

1

2 Received Date : 30-Dec-2015

3 Revised Date : 13-Jul-2016

4 Accepted Date : 29-Jul-2016

5 Article type : Research Article

6

7

8 Editor: David Orme

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

11

12 Jonathan S. Mitchell *

13 Daniel L. Rabosky

14

15 University of Michigan, Museum of Zoology & Department of Ecology and
16 Evolutionary Biology, Ann Arbor, MI 48109

17

18 *Corresponding author information:

19 Email: jonsmitc@umich.edu

20 Phone: 919 273 8800

21

22 RH: Inference of rate variation insensitive to prior

23

24 Word count: 7044

25 Abstract

26

27 1. Understanding variation in rates of speciation and extinction -- both among
28 lineages and through time -- is critical to the testing of many hypotheses about
29 macroevolutionary processes. BAMM is a flexible Bayesian framework for inferring
**This is the author manuscript accepted for publication and has undergone full peer review
but has not been through the copyediting, typesetting, pagination and proofreading process,
which may lead to differences between this version and the [Version of Record](#). Please cite
this article as [doi: 10.1111/2041-210X.12626](https://doi.org/10.1111/2041-210X.12626)**

30 the number and location of shifts in macroevolutionary rate across phylogenetic
31 trees and has been widely used in empirical studies. BAMM requires that
32 researchers specify a prior probability distribution on the number of diversification
33 rate shifts before conducting an analysis. The consequences of this "model prior" for
34 inference are poorly known but could potentially influence both the probability of
35 accepting models that are more (high error rate) or less (low power) complex than
36 the generating model.

37

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

44

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

54

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

57

58 Keywords: BAMM, macroevolution, rate variation, birth-death model, Bayesian

59 Introduction

60

61 Species richness is unequally partitioned across the tree of life, with some clades
62 having far more species than their corresponding sister lineages. Understanding the
63 root causes of this variation has long been a foundational research paradigm in
64 macroevolution (Sloss 1950; Raup 1985; Jetz et al 2012; Rabosky 2014). It is
65 increasingly clear that much of the variation in species richness among clades
66 involves differential rates of speciation and extinction (Jablonski 2008; Alfaro et al
67 2009). Hypotheses to explain patterns of species richness range from the
68 geographical complexity of regions in which different clades occur (e.g., Heaney
69 1986) to key innovations (e.g., Simpson 1953; Liem 1973; Coyne & Orr 2004;
70 Jablonski 2008). However, the stochastic nature of the diversification process can
71 lead to variation in species richness that is not associated with causal differences in
72 macroevolutionary rates (Gould et al. 1977). Hence, robust tests of
73 macroevolutionary hypotheses require methods that can identify differential rates
74 of speciation and extinction across the tree of life (Slowinski and Guyer 1989;
75 Phillimore and Price, 2008; Rabosky 2014).

76
77 A number of methods have recently been developed that allow researchers to model
78 heterogeneous rates of speciation and extinction across the branches of
79 phylogenetic trees (Maddison et al. 2007; FitzJohn et al 2009; Alfaro et al 2009;
80 Morlon et al. 2011; Etienne and Haegemann 2012; Beaulieu and O'Meara 2015).
81 Bayesian Analysis of Macroevolutionary Mixtures (BAMM; Rabosky 2014, Rabosky
82 et al. 2014) is a method for automatically identifying heterogeneous mixtures of
83 evolutionary rate regimes across time-calibrated phylogenetic trees of extant taxa
84 that has been widely applied to diverse empirical datasets. BAMM uses reversible-
85 jump Markov-Chain Monte Carlo to approximate posterior distributions of
86 diversification models, enabling researchers to reconstruct the number, magnitude
87 and locations of rate shifts on phylogenetic trees. Shifts in evolutionary rates can
88 occur along any branch of the phylogenetic tree, and the rates can vary through time
89 within a rate regime (Rabosky 2014). This framework enables researchers to
90 evaluate whether clades vary in their speciation or extinction rate without
91 specifying particular clades to test *a priori*, and can be used to assess the

92 relationship between character states and lineage diversification rates (Rabosky
93 and Huang, 2015).

94

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

107

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

118

119 Methods

120

121 *Prior probability of k shifts in BAMM*

122

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

129

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

141

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

145

$$146 \quad \Pr(k|\theta) = \int_0^{\infty} \Pr(k|\Lambda) \Pr(\Lambda|\theta) d\Lambda \quad (\text{eqn 1})$$

147

148 This can be expanded to

149

$$150 \quad \Pr(k|\theta) = \int_0^{\infty} \frac{\Lambda^k e^{-\Lambda}}{k!} \theta e^{-\theta\Lambda} d\Lambda \quad \text{eqn 2}$$

151

152 and reduced to

$$153 \quad = \frac{\theta}{k!} \int_0^{\infty} \Lambda^k e^{-\Lambda(1+\theta)} d\Lambda \quad (\text{eqn 3}).$$

154

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

157

$$158 \quad \int_0^{\infty} \Lambda^x e^{-a\Lambda} d\Lambda \rightarrow \frac{\Gamma(x+1)}{a^{x+1}} \quad (\text{eqn 4})$$

159

160 and thus the full expression becomes

161

$$162 \quad \Pr(k|\theta) = \frac{\Gamma(k+1)}{k!} \frac{\theta}{(1+\theta)^{k+1}} = \frac{\theta}{(\theta+1)^{k+1}} \quad (\text{eqn 5}).$$

163

164 Letting $\gamma = 1/\theta$ we have:

165

$$166 \quad \Pr(k|\gamma) = \frac{\frac{1}{\gamma}}{\left(\frac{1}{\gamma}+1\right)^{k+1}} = \frac{\frac{1}{\gamma}}{\left(\frac{1}{\gamma}+1\right)\left(\frac{1}{\gamma}+1\right)^k}$$

$$167 \quad = \left(\frac{1}{\gamma+1}\right)\left(\frac{\gamma}{\gamma+1}\right)^k \quad (\text{eqn 6})$$

168 which is simply a geometric distribution with parameter $p = 1 / (\gamma + 1)$. This well-
 169 known mathematical result facilitates more rapid and comprehensive evaluation of
 170 the prior. The mean of the distribution is $(1-p)/p$, meaning that the expected

171 number of shifts under a particular exponential hyperprior is γ . We now explicitly
172 reference the model prior in terms of the expected number of shifts, γ .

