Diversification rates and phylogenies: what are we estimating, and how good are the estimates?

Pascal O. Title[1,2], Daniel L. Rabosky[1]

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

[2]ptitle@umich.edu

Abstract

1. Species-specific diversification rates, or "tip rates", can be computed quickly from phylogenies and are widely used to study diversification rate variation in relation to geography, ecology, and phenotypes. These tip rates provide a number of theoretical and practical advantages, such as the relaxation of assumptions of rate homogeneity in trait-dependent modeling approaches. However, there is significant confusion in the literature regarding whether these metrics estimate speciation or net diversification rates. Additionally, no study has yet compared the relative performance and accuracy of tip rate metrics.
2. We compared the statistical performance of three model-free rate metrics (inverse terminal branch lengths; node density metric; DR statistic) and a model-based approach (BAMM). We applied each method to a large set of simulated phylogenies that had been generated under different diversification processes; scenarios included multi-regime time-constant and diversity-dependent trees, as well as trees where the rate of speciation evolves under a diffusion process. We summarized performance in relation to

the type of rate variation, the magnitude of rate heterogeneity and rate regime size. We also compared the ability of the metrics to estimate both speciation and net diversification rates.

3. We show decisively that model-free tip rate metrics estimate the rate of speciation and not net diversification. Error in net diversification rate estimates is high and increases dramatically as a function of the relative extinction rate. In contrast, error in speciation rate estimates is low and relatively insensitive to extinction. Across all diversification scenarios, BAMM inferred the most accurate tip rates and exhibited lower error than non-model-based approaches. DR was highly correlated with true speciation rates but exhibited high sample variance, and was the best metric for very small rate regimes.
4. We found that DR and BAMM are useful metrics for studying speciation rate dynamics and trait-dependent diversification. Although BAMM was more accurate than DR overall, the two approaches have complementary strengths. Because tip rate metrics are more reliable estimators of speciation rate, we recommend that empirical studies using these metrics exercise caution when drawing biological interpretations in any situation where the distinction between speciation and net diversification is important.

Keywords: tip rates, diversification, terminal branch length, node density, DR, BAMM, trait-dependent diversification

Introduction

Rates of speciation and extinction vary through time and among lineages (Nee, Mooers & Harvey 1992; Sanderson & Donoghue 1996; Etienne & Haegeman 2012; Jetz et al. 2012; Moen & Morlon 2014; Alfaro et al. 2018), contributing to dramatic heterogeneity in species richness across the tree of life (Alfaro et al. 2009; Jetz et al. 2012; Barker et al. 2013). By characterizing variation in rates of speciation and extinction, we can better understand the dynamics of biological diversity through time, across geographic and environmental gradients (Zink, Klicka & Barber 2004; Ricklefs 2006; Mittelbach et al. 2007; Silvestro, Schnitzler & Zizka 2011; Rabosky, Title & Huang 2015), and in relation to traits and key innovations (FitzJohn, Maddison & Otto 2009; Near et al. 2012; Beaulieu & O'Meara 2016). Consequently, there has been great interest in statistical methods for inferring rates of speciation and extinction from molecular phylogenies.

Although rates of diversification have traditionally been quantified for clades, there has been a growing interest in estimating species-specific rates of diversification, which we refer to here as "tip rates". Tip rates are increasingly used to describe patterns of geographic and trait-associated variation in diversification (Freckleton, Phillimore & Pagel 2008; Jetz et al. 2012; Kennedy et al. 2016; Harvey & Rabosky 2017; Quintero & Jetz 2018; Rabosky et al. 2018). It may seem strange to view evolutionary rates as a property of individual lineages, but such rates emerge naturally from the birth-death model we typically use to conceptualize the diversification process (Nee, Mooers & Harvey 1992; Nee, May & Harvey 1994). Under the birth-death process, individuals (species) are characterized by per-lineage rates of species origination (speciation, λ) and extinction (μ). For the purposes of inference, these rates are typically assumed to be constant among contemporaneous members of a focal clade. However, tip rates

69 can be viewed as our best estimate of the present-day rate of speciation or extinction for an
70 individual lineage, conditional on past (usually recent) evolutionary history. As such, they
71 provide information about the expected amount of time that will elapse before a lineage splits or
72 becomes extinct.

73 A number of approaches have been used to estimate tip rates, including both model-based
74 and non-model-based approaches. These approaches vary in terms of how much information
75 they derive from a focal species relative to the amount of information they incorporate from
76 other regions of the phylogeny. On one end of the spectrum, tree-wide estimates of speciation
77 and extinction rates under a constant-rate birth-death (CRBD) model provides tip rates that are
78 maximally auto-correlated across species in the clade; such rates for any given species are not
79 independent of rates for any other species in the group of interest. On the other end of the
80 spectrum, terminal branch lengths can be used to derive a censored estimate of the rate of
81 speciation that is minimally autocorrelated with rates for other species in the focal clade.
82 Terminal branch lengths are largely unique to each species (rates might be identical only for
83 sister taxa), but provide a noisy measure of speciation, due to the stochasticity inherent in the
84 diversification process (Nee, May & Harvey 1994). In contrast to single (terminal) branch
85 estimates, tree-wide estimates should be less susceptible to stochastic noise, because they
86 incorporate information from the entirety of the tree (e.g., multiple branches are used in the
87 estimates). Of course, the tree-wide estimate necessarily assumes that all tips share a common
88 underlying diversification process. Other tip rate metrics fall somewhere between these two
89 extremes, incorporating some tree-wide information but relaxing the assumption of
90 homogeneous rates across all lineages (node density metric: Freckleton, Phillimore & Pagel

2008; DR: Jetz et al. 2012). The estimation of tip-specific rates thus entails a tradeoff between the precision of individual estimates and the stochastic error associated with those estimates.

BAMM (Bayesian Analysis of Macroevolutionary Mixtures, Rabosky 2014) is a model-based approach that can accommodate heterogeneity in the rate of diversification through time and among lineages. BAMM simulates a posterior distribution of macroevolutionary rate shift configurations given a phylogeny of interest; marginal rates of speciation and extinction for individual taxa can then be extracted from this distribution. In this framework, the correlation in rates between any pair of species is a function of the posterior probability that they share a common macroevolutionary rate regime (Rabosky et al. 2014). If the tree-wide posterior probability of rate variation is low, the marginal rates estimates for individual species will be similar across the entire tree, as under a CRBD model. Likewise, any pair of taxa that are consistently assigned to the same macroevolutionary rate regime will necessarily have perfectly autocorrelated rates.

