust semi-parametric test for detecting trait-dependent diversification. *Systematic Biology*, **65**, 181–193.

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**, 1958. http://dx.doi.org/10.1038/ncomms2958

Rabosky, D.L., Donnellan, S.C., Grundler, M. & Lovette, I.J. (2014) Analysis and visualization of complex macroevolutionary dynamics: an example from Australian scincid lizards. *Systematic Biology*, **6**, 610–627.

Raup, D.M. (1985) Mathematical models of cladogenesis. *Paleobiology*, **11**, 42–52.

Simpson, G.G. (1953) *The Major Features of Evolution*. Columbia University Press, New York, NY, USA.

Sloss, L.L. (1950) Rates of evolution. *Journal of Paleontology*, **24**, 131–139.

Slowinski, J.B. & Guyer, C. (1989) Testing the stochasticity of patterns of organismal diversity: an improved null model. *The American Naturalist*, **134**, 907–921.

Stadler, T. (2011) Simulating trees with a fixed number of extant species. *Systematic Biology*, **60**, 676–684.

Received 10 April 2016; accepted 29 July 2016
Handling Editor: David Orme

Supporting Information

Additional Supporting Information may be found online in the supporting information tab for this article:

Appendix S1. Cohort Cohesion and Shift Accuracy in BAMM.

Appendix S2. Impact of BAMM changes and improvements on empirical inference.

Appendix S3. Relationship of tree size to model support.