173

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

188

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

201

202 *Model selection with Bayes factors*

203

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

213

$$214 \quad BF_{x,y} = \frac{\Pr(M_x) \pi(M_y)}{\Pr(M_y) \pi(M_x)} \quad (\text{eqn 7}).$$

215

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

226

227 There are at least two practical issues that we must address to use Bayes factors in
228 the BMM framework. First, we can only compute Bayes factors for sets of models
229 where both the posterior and prior probabilities are known (or estimated). The

230 analytical prior formulation above allows us to compute prior odds ratios for any
231 pair of models, but we may be unable to approximate the posterior probability for
232 models that are rarely (or never) sampled in the posterior. Second, for a set of N
233 models, we obtain an $N \times N$ matrix of pairwise Bayes factors, and it is not
234 immediately obvious how to select an overall "best" model (Rabosky et al. 2014,
235 MEE). We selected models in a step-wise fashion using Bayes factors. Beginning
236 with the least complex sampled model (e.g., M_0), and we used Bayes factors to
237 determine whether or not the next most complex model (e.g., M_1) was better
238 supported. If the Bayes factor evidence supported the more complex model, then the
239 procedure is continued up to the next level of complexity (e.g., comparing M_2 to M_1).
240 The most complex model supported was then chosen as the "best" model.

241

242 To increase the stringency of this test, a Bayes factor threshold can be chosen such
243 that more complex models are only selected if they have a minimum level of support
244 (e.g., Bayes factor > 5). Increasing the level of evidence needed to accept a more
245 complex model will decrease the probability of detecting too many shifts, but
246 increase the probability of detecting too few (a trade-off between Type I and Type II
247 error rates). Here, to rigorously test whether or not users could "stack the deck"
248 with their selection of a model prior in BAMM, we selected a more complex model if
249 the corresponding Bayes factor evidence relative to the less-complex model was
250 greater than 1.0.

251

252 *Effects of model prior: constant-rate trees*

253

254 We first tested the effects of the model prior on the inferred number of rate shifts
255 when phylogenies are simulated in the absence of diversification rate variation. We
256 simulated 100 constant-rate phylogenetic trees with 100 tips using the function
257 `sim.bd.taxa` from the TreeSim package (Stadler 2011). Values for the speciation rate
258 (λ) were drawn from an exponential distribution with a rate parameter of 1, and the
259 values for extinction rates (μ) were drawn such that the relative extinction rate ($\mu /$
260 λ) was uniformly distributed on the interval $[0, 0.9]$. We analyzed these trees with

261 the expected number of shifts (γ) set to 0.1, 0.5, 1, 2, 10, and 20 using BAMM v
262 2.5.0). We ran each analysis for 3,000,000 generations and discarded the first 10%
263 of samples as burn-in. We tabulated the posterior probabilities of all classes of
264 models sampled during the BAMM run, and -- for each model of order $i > 0$ -- we
265 computed the pairwise Bayes factor between M_i and M_0 , or $BF_{i,0}$. For comparison,
266 we performed a parallel analysis using an outdated version of BAMM that included an
267 error in the acceptance probability for MCMC moves that updated the Poisson rate
268 parameter Λ , because this error was present in all released versions of BAMM < 2.4 .
269 We included this comparison since many published empirical studies have used
270 BAMM v2.3 or lower. In addition to the model prior, BAMM also places priors on
271 speciation and extinction rates. The present study was focused solely on assessing
272 the impact of the model prior, and as such these other priors were held constant at
273 their default values (exponential distribution with mean of 1.0) across all
274 simulations.

275

276 *Effects of model prior: trees with rate shifts*

277

278 To assess the effects of the model prior on inference when diversification rate shifts
279 are present, we re-analyzed the original set of rate-variable phylogenies used in
280 Rabosky's (2014) validation of BAMM's performance; this distribution of trees is
281 available at Dryad (doi: 10.5061/dryad.hn1vn). The trees in this dataset were
282 simulated with one, two, three, or four shifts in diversification rate regimes and
283 range from 54 to 882 tips. Each "shift regime" is a distinct linear diversity-
284 dependent diversification process (speciation rate declines linearly with total clade
285 richness; see Rabosky 2014 for more simulation details), and diversification rates
286 thus vary among lineages and through time. A complete description of the
287 simulation algorithm used to generate these phylogenies is found in Rabosky
288 (2014). Due to computational resource availability, we analyzed the first 300 trees
289 for each number of simulated rate shifts (1, 2, 3 and 4) using priors on γ equal to 0.1,
290 1 and 100.

291

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

300
 301 Finally, we assessed the relationship between the model prior and the accuracy with
 302 which BAMM reconstructs both speciation rates and rate shift location. Speciation
 303 rate accuracy was measured as the ratio between the estimated and generating
 304 values of λ as per Rabosky (2014). To assess shift location, we created a pairwise
 305 cohort matrix (Rabosky et al 2014) for each tree. A cohort matrix is, for a phylogeny
 306 of N taxa, an $N \times N$ matrix describing the pairwise probability that the i 'th and j 'th
 307 taxa are assigned to the same evolutionary rate regime. We graphically describe the
 308 use of cohort matrices for measuring shift accuracy in Appendix I. For the "true"
 309 cohort matrix, each value of the cohort matrix takes a value of 1 (if a given pair of
 310 taxa are in the same rate regime) and 0 (if the pair of taxa are in different rate
 311 regimes). We denote the true probability that two taxa are in the same regime with
 312 $C_{i,k}$. Each element $D_{i,k}$ of the "observed" cohort matrix, derived from BAMM analysis,
 313 is computed as

$$314 \quad D_{i,k} = \frac{1}{V} \sum_{z=1}^V I_{i,k,z} \quad (\text{eqn 8})$$

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

$$320 \quad \frac{2}{N(N-1)} \sum_{k=2}^N \sum_{i=1}^{k-1} |C_{i,k} - D_{i,k}| \quad (\text{eqn 9})$$

321 This statistic represents the average probability that BAMM has correctly
 322 determined the relationship between any two taxa ("same regime" or "different
 323 regime") in the phylogeny. An overall value of 1.0 indicates that all pairs of species
 324 have been correctly assigned; this value can only be achieved if BAMM recovers the
 325 true locations of rate shifts with 100% accuracy in all samples from the posterior.
 326 Conversely, a value of 0 implies that all pairs of taxa are incorrectly assigned (e.g.,
 327 species from different regimes are consistently assigned to the same evolutionary
 328 rate regime and species from the same regime are placed in different ones).

329

330 We computed this index for trees that were analyzed with different model priors (γ
 331 = 0.1, 1, and 100). We then compared the accuracy of BAMM shift reconstructions to
 332 randomized shift placements. For a given BAMM analysis, a single such
 333 randomization involved sampling a shift configuration from the posterior and
 334 probabilistically assigning the observed number of shifts to branches based on the
 335 branch-specific prior probability of a shift; shifts were thus randomly and uniformly
 336 distributed across trees.