Tip rates are best suited to a host of questions and hypotheses where the diversification dynamics over the evolutionary history of a group are either less relevant, or no more relevant, than the rates of diversification closer to the present day. For example, many hypotheses involving trait-dependent diversification implicitly assume a time-homogeneous effect of the trait on diversification rate (Coyne & Orr 2004; Kay et al. 2006; Jablonski 2008; FitzJohn 2010; Claramunt et al. 2011). Harvey & Rabosky (2017) found that the use of tip rates for assessing correlations between continuous traits and diversification has good performance across a range of diversification scenarios. Furthermore, hypotheses pertaining to non-historical geographic patterns of diversity are also better addressed with recent rates of diversification. For example, many hypotheses for the latitudinal diversity gradient propose time-homogeneous effects of

114 particular environmental factors (temperature, energy, geographic area) on rates of
115 diversification (Mittelbach et al. 2007; Kennedy et al. 2014; Rabosky, Title. & Huang 2015;
116 Rabosky et al. 2018). Put simply, if such time-homogeneous processes have shaped the
117 latitudinal diversity gradient (e.g., correlation between speciation and temperature: Rohde 1992),
118 then the effect should be manifest in the distribution of present-day evolutionary rates.

119 At present, there is significant confusion in the literature over what quantity various tip
120 rate metrics actually measure. The DR statistic (Jetz et al. 2012) was originally described as a
121 measure of net diversification rate, where net diversification rate (r) is the difference between the
122 rate of speciation (λ) and extinction (μ). However, subsequent work suggested that DR was a
123 better measure of speciation rate (Belmaker & Jetz 2015). Many studies have nonetheless
124 continued to describe DR as an estimate of the lineage-level net diversification rate (Marin &
125 Hedges 2016; Oliveira et al. 2016; Cai et al. 2017; Quintero & Jetz 2018; and many others). The
126 node density metric of Freckleton, Phillimore & Pagel (2008) has also been described as a
127 measure of net diversification. Whether these metrics more accurately measure speciation or net
128 diversification is critically important for interpreting biodiversity patterns (e.g., two regions
129 might differ dramatically in speciation rate, but net diversification rates in each might
130 nonetheless be zero). An initial objective of our study is thus to compare the ability of DR, node
131 density, and other metrics to estimate speciation and net diversification rates.

132 Despite the potential utility of tip rates in geographic and trait-based analyses of
133 speciation rate heterogeneity (Jetz et al. 2012; Belmaker & Jetz 2015; Oliveira et al. 2017;
134 Quintero & Jetz 2018), there has yet been no comparative assessment of the accuracy and
135 precision of the estimates. BAMM has low power to infer small rate regimes (Rabosky, Mitchell
136 & Chang 2017; Meyer & Wiens 2017), leading to the possibility that other approaches might

perform better for smaller phylogenies or when the variation in rates among clades is subtle. However, DR and related methods will always identify variation in tip rates, even when none exists, provided there is stochastic variation in branch lengths. A goal of this study is therefore to evaluate the trade-off between the stochastic noise inherent in non-model-based approaches, and the conservative but less noisy estimates from model-based metrics. We compare the performance of these metrics across a range of simulation scenarios, which include both discrete and continuous variation in rates.

Methods

Tip rate metrics

We assessed the accuracy of four tip rate metrics in this study at quantifying rates of speciation. As we demonstrate below (see also Belmaker & Jetz 2015), these metrics are estimators of speciation rate and not net diversification rate, and we refer to them as such throughout. The first metric is the inverse of the equal splits measure (Redding and Mooers 2006), also called the *DR statistic* (Jetz et al. 2012), *DivRate* (Belmaker & Jetz 2015; Oliveira et al. 2017), or *tip DR* (Quintero & Jetz 2018), which we denote in this study as λ_{DR}. This species-specific measure incorporates the number of splitting events and the internode distances along the root-to-tip path of a phylogeny, while giving greater weight to branches closer to the present (Redding & Mooers 2006; Jetz et al. 2012). λ_{DR} is computed as:

$$\lambda_{DR_i} = \sum_{j=1}^{N_i} l_j \frac{1}{2^{j-1}}$$

where λ_{DR_i} is the tip rate for species i, N_i is the number of branches between species i and the root, l_j is the length of branch j, starting at the terminal branch (j = 1) and ending with the root.

159 We also considered a simpler metric, node density (Freckleton, Phillimore & Pagel 2008;
160 denoted by λ_{ND}). This is simply the number of splitting events along the path between the root
161 and tip of a phylogeny, divided by the age of the phylogeny. While λ_{DR} down-weights the
162 contribution of branch lengths that are closer to the root, λ_{ND} equally weights the contributions of
163 all branches along a particular root-to-tip path, regardless of where they occur in time. Under a
164 pure-birth model ($\mu = 0$), both λ_{DR} and λ_{ND} should yield unbiased estimates of the rate of
165 speciation.

166 The third measure we considered is the inverse of the terminal branch lengths (λ_{TB}).
167 Rapid speciation rates near the present should be associated with proportionately shorter terminal
168 branches; smaller values of λ_{TB} should thus characterize species with faster rates of speciation.
169 This measure has recently been used as a summary statistic to assess model adequacy in trait-
170 dependent diversification studies (Bromham, Hua & Cardillo 2016; Gomes, Sorenson & Cardoso
171 2016; Harvey & Rabosky 2017). It should be noted that λ_{TB} is theoretically expected to
172 overestimate the rate of speciation. Under a pure-birth process, the set of waiting times between
173 successive speciation events can be thought of as draws from an exponential distribution with
174 rate λ. However, terminal branches are not waiting times between successive events: they are
175 censored observations, in that they are random samples of times that are necessarily less than the
176 next speciation event, which has not yet occurred at the present.

177 Finally, we considered a Bayesian, model-based approach to estimating tip rates. BAMM
178 (Rabosky 2014) assumes that phylogenies are generated by set of discrete diversification
179 regimes. Using MCMC, the program simulates a posterior distribution of rate shift regimes, from
180 which marginal posterior rate distributions can be extracted for each tip in the phylogeny. We
181 denote BAMM tip speciation rates (mean of the marginal posterior) as λ_{BAMM}. As BAMM also

182 estimates extinction rates for each regime, we also calculated tip-specific net diversification rate
183 as λ_{BAMM} - μ_{BAMM}, denoted as r_{BAMM}.
184
185 Tip rate metrics estimate speciation, not net diversification
186 As suggested previously (Belmaker & Jetz 2015), DR and presumably other tip-based
187 measurements, more accurately estimate the rate of speciation than the rate of net diversification.
188 However, numerous studies continue to refer to DR as a measure of net diversification (Marin &
189 Hedges 2016; Oliveira et al. 2016; Cai et al. 2017; Quintero & Jetz 2018; and many others). This
190 is incorrect and it is straightforward to demonstrate that λ_{TB}, λ_{ND} and λ_{DR} are more reliable
191 measures of speciation rates and not net diversification rates, at least when extinction is moderate
192 to high.
193 To illustrate this property of the metrics, we applied all approaches to constant-rate birth-
194 death phylogenies simulated across a range of relative extinction rates ($\varepsilon = \mu / \lambda$) , including pure-
195 birth trees ($\varepsilon = 0$) as well as trees exhibiting very high turnover ($\varepsilon = 1$). To evaluate accuracy of
196 speciation estimates as a function of ε, we generated 1000 phylogenies with 100 tips each, where
197 λ and ε were drawn from uniform distributions (λ: [0.05, 0.3]; ε: [0, 1]). Importantly, when λ is
198 sampled uniformly with respect to ε, the distribution of *r* is not uniform: the mean, range and
199 variance in *r* decrease dramatically as ε increases. To evaluate the accuracy of *r* as a function of
200 ε, we thus generated a second set of trees by sampling *r* and ε from uniform distributions (*r*:
201 [0.05, 0.3], ε [0, 1]). As a result, λ has constant mean and variance with respect to ε in the first
202 set of simulations, and the same is true for *r* in the second set of simulations (Figure S1). All
203 phylogeny simulations were conducted with the TreeSim package in R (Stadler 2011).

We compared tip rate metrics to true speciation rates λ_{TRUE} (with the first simulation set) and to true net diversification rates r_{TRUE} (with the second simulation set). We evaluated mean per-tip accuracy of the tip rate metrics with two measures of error:

$$mean\ absolute\ error = \sum_{i=1}^{N_i} \left|\lambda_i - \lambda_{TRUE_i}\right| \Big/ N$$

$$mean\ proportional\ error = \sum_{i=1}^{N_i} \frac{\lambda_i - \lambda_{TRUE_i}}{\lambda_{TRUE_i}} \Big/ N$$

where λ_i is the estimated tip rate for species *i* out of N total species, λ_{TRUE} is the true tip rate. Mean absolute error captures the magnitude in error in tip rates, and mean proportional error quantifies the bias in tip rates, as a function of the true tip rates (Rabosky et al. 2014).

Assessment of tip rate metrics

We tested the performance of the metrics by compiling publicly-available datasets from a number of simulation-based studies (Table 1). These simulated trees include rate heterogeneity in time and across lineages. Together, these phylogenies present a wide range of tree sizes and diversification rate shifts, providing an ideal comparative dataset for our purposes. To more easily distinguish between these tree types in the text, we refer to the BAMM-type, multi-regime time-constant phylogenies simply as “multi-regime”, and the multi-regime diversity-dependent phylogenies simply as “diversity-dependent”, even though discrete rate shifts are present in both types of trees. In addition to discrete-shift scenarios (e.g., BAMM-type process), we simulated phylogenies under an "evolving rates" model of diversification (Rabosky 2010; as corrected in Beaulieu & O’Meara 2015) to explore performance of tip rate metrics when diversification rates change continuously and independently along branches, as might occur if diversification rates

225 are correlated with an underlying continuous trait (FitzJohn 2010). In these simulations, we
226 allowed the logarithm of λ to evolve across the tree under a Brownian motion process, while
227 holding ε constant. The magnitude of rate heterogeneity among branches is controlled by the
228 diffusion parameter σ, where greater values lead to greater heterogeneity in speciation rates.
229 Although published phylogenies with rate data were unavailable for this simulation scenario, we
230 used simulation code and parameters taken directly from Beaulieu & O'Meara (2015) to generate
231 trees with similar statistical properties to those in their study. Simulations were performed with
232 the following parameters: λ = 0.078, 0.103, 0.145, 0.249 and ε = 0.0, 0.25, 0.50, 0.75. We
233 simulated 100 phylogenies for each (λ , ε) pair, and for three values of σ (σ = 0.03, 0.06, 0.12).
234 We evaluated tip rate accuracy by comparing estimated to true tip rates, using the absolute and
235 proportional error metrics described above. We also examined the correlation between true and
236 estimated tip rates, combining tip rates from all phylogenies generated under the same class of
237 diversification process, and visualizing these data as density scatterplots, generated with the LSD
238 package in R (Schwalb et al. 2018), where colors indicate the density of points.
239 Size of diversification rate regimes might be an important factor in a tip rate metric's
240 ability to accurately estimate rates. For example, BAMM's statistical power in detecting a shift
241 to a new rate regime is a function of the number of taxa in that rate regime, and tip rates for taxa
242 from small regimes will more likely be parameterized according to the larger parent regime or
243 the tree-wide average rate (Rabosky, Mitchell & Chang 2017); this is the expected behavior
244 when BAMM fails to identify a rate shift. However, non-model-based approaches such as those
245 examined in this study might be more accurate for small regimes. To explore how rate regime
246 size influences the accuracy of tip rate metrics, we calculated the mean tip rate for each true rate
247 regime from all multi-regime phylogenies (simulation datasets from Moore et al. 2016; Rabosky,

Mitchell & Chang 2017; Meyer & Wiens 2017; Mitchell, Etienne & Rabosky 2018). We then calculated the Pearson correlation coefficient and the slope of a linear model between true and estimated mean regime rates. We explored the performance of all metrics with respect to regime sample size, as in Rabosky, Mitchell & Chang (2017: Figure 13). For comparison, we repeated all performance summaries on tip rates estimated by applying a simple constant-rate birth-death (CRBD) process to each simulated phylogeny. This exercise is an important control, because it indicates how much error we would expect for each simulated phylogeny under the simplifying (incorrect) assumption that rates are constant among lineages and through time for each dataset.

Results

Speciation or net diversification?

As expected, the tip rate metrics examined in this study are more accurate estimators of the rate of speciation (λ) and not the net rate of species diversification (r). Mean absolute error increased exponentially with respect to the relative extinction rate ε (Figure 1). However, mean absolute error in speciation rate was largely invariant with respect to ε (0.95 quantile of r-based and λ-based mean absolute error for λ_{DR}: 2.28 and 0.17, respectively). Note that r and λ for these simulations were drawn from identical uniform distributions, and absolute error in the rates is thus comparable. Proportional error generally exhibited the same pattern, and in terms of λ versus r, differences in speciation-based error varied across ε (Figure S2). λ_{ND} and λ_{DR} had a tendency to overestimate λ when relative extinction was low, and underestimate λ when relative extinction was high. This trend was not present in λ_{BAMM}. Overall, error was highest for λ_{TB} by two orders of magnitude (Figure S3), and decreased progressively with λ_{ND} and λ_{DR}, with the

lowest overall error in λ_{BAMM}. BAMM estimates of net diversification rate were relatively accurate, except at the highest values of ε (Figures 1, S2).