337

338 Results

339

340 For constant-rate (zero-shift) simulations, when we compared the Bayes factor
 341 evidence for model M_1 to model M_0 , we found no effect of the model prior (Fig. 2A).
 342 However, there is a relatively modest effect of the model prior on the posterior
 343 probability of model M_0 which approaches an asymptote of approximately 0.5 for γ
 344 > 5, which also did not lead to the rejection of the constant-rate model (Fig. 2B).
 345 Model inference is thus not sensitive the prior across a broad range of expected shift
 346 numbers ($\gamma = 0.1$ to $\gamma = 100$). We did not observe positive evidence (Bayes factors >
 347 1) for one or more shifts in any of the 100 simulated constant-rate phylogenies, thus
 348 indicating a very low Type I error rate for BAMM on constant-rate phylogenies. In

349 contrast, model selection under the incorrect prior (BAMM v2.3 and earlier) is
350 substantially influenced by the prior parameterization, regardless of whether model
351 selection is performed using Bayes factors (Fig. 3A) or posterior probabilities (Fig.
352 3B). However, even with the incorrect prior, we found no evidence of bias towards
353 (spurious) rate heterogeneity under BAMM's default prior ($\gamma = 1$; Fig S1).
354 The incorrect model prior (BAMM v. 2.3 and earlier) is now dropped from further
355 consideration; all results below reflect only the correct implementation of the model
356 prior in BAMM (see Appendix II for a comparison of a previous dataset analyzed
357 using both the old, incorrect model prior and the new; Figures S2 and S3).

358
359 For each simulation scenario (e.g., constant-rate; 1-shift), we found the average
360 posterior probability for each value of k across all simulated phylogenies under
361 three prior parameterizations (Fig. 4). The best-fitting model was chosen using
362 Bayes factor comparisons, and this best model was most often equal to the
363 generating model (Fig. 5). The stepwise Bayes factor procedure selected models that
364 were more complex than the true (generating) model in fewer than 2% of all trees
365 with shifts (the highest rate was 3.1% in the 4-shift trees with $\gamma = 0.1$; Fig. 5). As the
366 number of shifts increased, the probability of selecting a less-complex model
367 increased, suggesting that even under very liberal priors BAMM is conservative and
368 more prone to low power than to the inference of spurious rate regimes. Critically,
369 we find no evidence that use of flatter prior values (e.g., high number of expected
370 shifts) can lead to biased inference with BAMM when Bayes factors are used for
371 model selection.

372
373 In general, the very liberal prior ($\gamma = 100$) produced better convergence
374 performance with higher effective sample sizes (Fig. 6). However, the model prior
375 does not appear to impact the accuracy with which BAMM reconstructs shift
376 locations (Fig. 7). For all three prior parameterizations, BAMM consistently
377 identified the correct pairwise relationship between taxa ("same regime" or
378 "different regime") for the overwhelming majority of such comparisons. Likewise,

379 the mean proportional error in speciation rate (Table 1) did not vary substantially
380 with the model prior. The error associated with speciation rates is impacted by the
381 number, location and magnitude of the shifts as well as the assumptions of the
382 model, so low error in the rate estimate means that BAMM is performing well. To
383 test consistency among runs, we compared the estimates of the tip rates (λ and μ)
384 for the 4-shift trees between model priors of $\gamma = 0.1$ and $\gamma = 100$, and found that
385 separate runs produced highly correlated results for a clear majority of trees (Fig.
386 8). These results on accuracy and consistency between runs with different values of
387 the model prior suggest that γ has little impact on estimates of tip rates. It is still
388 possible that researchers could bias their estimate of tip rates by placing extremely
389 strong priors on the number of shifts (e.g., $\gamma = 0.00001$), but we find no evidence
390 that rates are biased across prior parameterizations that varied by three orders of
391 magnitude.

392

393 Discussion

394

395 We demonstrate that diversification model selection with BAMM is largely robust to
396 choice of model prior (Fig. 2). BAMM successfully detected the correct number of
397 shifts in the majority of simulated datasets and rarely selected overly complex
398 models (Fig. 2 - 3). However, as the number of shifts increased, the probability of
399 selecting a less-complex model increased (Fig. 4-5). This result implies that BAMM is
400 slightly conservative, even under very liberal priors. Critically, using a flatter prior
401 value (high number of expected shifts) did not “stack the deck” in favor of selecting
402 excessively complex models when using Bayes factors for model selection. When
403 constant-rate phylogenies were analyzed with BAMM, we found a striking
404 invariance of Bayes factors to the model prior (Fig. 2).

405

406 BAMM underestimated the number of rate shifts in a substantial fraction of the
407 simulated trees (low power to detect some shifts). However, the simulation
408 algorithm allowed multiple shifts to occur in close temporal and topological

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

419

420 The effects of the model prior on speciation and extinction rates appear to be
421 limited, as the two most extreme model priors ($\gamma = 0.1$ vs $\gamma = 100$) produced highly
422 correlated diversification rate estimates for trees with four rate shifts. These results
423 pertain specifically to the model prior (i.e., the number of expected shifts, γ), as we
424 did not explore the impact of the rate parameter priors. The effects of the rate
425 parameter priors on posterior estimates of speciation and extinction rates in BAMM
426 remain largely unexplored (but see Callahan and McPeck 2016 for an empirical
427 example).

428

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

437

438 The simple analytical form of the prior (geometric; eqn 6) allows us to calculate the
439 prior probability of any number of shifts precisely. This is a substantial advance

440 relative to earlier versions of BAMM, which relied on explicit simulation of the prior
441 distribution on the number of rate shifts. For very large trees it may be the case that
442 samples from the posterior never include the no-shift model (e.g., 6000+ tip trees
443 for birds and fish; Rabosky et al. 2013; Rabosky et al. 2015), leading to difficulties in
444 computing Bayes factors where the posterior probability of one model is poorly
445 estimated. The model prior in BAMM decreases monotonically from zero shifts,
446 which means that model M_0 (zero shifts) always has the highest prior probability
447 regardless of γ . This simple property of the prior distribution implies that failure to
448 sample model M_0 in the posterior is evidence for rate heterogeneity when γ is low.
449 However, if M_0 is unsampled, it is difficult to estimate the corresponding posterior
450 probability of the model with any degree of accuracy, and estimates of model
451 posterior probabilities are essential for computing Bayes factors.

452
453 Our step-wise procedure for selecting the best-fit model using Bayes factors always
454 selected the better-supported model, even when the difference in support was small
455 relative to the increase in complexity (i.e., a threshold of 1.0). Although we observed
456 good statistical performance for constant- and variable-rate trees, researchers may
457 want to choose a higher Bayes factor threshold to be more conservative in some
458 cases. Also, this procedure assumes that unsampled models are so poorly supported
459 that they can be ignored. It is possible to relax this assumption by approximating the
460 “maximum” posterior probability of the unsampled model: $P_{MAX} = 1 / (Z + 1)$, where
461 Z is the number of samples in the posterior simulated with MCMC. In general, we
462 expect that the true posterior probability of a model that is never sampled is less
463 than this value, making this a reasonably conservative and quick approximation.