Tip rate accuracy across rate-variable phylogenies

Tip rates estimated with BAMM were consistently more accurate than those obtained using the other methods across all diversification scenarios considered, including multi-regime, diversity-dependent and evolving rates trees (Figure 2). λ_{DR} was the second-most accurate metric, although its relationship with true rates was substantially weaker than λ_{BAMM}. λ_{ND} and λ_{TB} were correlated with true rates but performed relatively poorly in all scenarios, with λ_{TB} massively overestimating tip rates (Figure S4). All metrics performed best for multi-regime trees, followed by evolving rates and diversity-dependent trees, respectively. For diversity-dependent trees, λ_{ND} rates are effectively uncorrelated with the true rates (Figure 2). Additionally, the performance of the different tip rate metrics for multi-regime phylogenies is not sensitive to the source of the simulated phylogenies (Figure S5). We found that BAMM substantially outperformed all other metrics on datasets from studies that independently assessed BAMM's performance (Figure S5: Moore et al. 2016; Meyer & Wiens 2017). Tip rates were also generally but more weakly correlated with true net diversification rates, with the exception of λ_{ND}, which was not at all correlated with true rates (Figure S6).

In terms of mean per-tip error, λ_{BAMM} consistently outperformed the other metrics for multi-regime, diversity-dependent and evolving rates trees (Figure 3). Error in λ_{BAMM} increased as a function of rate heterogeneity for evolving rate phylogenies, but was largely independent of the magnitude of rate heterogeneity for the other scenarios. λ_{DR} generally exhibited greater error than λ_{BAMM}, and this error increased as a function of the level of heterogeneity for both the

evolving rates and multi-regime trees. Error in λ_{DR} was generally invariant to the number of rate regimes for the diversity-dependent scenarios. However, λ_{DR} tended to have greater error than tree-wide estimates of speciation rates from a simple model that assumes no variation in rates through time or among lineages (λ_{CRBD}). λ_{ND} performed somewhat similarly to λ_{DR} for constant-rate and evolving rates trees, but worse for diversity-dependent trees. Error in λ_{TB} increased with increasing rate heterogeneity for constant-rate and evolving rates trees, but was relatively unaffected by rate heterogeneity in diversity-dependent trees (Figure S7). However, error for this metric was far greater than for all other tip metrics.

<u>Effects of regime size on performance</u>

Both metrics of performance assessment – the Pearson correlation and OLS slope – generally increased with increasing regime size (Figure 4). This was found to be true for all tip rate metrics, although λ_{TB} and λ_{ND} never achieved high performance. λ_{DR} tended to perform better than other metrics when small rate regimes were included (e.g., 10 tips or fewer); however, the slope between estimated and true rates was greater than 1 across the majority of minimum regime sizes, indicating that λ_{DR} overestimates speciation rates (see also Figure S2). Similar patterns were observed for net diversification rates with λ_{DR}, but the magnitude of the overestimation was greater than for speciation (Figure S8). λ_{BAMM}, in contrast, approached a slope of 1 when estimating speciation rates and slightly underestimated net diversification rates (regimes with > 30 tips: OLS slope = 0.96 for λ, 0.87 for r).

Absolute error in regime mean tip rates was lowest for λ_{DR} and λ_{BAMM}, regardless of the size of the rate regime (Figure 5). BAMM's ability to accurately estimate tip rates improved with regime size, whereas absolute error was relatively consistent across regime sizes for λ_{DR} for

regimes greater than 10 species. We also found that λ_{DR} slightly outperformed λ_{BAMM} for small rate regimes.

Note that, in Figures 4 and 5, each rate regime is treated as a single data point. Rate regimes of sizes 1000, 100, and 1 tip are equivalent under this method of error assessment. Figure 4 assesses how well these methods estimate rates for individual regimes, regardless of the size of those regimes. In contrast, Figures 1-3 ask how well these methods perform at estimating rates for a given tip.

Discussion

We assessed several tip rate metrics and confirmed that these are more accurate estimators of the rate of speciation, rather than net diversification (Figures 1, 4, S6, S8; see also Belmaker & Jetz 2015). This distinction was especially pronounced at high relative extinction rates, where the rate of lineage turnover is high, and rates of speciation and net diversification have the potential to be most differentiated. Net diversification rate is a critical determinant of species richness, yet this quantity is potentially independent of the underlying rate of speciation. Misinterpretation of tip rate metrics could therefore lead to highly misleading perspectives on large-scale diversity dynamics. As we demonstrate (Figures 1, S2), tip rate metrics (λ_{ND}, λ_{DR}) provide relatively little information about net diversification, and high values of these metrics are fully consistent with equilibrial models of speciation where the true net diversification rate is zero. Thus, λ_{DR} and λ_{ND} should not be used to support claims about the dynamics of species richness or net diversification *per se* without independent evidence bearing on plausible levels of extinction.

In terms of accuracy, we found that BAMM performed better than non-model-based metrics across all datasets we considered: estimated tip rates were most highly correlated with true tip rates, and mean per-tip error in rates was lower across a range of rate-variable simulation scenarios. BAMM is expected to perform well for phylogenies with discrete shifts in diversification rates as this type of rate variation is most consistent with BAMM's assumptions (Rabosky 2014; Mitchell & Rabosky 2016; Rabosky, Mitchell & Chang 2017; Mitchell, Etienne & Rabosky 2018). However, BAMM performed surprisingly well for the evolving rates phylogenies, which conform poorly to the assumptions of the inference model. In these trees, the rate of speciation changes continuously under a diffusion process, and as a result, the phylogeny exhibits rate heterogeneity without discrete rate shifts.

On evolving rates phylogenies, λ_{BAMM} performed better than λ_{DR} (Figure 2; Spearman's ρ for $\lambda_{BAMM} = 0.83$, ρ for $\lambda_{DR} = 0.62$), despite the fact that λ_{DR} does not rely on the detection of distinct rate regimes to estimate tip rates (Figure 5). λ_{BAMM} also exhibited the lowest mean per-tip error across varying levels of rate heterogeneity (Figure 3).

Why do λ_{BAMM} and λ_{DR} exhibit such striking differences in performance across the simulation scenarios considered here? To illustrate the differences between inference under these metrics, we compared true tip rates to λ_{BAMM} and to λ_{DR} on a simulated birth-death tree with a single rate shift (Figure 6), as well as on one evolving rates tree simulated for this study (Figure 7). It is clear that if BAMM has the statistical power to detect true rate shifts, then it will perform well under rate shift scenarios. In contrast, λ_{DR} tracks true rate shifts but exhibits high sample variance. With an evolving rates tree (Figure 7), the simulation model is very different from the inference model in BAMM. However, it conservatively places rate shifts in order to accommodate rate heterogeneity that is spread across the phylogeny under a rather different

361 model of rate variation. λ_{DR} also broadly tracks the overall pattern of the true rates, but the
362 variance in the corresponding estimates is so high that performance is negatively affected. If we
363 calculate mean (absolute) per-tip error in λ_{BAMM} and $\lambda_{DR,}$ the error is relatively similar between
364 λ_{BAMM} and λ_{DR}, but the variance in per-tip error for λ_{DR} is higher. Overall, BAMM exhibited
365 substantially lower error than λ_{DR} under precisely this scenario (Figure 3).

366 Thus, although BAMM is conservative in the estimation of tip rates relative to λ_{DR}, the
367 method exhibits lower overall error. It appears that λ_{DR} can recover more subtle rate
368 heterogeneity relative to BAMM (see Rabosky, Mitchell & Chang 2017 for discussion of power
369 in BAMM), but this apparent power advantage comes at the cost of increased variance (error) in
370 the resulting estimates. Remarkably, on a per-tip basis, we find that a simple constant-rate birth-
371 death process (λ_{CRBD}) frequently yields tip estimates that are more accurate than those obtained
372 with λ_{DR} (Figure 3), despite the simplifying (and incorrect) assumption that rates are identical
373 across all tips in a given tree. Given that λ_{DR} can and does track true heterogeneity in speciation
374 rate (Figures 6, 7), this pattern suggests that the metric is especially sensitive to the stochastic
375 variation in branch lengths that can emerge even when all tips have the same underlying
376 speciation rate.

377 Regardless of the performance summaries presented in this article, important questions
378 remain with respect to how well tip rate metrics can estimate the true rate of speciation from
379 empirical phylogenies. The phylogenies analyzed in this study were simulated under idealized
380 processes and neglect potential biases and sources of uncertainty that are present in real datasets.
381 For example, if the process of speciation takes time to complete, as is generally believed to be
382 the case (i.e., the protracted speciation process; Rosindell et al. 2010; Etienne and Rosindell
383 2012), then the most recent speciation events may still be on-going at the present and typical

384 species-level molecular phylogenies may fail to recognize these events. This will lead to an
385 overestimation of terminal branch lengths, as some terminal branches potentially include
386 incipient species. A related bias might arise due to incomplete taxon sampling, which
387 disproportionately affects the length of terminal (or otherwise recent) branch lengths (Pybus &
388 Harvey 2000). Likewise, variation in taxonomic practice across a phylogeny might lead to
389 spurious rate variation, particularly if different species concepts are used, or if some clades in the
390 phylogeny – but not others – have been subject to population genetic analysis or screens for
391 cryptic species diversity. Additionally, it has been shown that BAMM and other methods may
392 fail to infer accurate speciation rate dynamics if the phylogeny is in diversity decline – that is,
393 when extinction rates increase towards the present and ultimately exceed speciation rates
394 (Quental & Marshall 2011; Burin et al. 2018). A major, if obvious, caveat in the interpretation of
395 tip rates is that they apply to recent speciation rates and are necessarily limited with respect to
396 inferences about historical variation in speciation rate.

397 The greater the importance of the terminal branches in tip rate metrics, the greater the
398 impact these biases might have on tip rate estimates. On one end of the spectrum, metrics such as
399 λ_{TB} will be very sensitive to such biases as they rely exclusively on terminal branch lengths.
400 Such approaches may retain utility as summary statistics (e.g., Bromham, Hua & Cardillo 2016),
401 but we did not find λ_{TB} to be an accurate estimate of speciation rates in any of our analyses. On
402 the other end of the spectrum, a metric like λ_{ND} would be minimally impacted as this metric is
403 attempting to capture an average speciation rate over an entire root-to-tip path and does not
404 upweight the contribution of recent branch lengths. λ_{DR} is likely somewhere in the middle of this
405 spectrum, as it gives decreasing weight to branches towards the root. λ_{BAMM} is potentially
406 sensitive to such issues as well, although it may be possible to analytically correct for some

407 biases in the mechanics of the model itself (e.g., Rosindell et al. 2010; Etienne and Rosindell
408 2012).

409 Potential empirical biases aside, tip rates present a number of practical advantages in the
410 study of diversification rate variation. First, tip rates can be summarized and compared across
411 non-monophyletic assemblages of species (Jetz et al. 2012; Kennedy et al. 2016; Belmaker &
412 Jetz 2015; Oliveira et al. 2017; Quintero & Jetz 2018; Rabosky et al. 2018), making it possible to
413 summarize rate characteristics of entire communities or regional assemblages of species. Second,
414 estimation of rates at the present should be more robust to the influence of extinction, as
415 extinction can erase the history of lineage splitting deeper in the phylogeny (Nee et al. 1994;
416 Nee, May & Harvey 1994; Rabosky & Lovette 2008). Third, tip-specific rates can be paired with
417 species-specific trait values or geographic attributes in order to test potential trait- or geography-
418 dependent speciation rates (Freckleton, Phillimore & Pagel 2008; Jetz et al. 2012, Rabosky &
419 Goldberg 2017; Harvey & Rabosky 2017). Tip rates make it possible to relax strong assumptions
420 of rate homogeneity within character states, which are inherent to certain trait-dependent models,
421 including BiSSE and GeoSSE (Maddison, Midford & Otto 2007; Goldberg, Lancaster & Ree
422 2011; Ng & Smith 2014). Recent work has provided a conceptually rich and robust interpretive
423 framework for SSE models that does not assume rate-constancy within character states (Beaulieu
424 & O'Meara 2016; Caetano, O'Meara & Beaulieu 2018), but tip rates nonetheless can provide an
425 important check on results obtained with SSE models by providing a direct means of visualizing
426 the relationship between branch lengths and character states (Bromham, Hua & Cardillo 2016;
427 Hua & Bromham 2016; Harvey & Rabosky 2017). Visual inspection of data in this fashion has
428 the potential to reduce false positives by calling attention to potential outliers and other sources
429 of model inadequacy (Maddison & FitzJohn 2014; Rabosky & Goldberg 2015). A final

advantage for non-model based tip rates, especially λ_{DR}, is that they can profitably be applied to extremely large phylogenies: there are few computational limits to using them on phylogenies with tens of thousands of tips or more, in contrast to formal model-based approaches for which BAMM, HiSSE, and other methods are poorly suited.

In summary, tip rates offer a number of theoretical and practical advantages, particularly in the study of associations between traits and diversification. We found that λ_{BAMM} outperformed other metrics evaluated in this study and proved to be relatively accurate, even under diversification scenarios that depart from the BAMM inference model. λ_{DR} underperformed in comparison to λ_{BAMM}, but in many cases still did reasonably well, particularly for small rate regimes. Despite our performance results, λ_{DR} is likely to remain a useful tool in the study of trait- and geography-dependent diversification (Rabosky & Goldberg 2017; Harvey & Rabosky 2017).

Acknowledgements

We thank Jonathan Mitchell for help with compiling the simulation datasets evaluated in this study. We also thank Michael Grundler, as well as all other members of the Rabosky and Davis Rabosky labs at the University of Michigan for thoughtful discussion. This work was supported in part by a University of Michigan Rackham Predoctoral Fellowship (P.O.T.) and by a Fellowship from the David and Lucile Packard Foundation (D.L.R.).

Data accessibility

All data and code necessary to repeat the analyses have been made available through the Dryad digital data repository at XXXXXX.

Author contributions

P.O.T. and D.L.R. designed the project. P.O.T. assembled the datasets and performed all analyses. P.O.T. and D.L.R. wrote the manuscript. Both authors contributed critically to subsequent drafts and approved the final publication.