464
465 Alternatively, as a more rigorous but computationally-demanding method,
466 Huelsenbeck et al. (2004) provide a framework for approximating the posterior
467 probability of an unsampled model. In their approach, a second posterior
468 distribution is generated using a seeded prior, in which the unsampled model of
469 interest is very strongly favored. Huelsenbeck et al. (2004) provide an equation,

470 reproduced below, for calculating the posterior probability of the unsampled model
 471 given the prior and posterior from the seeded and unseeded priors:

472

$$473 \quad Pr^{\prime}(M|X) = \frac{\Pr(M_i|X) \frac{Pr^{\prime}(M_i)}{\Pr(M_i)}}{\sum_j \Pr(M_j|X) \frac{Pr^{\prime}(M_j)}{\Pr(M_j)}}$$

474 (eqn 10)

475

476 where X is the observed data, $Pr^{\prime}(M)$ is the posterior distribution of a model under
 477 the unseeded prior and $\Pr(M)$ is the posterior of a model under the seeded prior.
 478 This allows a researcher interested in examining the probability of a zero-shift
 479 model to simulate a posterior distribution of shift configurations under a seeded
 480 prior that will maximize the probability a no shift model is sampled in the posterior.
 481 Using the prior and posterior from the seeded run, it is possible to use eqn 10 to
 482 compute the posterior of the unsampled model for the unseeded run. The downside
 483 to this approach is that, for very large empirical trees with high levels of rate
 484 variation, an extraordinarily restrictive prior must be used, and -- as we have shown
 485 -- use of more restrictive priors can decrease MCMC efficiency and lead to
 486 convergence problems.

487

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

499 the mean of the prior distribution on the number of shifts, we recommend the use of
500 a liberal model prior in studies using BAMB.

501

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

507

508 Data Accessibility: All input and output files, as well as R scripts used to analyze
509 these data, are available on Dryad at doi: 10.5061/dryad.0564q.

510

511 Supporting Information: Three online appendices are included as supporting
512 materials. Appendix I details the rationale behind, and method of calculating, the
513 cohort-matrix metric of shift location accuracy. Appendix II is a reanalysis of a
514 previously-published empirical phylogeny of whales, documenting how the changes
515 to the prior described in this paper alter empirical inference. Appendix III is an
516 illustration and discussion of how model support varies with increasing tree size.

517 References

518

519 Alfaro M.E., Santini F., Brock C., Alamillo H., Dornburg A., Rabosky D.L., Carnevale G.,
520 and Harmon L.J. (2009) Nine exceptional radiations plus high turnover explain
521 species diversity in jawed vertebrates. *Proceedings of the National Academy of*
522 *Sciences*, 106: 13410-13414.

523

524 Beaulieu J.M. and O'Meara B.C. (2015) Detecting hidden diversification shifts in
525 models of trait-dependent speciation and extinction. *bioRxiv* doi:
526 <http://dx.doi.org/10.1101/016386>

527

- 528 Callahan M.S. and McPeck M.A. (2016) Multi-locus phylogeny and divergence time
529 estimates of *Enallagma* damselflies (Odonata: Coenagrionidae). *Molecular*
530 *Phylogenetics and Evolution* 94: 182-195.
- 531
- 532 Coyne J.A. and Orr H.A. (2004) *Speciation*. Sinauer Associates, Sunderland, MA 545
533 pp.
- 534
- 535 Etienne R.S., and Haegeman B. (2012) A conceptual and statistical framework for
536 adaptive radiations with a key role for diversity dependence. *The American*
537 *Naturalist* 180: E75-E89.
- 538
- 539 FitzJohn R.G., Maddison W.P. and Otto S.P. (2009) Estimating trait-dependent
540 speciation and extinction from incompletely resolved phylogenies. *Systematic*
541 *Biology*, 58, 595-611.
- 542
- 543 Gould S.J., Raup D.M., Sepkoski J.J., Schopf T.J.M., and Simberloff D.S. (1977) The
544 shape of evolution: a comparison of real and random clades. *Paleobiology* 3: 22-40
545
- 546 Greenwood, M., and G. U. Yule. (1920). An inquiry into the nature of frequency
547 distributions representative of multiple happenings with particular reference to the
548 occurrence of multiple attacks of disease or of repeated accidents. *Journal of the*
549 *Royal Statistical Society* 83: 255-279.
- 550
- 551 Grimmett G.R., and Stirzaker D.R. (2001) Probability and Random Processes, Third
552 Edition. Oxford University Press Inc., New York.
- 553
- 554 Heaney, L. R. (1986) Biogeography of mammals in SE Asia: estimates of rates of
555 colonization, extinction and speciation. *Biological Journal of the Linnean Society*, 28,
556 127-165.
- 557

- 558 Huelsenbeck J.P., Larget B., & Alfaro M.E. (2004) Bayesian phylogenetic model
559 selection using reversible jump Markov Chain Monte Carlo. *Molecular Biology and*
560 *Evolution*, 21, 1123-1133, doi:10.1093/molbev/msh123.
- 561
- 562 Jablonski D. (2008) Species selection: Theory and data. *Annual Reviews in Ecology,*
563 *Evolution and Systematics* 39: 501-524
- 564
- 565 Jeffreys, H. (1935) Some tests of significance, treated by the theory of probability.
566 *Proceedings of the Cambridge Philosophy Society*, 31, 203-222.
- 567
- 568 Jetz, W., Thomas, G.H., Joy, J.B., Hartmann, K. & Mooers, A.O. (2012) The global
569 diversity of birds in space and time. *Nature*, 491, 444-448.
- 570
- 571 Kass, R.E. & Raftery, A.E. (1995) Bayes factors. *Journal of the American Staistical*
572 *Association*, 90, 773-795.
- 573
- 574 Liem, K.F. (1973) Evolutionary strategies and morphological innovations: cichlid
575 pharyngeal jaws. *Systematic Zoology*, 22, 425-441.
- 576
- 577 Maddison W.P., Midford P.E., & Otto S.P. (2007) Estimating a binary character's
578 effect on speciation and extinction. *Systematic Zoology*, 56, 701-710.
- 579
- 580 Morlon H., Parsons T.L., & Plotkin J.B. (2011) Reconciling molecular phylogenies
581 with the fossil record. *Proceedings of the National Academy of Sciences* 108: 16327-
582 16332
- 583
- 584 Phillimore, A.B. & Price, T.D. (2008) Density-dependent cladogenesis in birds. *PLOS*
585 *Biology*, 6:e71.
- 586
- 587 Plummer M., Best N., Cowles K., & Vines K. (2006) CODA: Convergence diagnosis and
588 output analysis for MCMC. *R News*, 6, 7-11.

589

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

592

593 Rabosky D.L, Santini F., Eastman J., Smith S.A., Sidlauskas B., Chang J & Alfaro M.E.
594 (2013) Rates of speciation and morphological evolution are correlated across the
595 largest vertebrate radiation. *Nature Communications* 4:, doi:10.1038/ncomms2958

596

597 Rabosky D.L. (2014) Automatic detection of key innovations, rate shifts, and
598 diversity-dependence on phylogenetic trees. *PLoS ONE*, 9, e89543,
599 doi:10.1371/journal.pone.0089543.