References

Alfaro, M.E., Faircloth, B.C., Harrington, R.C., Sorenson, L., Friedman, M., Thacker, C.E., Oliveros, C.H., Černý, D. & Near, T.J. (2018). Explosive diversification of marine fishes at the Cretaceous–Palaeogene boundary. *Nature Ecology & Evolution*, **2**, 688–696.

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.

Barker, F.K., Burns, K.J., Klicka, J., Lanyon, S.M. & Lovette, I.J. (2013). Going to extremes: contrasting rates of diversification in a recent radiation of new world passerine birds. *Systematic Biology*, **62**, 298–320.

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.

Beaulieu, J.M. & O'Meara, B.C. (2015). Extinction can be estimated from moderately sized molecular phylogenies. *Evolution*, **69**, 1036–1043.

Belmaker, J. & Jetz, W. (2015). Relative roles of ecological and energetic constraints, diversification rates and region history on global species richness gradients. *Ecology Letters*, **18**, 563–571.

Bromham, L., Hua, X. & Cardillo, M. (2016). Detecting Macroevolutionary Self-Destruction from Phylogenies. *Systematic Biology*, **65**, 109–127.

Burin, G., Alencar, L.R.V., Chang, J., Alfaro, M.E. & Quental, T.B. (2018). How Well Can We Estimate Diversity Dynamics for Clades in Diversity Decline? *Systematic Biology*, **105**, 1–17.

Caetano, D., O'Meara, B., Beaulieu, J.2018. Hidden state models improve the adequacy of state-dependent diversification approaches using empirical trees, including biogeographical models. *biorxiv.org*.

Cai, T., Fjeldså, J., Wu, Y., Shao, S., Chen, Y., Quan, Q., Li, X., Song, G., Qu, Y., Qiao, G. & Lei, F. (2017). What makes the Sino-Himalayan mountains the major diversity hotspots for pheasants? *Journal of Biogeography*, **45**, 640–651.

Claramunt, S. (2010). Discovering exceptional diversifications at continental scales: the case of the endemic families of Neotropical suboscine passerines. *Evolution*, **64**, 2004–2019.

Coyne, J.A. & Orr, H.A. (2004). *Speciation*. Sinauer, Cambridge.

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.

Etienne, R.S. & Rosindell, J. (2012). Prolonging the Past Counteracts the Pull of the Present: Protracted Speciation Can Explain Observed Slowdowns in Diversification. *Systematic Biology*, **61**, 204–213.

FitzJohn, R.G. (2010). Quantitative traits and diversification. *Systematic Biology*, **59**, 619–633.

FitzJohn, R.G., Maddison, W.P. & Otto, S.P. (2009). Estimating Trait-Dependent Speciation and Extinction Rates from Incompletely Resolved Phylogenies. *Systematic Biology*, **58**, 595–611.

Freckleton, R.P., Phillimore, A.B. & Pagel, M. (2008). Relating Traits to Diversification: A Simple Test. *The American Naturalist*, **172**, 102–115.

Goldberg, E.E., Lancaster, L.T. & Ree, R.H. (2011). Phylogenetic inference of reciprocal effects between geographic range evolution and diversification. *Systematic Biology*, **60**, 451–465.

Gomes, A.C.R., Sorenson, M.D. & Cardoso, G.C. (2016). Speciation is associated with changing ornamentation rather than stronger sexual selection. *Evolution*, **70**, 2823–2838.

Harvey, M.G. & Rabosky, D.L. (2017). Continuous traits and speciation rates: Alternatives to state-dependent diversification models (N. Cooper, Ed.). *Methods in Ecology and Evolution*, **12**, 751–10.

Hua, X. & Bromham, L. (2016). PHYLOMETRICS: an R package for detecting macroevolutionary patterns, using phylogenetic metrics and backward tree simulation. *Methods in Ecology and Evolution*, **7**, 806–810.

Jablonski, D. (2008). Species selection: Theory and data. *Annual Review of Ecology, Evolution, and Systematics*, **39**, 501–524.

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.

Kay, K., Voelckel, C., Yang, J.Y., Hufford, K.M., Kaska, D.D. & Hodges, S.A. (2006). Floral characters and species diversification. *Ecology and Evolution of Flowers* (eds L. Harder & S. Barrett), pp. 311–325. Oxford.

Kennedy, J.D., Borregaard, M.K., Jønsson, K.A., Holt, B., Fjeldså, J. & Rahbek, C. (2016). Does the colonization of new biogeographic regions influence the diversification and accumulation of clade richness among the Corvides (Aves: Passeriformes)? *Evolution*, 38–50.

Maddison, W.P. & FitzJohn, R.G. (2014). The Unsolved Challenge to Phylogenetic Correlation Tests for Categorical Characters. *Systematic Biology*, **64**, 127–136.

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

Marin, J. & Hedges, S.B. (2016). Time best explains global variation in species richness of amphibians, birds and mammals. *Journal of Biogeography*, **43**, 1069–1079.

Meyer, A.L.S. & Wiens, J.J. (2017). Estimating diversification rates for higher taxa: BAMM can give problematic estimates of rates and rate shifts. *Evolution*, **72**, 39–53.

Mitchell, J.S. & Rabosky, D.L. (2016). Bayesian model selection with BAMM: effects of the model prior on the inferred number of diversification shifts. *Methods in Ecology and Evolution*, **8**, 37–46.

Mitchell, J.S., Etienne, R.S. & Rabosky, D.L. (2018). Inferring Diversification Rate Variation From Phylogenies With Fossils. *Systematic Biology*.

Mittelbach, G.G., Schemske, D.W., Cornell, H.V., Allen, A.P., Brown, J.M., Bush, M.B., Harrison, S.P., Hurlbert, A.H., Knowlton, N., Lessios, H.A., McCain, C.M., McCune, A.R., McDade, L.A., McPeek, M.A., Near, T.J., Price, T.D., Ricklefs, R.E., Roy, K., Sax, D.F., Schluter, D., Sobel, J.M. & Turelli, M. (2007). Evolution and the latitudinal diversity gradient: speciation, extinction and biogeography. *Ecology Letters*, **10**, 315–331.

Moen, D. & Morlon, H. (2014). Why does diversification slow down? *Trends in Ecology and Evolution*, **29**, 190–197.

Moore, B.R., Höhna, S., May, M.R., Rannala, B. & Huelsenbeck, J.P. (2016). Critically evaluating the theory and performance of Bayesian analysis of macroevolutionary mixtures. *Proceedings of the National Academy of Sciences*, **113**, 9569–9574.

Near, T.J., Dornburg, A., Kuhn, K.L., Eastman, J.T., Pennington, J.N., Patarnello, T., Zane, L., Fernandez, D.A. & Jones, C.D. (2012). Ancient climate change, antifreeze, and the evolutionary diversification of Antarctic fishes. *Proceedings of the National Academy of Sciences*, **109**, 3434–3439.

Nee, S., Holmes, E.C., May, R.M. & Harvey, P.H. (1994a). Extinction Rates Can Be Estimated From Molecular Phylogenies. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **344**, 77–82.

Nee, S., May, R.M. & Harvey, P.H. (1994b). The reconstructed evolutionary process. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **344**, 305–311.

569 Nee, S., Mooers, A. & Harvey, P.H. (1992). Tempo and Mode of Evolution Revealed From
570 Molecular Phylogenies. *Proceedings of the National Academy of Sciences*, **89**, 8322–8326.

571 Ng, J. & Smith, S.D. (2014). How traits shape trees: new approaches for detecting character
572 state-dependent lineage diversification. *Journal of Evolutionary Biology*, **27**, 2035–2045.

573 Oliveira, B.F., Machac, A., Costa, G.C., Brooks, T.M., Davidson, A.D., Rondinini, C. & Graham,
574 C.H. (2016). Species and functional diversity accumulate differently in mammals. *Global*
575 *Ecology and Biogeography*, **25**, 1119–1130.

576 Pybus, O.G. & Harvey, P.H. (2000). Testing macro-evolutionary models using incomplete
577 molecular phylogenies. *Proceedings of the Royal Society of London. Series B: Biological*
578 *Sciences*, **267**, 2267–2272.

579 Quental, T.B. & Marshall, C.R. (2011). The Molecular Phylogenetic Signature of Clades in
580 Decline. *PLoS ONE*, **6**, e25780–9.

581 Quintero, I. & Jetz, W. (2018). Global elevational diversity and diversification of birds. *Nature*,
582 **555**, 246–250.

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

585 Rabosky, D.L. (2010). Extinction rates should not be estimated from molecular phylogenies.
586 *Evolution*, **64**, 1816–1824.

587 Rabosky, D.L. & Goldberg, E.E. (2017). FiSSE: A simple nonparametric test for the effects of a
588 binary character on lineage diversification rates. *Evolution*, **106**, 13410–11.

589 Rabosky, D.L. & Goldberg, E.E. (2015). Model Inadequacy and Mistaken Inferences of Trait-
590 Dependent Speciation. *Systematic Biology*, **64**, 340–355.

591 Rabosky, D.L. & Lovette, I.J. (2008). Explosive evolutionary radiations: decreasing speciation or
592 increasing extinction through time? *Evolution*, **62**, 1866–1875.

593 Rabosky, D.L., Chang, J., Title, P.O., Cowman, P.F., Sallan, L., Friedman, M., Kaschner, K., Garilao,
594 C., Near, T.J., Coll, M. & Alfaro, M.E. (2018). An inverse latitudinal gradient in speciation
595 rate for marine fishes. *Nature*, 1–20.

596 Rabosky, D.L., Donnellan, S.C., Grundler, M. & Lovette, I.J. (2014). Analysis and Visualization of
597 Complex Macroevolutionary Dynamics: An Example from Australian Scincid Lizards.
598 *Systematic Biology*, **63**, 610–627.

599 Rabosky, D.L., Mitchell, J.S. & Chang, J. (2017). Is BAMM Flawed? Theoretical and Practical
600 Concerns in the Analysis of Multi-Rate Diversification Models. *Systematic Biology*, **66**, 477–
601 498.

602 Rabosky, D.L., Title, P.O. & Huang, H. (2015). Minimal effects of latitude on present-day
603 speciation rates in New World birds. *Proceedings of the Royal Society of London. Series B:*
604 *Biological Sciences*, **282**, 20142889.

605 Redding, D.W. & Mooers, A.O. (2006). Incorporating Evolutionary Measures into Conservation
606 Prioritization. *Conservation Biology*, **20**, 1670–1678.

607 Ricklefs, R.E. (2006). Global variation in the diversification rate of passerine birds. *Ecology*, **87**,
608 2468–2478.

609 Rohde, K. (1992). Latitudinal Gradients in Species Diversity: The Search for the Primary Cause.
610 *Oikos*, **65**, 514.

611 Rosindell, J., Cornell, S.J., Hubbell, S.P. & Etienne, R.S. (2010). Protracted speciation revitalizes
612 the neutral theory of biodiversity. *Ecology Letters*, **13**, 716–727.

613 Sanderson, M.J. & Donoghue, M.J. (1996). Reconstructing shifts in diversification rates on
614 phylogenetic trees. *Trends in Ecology and Evolution*, **11**, 15–20.

615 Silvestro, D., Schnitzler, J. & Zizka, G. (2011). A Bayesian framework to estimate diversification
616 rates and their variation through time and space. *BMC Evolutionary Biology*, **11**, 311.

617 Stadler, T. (2011). Simulating Trees with a Fixed Number of Extant Species. *Systematic Biology*,
618 **60**, 676–684.

619 Zink, R.M., Klicka, J. & Barber, B. (2004). The tempo of avian diversification during the
620 Quaternary. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **359**,
621 215–220.

622

623

Figures/Captions

Table 1. Summary of simulated phylogenies used in this study.

simulation model	number of trees	tree size	regime number	source
single-regime, constant-rate birth-death	100	100	1	Mitchell & Rabosky 2016
single- and multi-regime, constant-rate birth-death	100	51-148	1-6	Moore et al. 2016
single- and multi-regime, constant-rate birth-death	400	10-4296	1-67	Rabosky, Mitchell & Chang 2017
multi-regime, constant-rate birth-death	20	939-3708	11	Meyer & Wiens 2017
single- and multi-regime, constant-rate birth-death	188	4-3955	1-73	Mitchell, Etienne & Rabosky 2018
single-regime, constant-rate birth-death, lambda uniform	1000	100	1	this study
single-regime, constant-rate birth-death, net diversification uniform	1000	100	1	this study
pure birth root regime, 1-4 discrete shifts to diversity-dependent regimes	1200	54-882	1-5	Rabosky 2014; Mitchell & Rabosky 2016
speciation rate evolves via diffusion process	1200	25-1208	1	Rabosky 2010; Beaulieu & O'Meara 2015; Rabosky 2016; this study