600

601 Rabosky, D.L., Donnellan, S.C., Grudler, M., & Lovette, I.J. (2014) Analysis and
602 visualization of complex macroevolutionary dynamics: An example from Australian
603 scincid lizards. *Systematic Biology*, 6, 610-627.

604

605 Rabosky D.L. & Huang H. (2015) Minimal effects of latitude on present-day
606 speciation rates in New World birds. *Proceedings of the Royal Society B* 282,
607 20142889.

608

609 Rabosky, D.L. & Huang, H. (2015). A robust semi-parametric test for detecting trait-
610 dependent diversification. *Systematic Biology*, Advanced Access, doi:
611 10.1093/sysbio/syv066.

612

613 Raup, David M. (1985) Mathematical models of cladogenesis. *Paleobiology*, 11, 42-
614 52.

615

616 Sanderson, M.J. & Donoghue, M.J. (1994) Shifts in diversification rate with the origin
617 of angiosperms. *Science*, 264, 1590-1593.

618

- 619 Simpson, G.G. (1953) The major features of evolution. Columbia Univ. Press, New
 620 York.
 621
 622 Sloss L.L. (1950) Rates of evolution. *Journal of Paleontology* 24: 131-139
 623
 624 Slowinski J.B. & Guyer C. (1989) Testing the stochasticity of patterns of organismal
 625 diversity: An improved null model. *The American Naturalist* 134: 907-921
 626
 627 Stadler, Tanja (2011) Simulating trees with a fixed number of extant species.
 628 *Systematic Biology*, 60, 676-684. doi:10.1093/sysbio/syr029.
 629
 630 Vrba, ES (1984) Evolutionary pattern and process in the sister-group Alcelaphini-
 631 Aepycerotini (Mammalia: Bovidae) in *Living Fossils*, N Eldredge and S Stanley, eds
 632 Springer-Verlag, New York.

633

Model	$\gamma = 0.1$	$\gamma = 1$	$\gamma = 100$
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

640

Table 1. Proportionality ratios for
 speciation rates (estimated versus true)

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

646 Figure 1. Prior probability of k shifts as a function of the prior mean (γ) for the old
 647 (A) and new (B) model priors. Use of the model prior implemented in BAMM v2.3
 648 and earlier results in greater prior probability of large shift numbers when γ is large.

649 However, the difference between these implementations is relatively minor for the
650 default parameterization of $\gamma = 1$ (Fig 2 - 3; Fig. S1)

651

652 Figure 2. Distribution (median and 5-95% quantiles) of model support values
653 across constant rate trees as a function of the model prior (expected number of
654 shifts; γ) in recent versions (v 2.4+) of BAMM. (A) Bayes factor evidence favoring a
655 model with rate variation (one-shift) relative to the true (zero-shift) model. Bayes
656 factors greater than one indicate support for a model with rate variation; horizontal
657 dashed line corresponds to strong or "significant" Bayes factor support (BF = 20) in
658 favor of rate variation. (B) Posterior probabilities of the zero-shift model as a
659 function of γ . Bayes factor evidence for rate variation is not sensitive with respect to
660 the prior (A), and even liberal prior distributions (e.g., $\gamma = 100$) yield no evidence for
661 rate variation for constant-rate phylogenies. Posterior probabilities are influenced
662 by the model prior (B) but did not achieve conventional thresholds ($p = 0.05$) for
663 rejecting the true (zero-shift) model, even with liberal model priors.

664

665 Figure 3. Distribution (median and 5-95% quantiles) of model support values across
666 constant rate trees as a function of the model prior for old (v 2.3 and earlier)
667 versions of BAMM, which contained an error in the Hastings ratio calculation for
668 MCMC moves that updated the Poisson rate parameter Λ . (A) Bayes factor evidence
669 favoring a model with rate variation (one-shift) relative to the true (zero-shift)
670 model. (B) Posterior probabilities of the zero-shift model as a function of γ . The
671 incorrect implementation magnified the effects of the prior on the posterior relative
672 to the correct MCMC implementation. However, across the range of model priors (γ
673 = 0.1 to $\gamma = 100$), Bayes factors did not result in strong evidence (BF = 20; horizontal
674 dashed line) for models with rate variation, despite increasing support for overly
675 complex models with increasing γ . The default model prior in BAMM ($\gamma = 1$ for all
676 versions) does not typically support models with too many shifts, even when
677 posterior probabilities alone are used for model selection (B), although increasing γ
678 did increase support for overly-complex models in the old version of BAMM.

679

680 Figure 4. Marginal posterior probability distributions on the number of shifts for
681 phylogenies simulated with and without rate heterogeneity, under three prior
682 parameterizations ($\gamma = 0.1, 1, \text{ and } 100$). Histograms represent the mean of the
683 corresponding marginal posterior distributions across all simulated phylogenies
684 with a specified level of rate heterogeneity (rows). Top row consists of 100
685 constant-rate trees; rows 2 - 5 correspond to distributions of phylogenies with 1, 2,
686 3, and 4 shifts, respectively. Phylogenies with rate variation are taken from Rabosky
687 (2014). The prior distribution on the number of shifts (red line) is illustrated in
688 each panel; filled histogram bars correspond to the true number of shifts for each
689 simulation scenario.

690

691 Figure 5. Frequency distribution of the "best model" across sets of phylogenies
692 simulated under 5 diversification scenarios (rows), selected using the step-wise
693 Bayes factor procedure described in the text. Each column represents analyses done
694 with a different model prior (left column $\gamma = 0.1$, middle column $\gamma = 1$, right column
695 $\gamma = 100$); rows (top to bottom) denote sets of phylogenies with 0, 1, 2, 3, and 4 shifts,
696 respectively. Black bars in each panel indicate the proportion of analyses where
697 BAMM recovered the true number of rate shifts in the simulated datasets. The best-
698 supported model contained an excessive number of shifts (e.g., Type I error) in 2%
699 of trees for each set of analyses. Panels only show trees that had reached
700 convergence (effective sample size > 200) with a minimum of 150 trees in each
701 panel (other than the constant rate panels where all trees converged). Convergence
702 problems arose for some analyses with $\gamma = 0.1$ (see Figure 6).

703

704 Figure 6. Spindle plots illustrating effects of model prior on convergence properties
705 of the Markov chain Monte Carlo simulation in BAMM. Each panel shows the
706 distribution of effective sample sizes in the number of rate shifts for trees simulated
707 with $k = 1, 2, 3, \text{ or } 4$ rate shifts (described by Rabosky 2014 and in the text) and
708 analyzed with three different model prior values ($\gamma = 0.1, 1 \text{ and } 100$). Width of bars

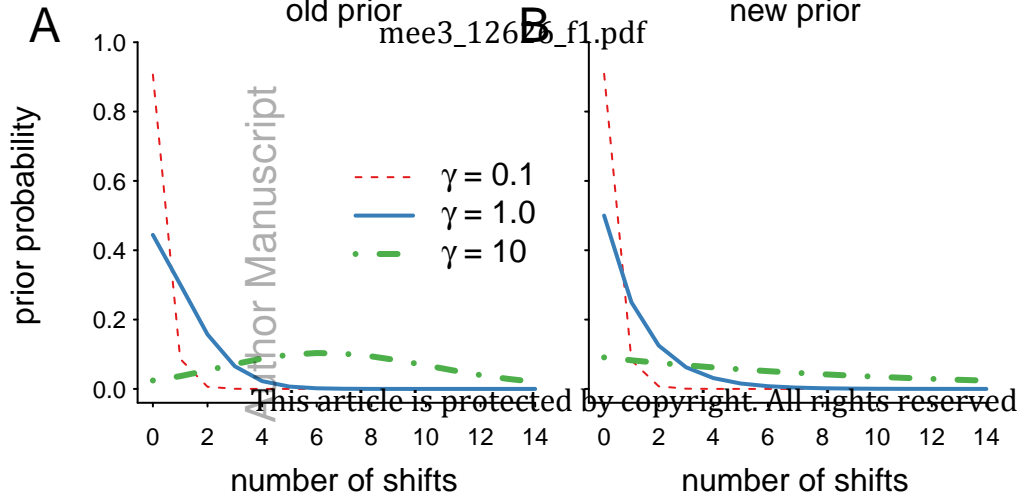
709 is proportional to the number of trees with effective sample sizes that fall into each
710 bin. A set of analyses with good convergence properties would appear "top-heavy";
711 conversely, a set of analyses with poor convergence properties would appear
712 "bottom-heavy", reflecting a high proportion of analyses with low effective sample
713 sizes. Analyses that specified larger values of γ (expected numbers of shifts) result
714 in larger effective sample sizes (i.e., chains run with liberal priors were more likely
715 to converge and converged more quickly), relative to analyses with small values of γ .

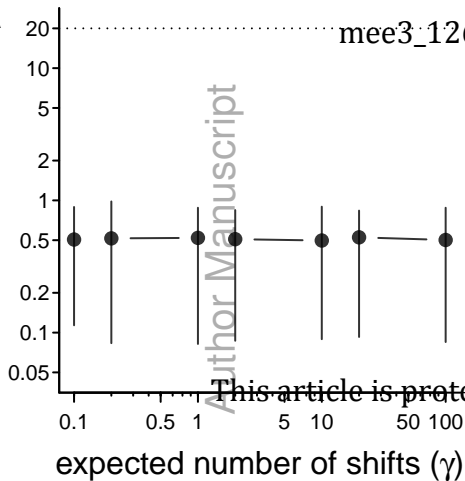
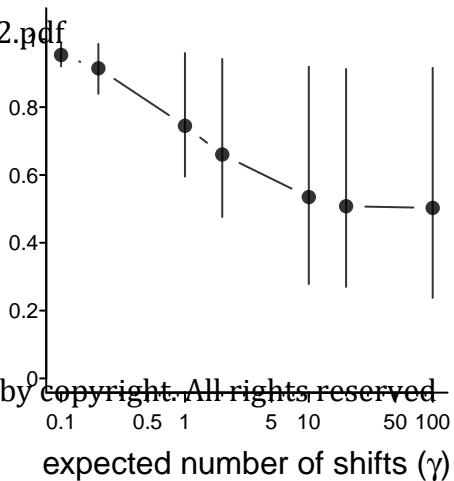
716

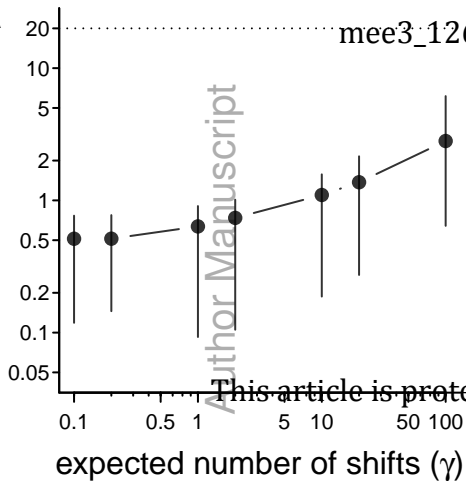
717 Figure 7. Shift location accuracy is independent of model prior. For each tree, we
718 computed the mean cohort assignment accuracy, a measure of the extent to which
719 BAMM correctly assigns taxa to the same (or different) rate regime. On average,
720 cohort accuracy under each prior exceeded 0.95. Values of 1.0 can only be obtained
721 when BAMM correctly infers the correct location of all rate shifts for each sample
722 from the posterior. For comparison, the distribution of mean cohort assignment
723 accuracies are shown after randomizing shift locations across the focal phylogenies.

724

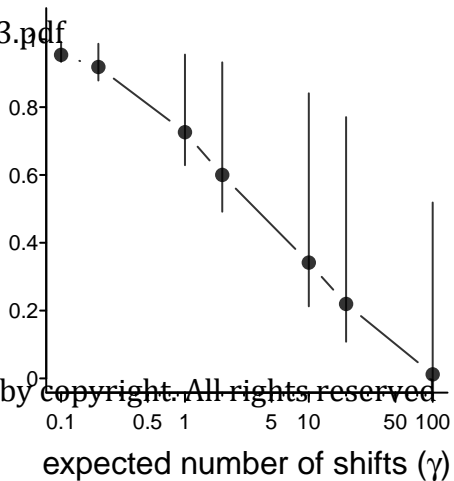
725 Figure 8. Estimates of speciation and extinction rates are highly correlated across
726 different model priors. Each phylogeny from the 4-shift dataset was analyzed with
727 BAMM under model priors of $\gamma = 0.1$ and $\gamma = 100$. For each tree, we computed the
728 Pearson correlation between tip-specific estimates of the rate of speciation (A) or
729 extinction (B) under the two priors; highly correlated estimates indicate that BAMM
730 runs with these very different model priors resulted in concordant estimates of
731 evolutionary rates. A small number of trees showed low correlations (~ 0) in tip rate
732 estimates; these analyses generally involved runs where most of the posterior shift
733 distribution for the conservative model prior ($\gamma = 0.1$) was centered on 0 or 1 shift.

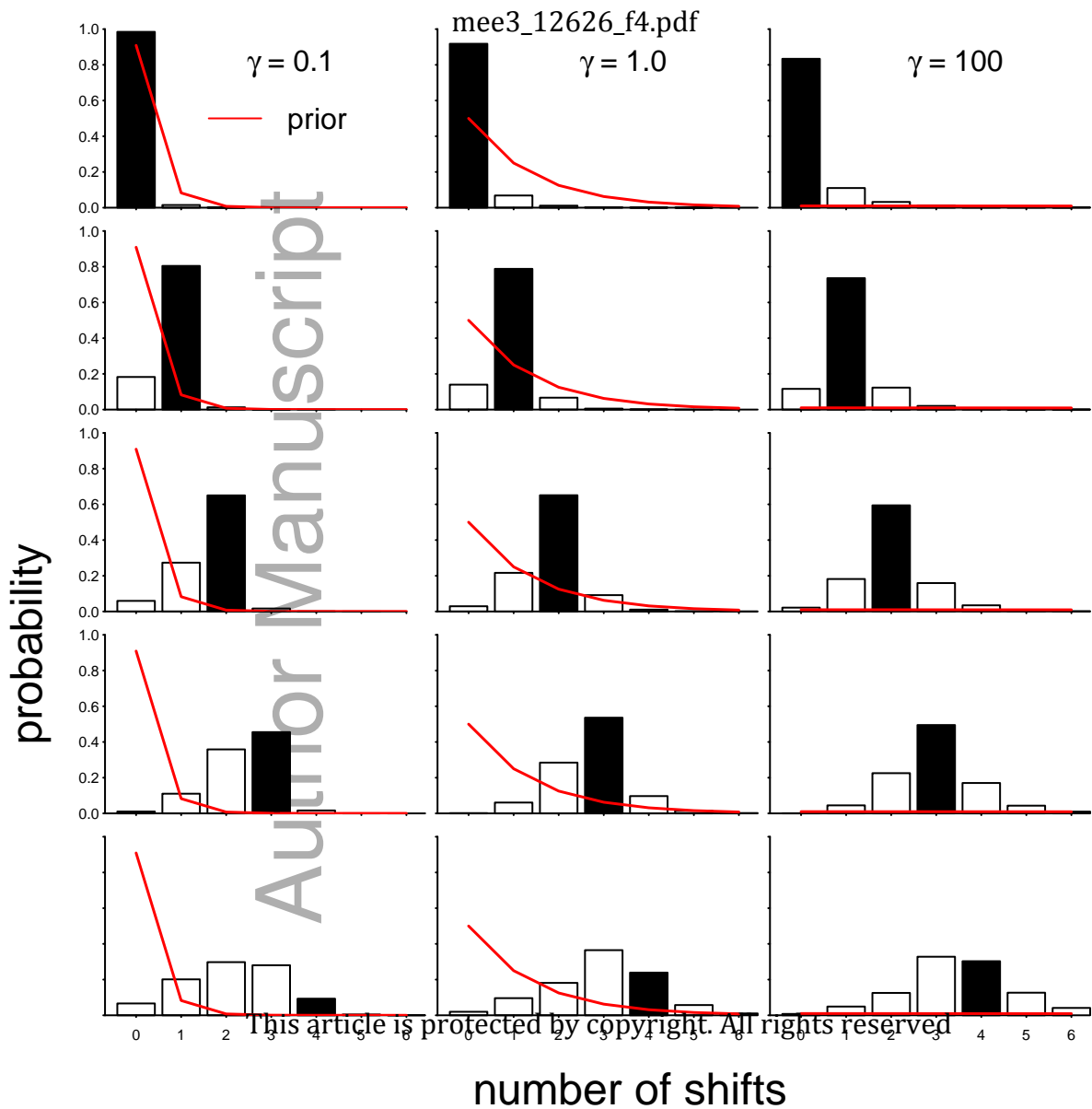


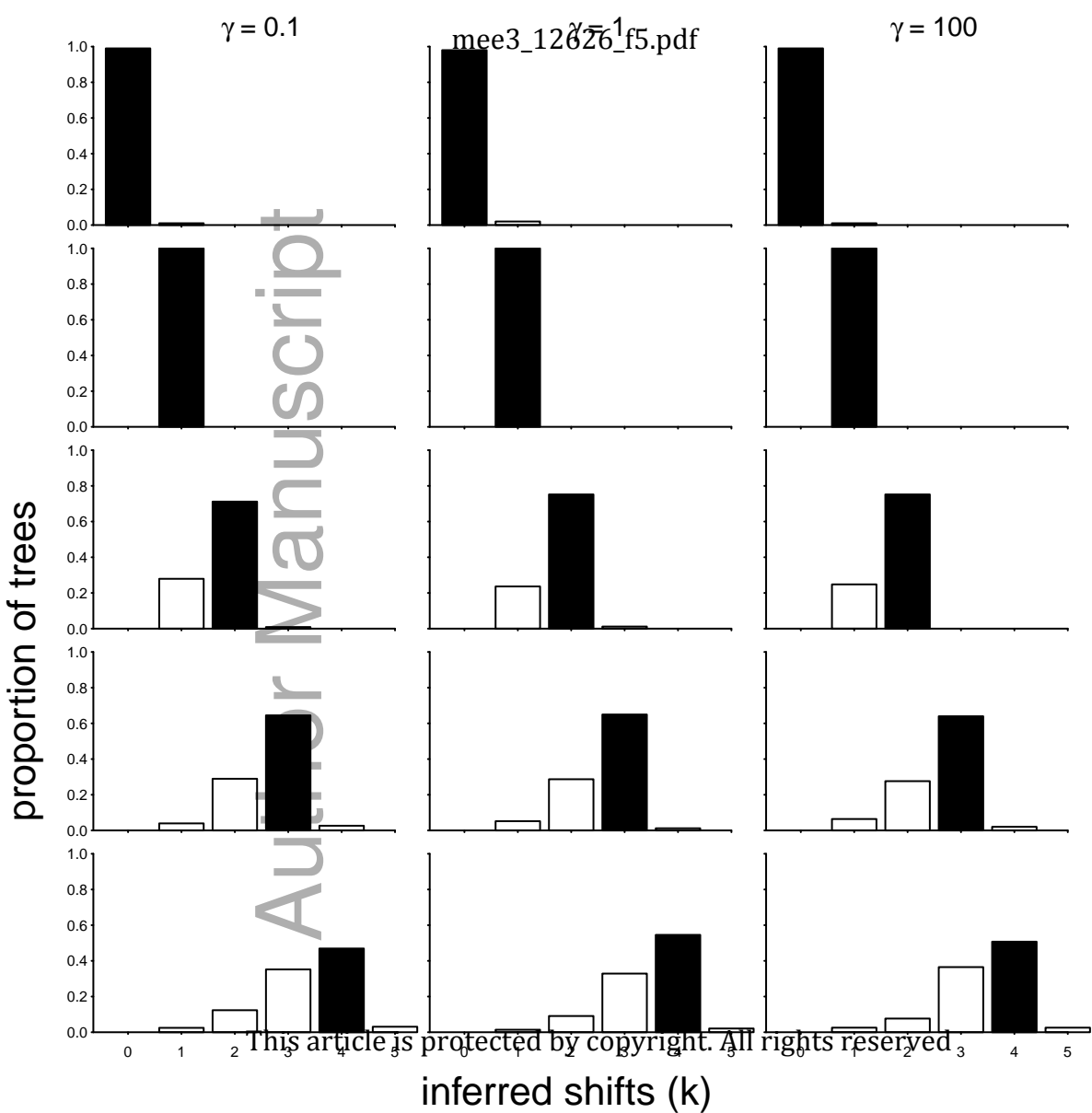
A
Bayes factor M_1 vs M_0 **B**
post. prob. 0 shifts

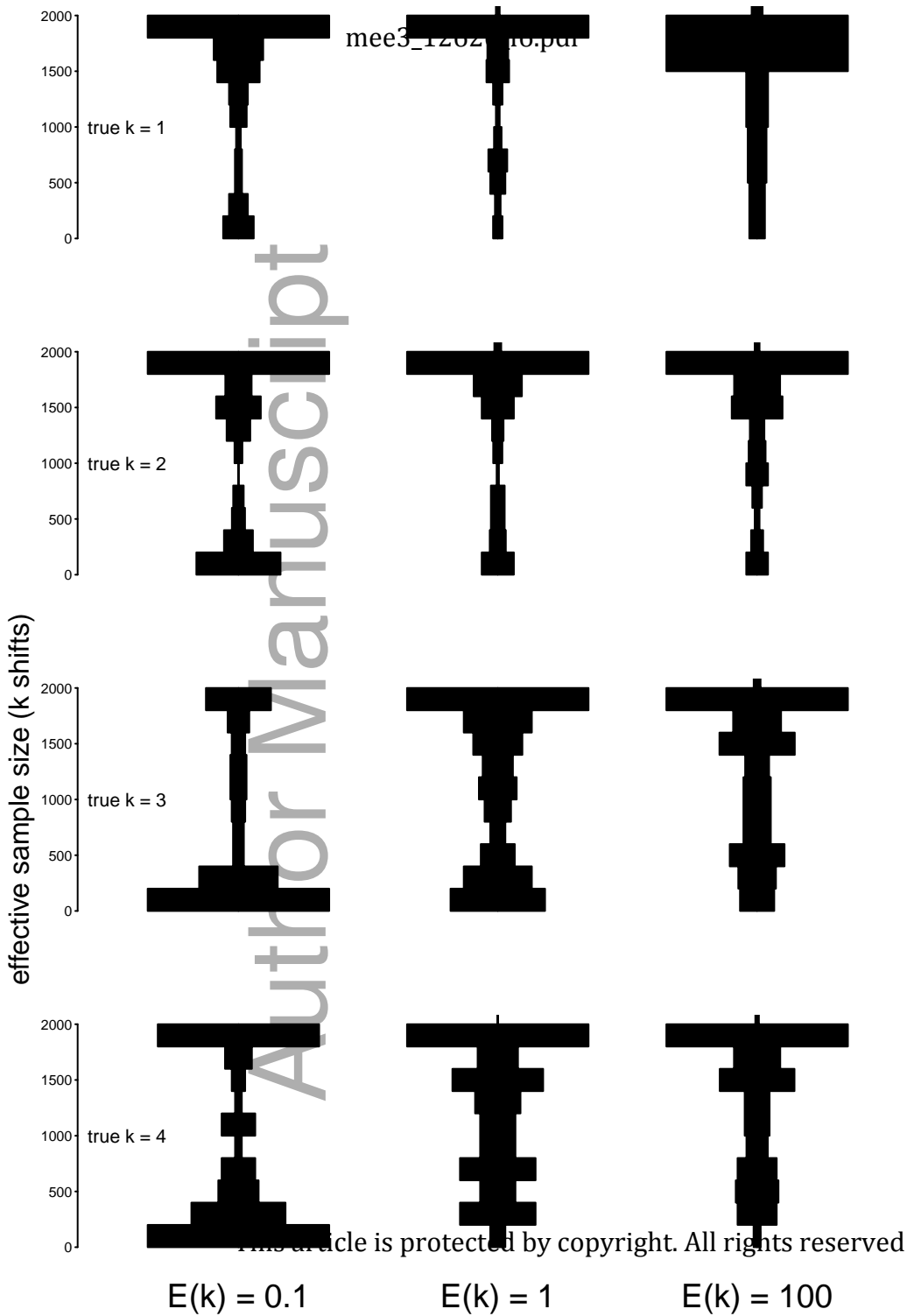
ABayes factor M_1 vs M_0 **B**

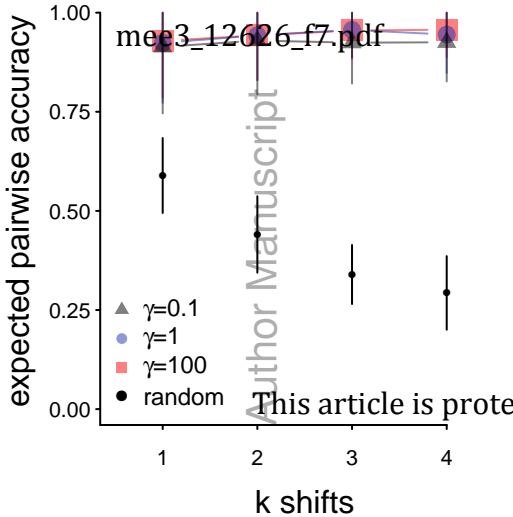
post. prob. 0 shifts





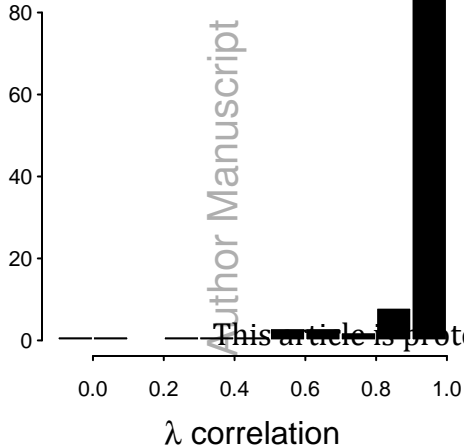




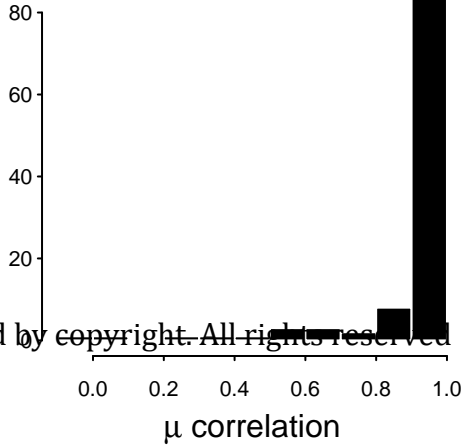


A

of 4-shift trees



B



mee3-2626_f8.pdf

Author Manuscript

This article is protected by copyright. All rights reserved