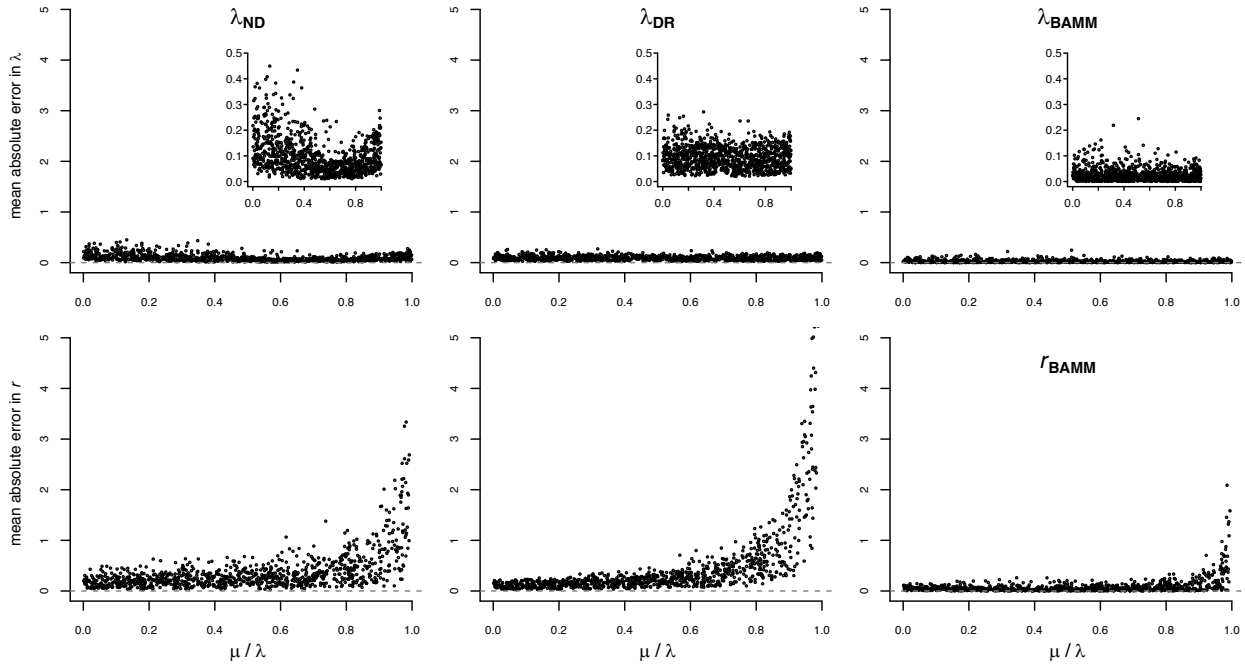
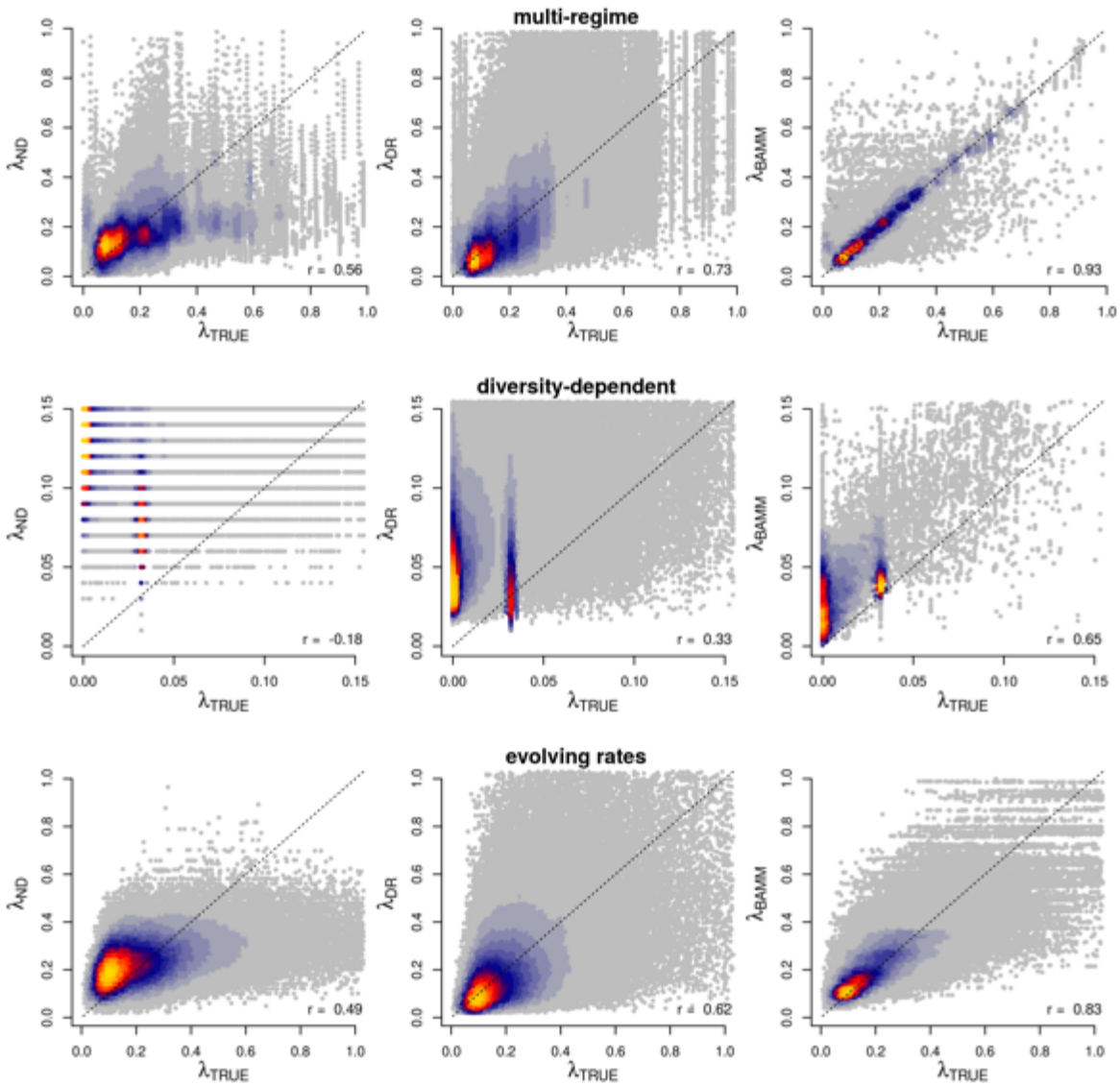


Figure 1. Mean absolute error in λ (top) and *r* (bottom) for three different tip rate metrics, across a range of relative extinction rates. For BAMM, the estimated speciation and net diversification rates are presented in the top and bottom panels, respectively. Absolute error of zero implies perfect accuracy. Inset plots show error in λ with truncated y-axis scale to facilitate comparison among metrics. All tip rate metrics track λ more accurately than they track *r*. See Figure S3 for λ_{TB}, which performed much worse than the other metrics.



650

651 **Figure 2.** True tip rates (λ_{TRUE}) in relation to estimated tip rates. Tip rates were compared

652 separately for different major categories of phylogeny simulations (rows). Plotting region is

653 restricted to the 99th percentile of true rates, but Spearman correlations between true and

654 estimated rates (lower right of each figure panel) are based on the full range of the data. Colors

655 indicate the density of points in the scatter plots. The horizontal gaps in λ_{ND} for diversity-

656 dependent trees are an artefact of all trees having the same crown age. λ_{BAMM} exhibited the

657 strongest correlation with true rates for all simulation categories.

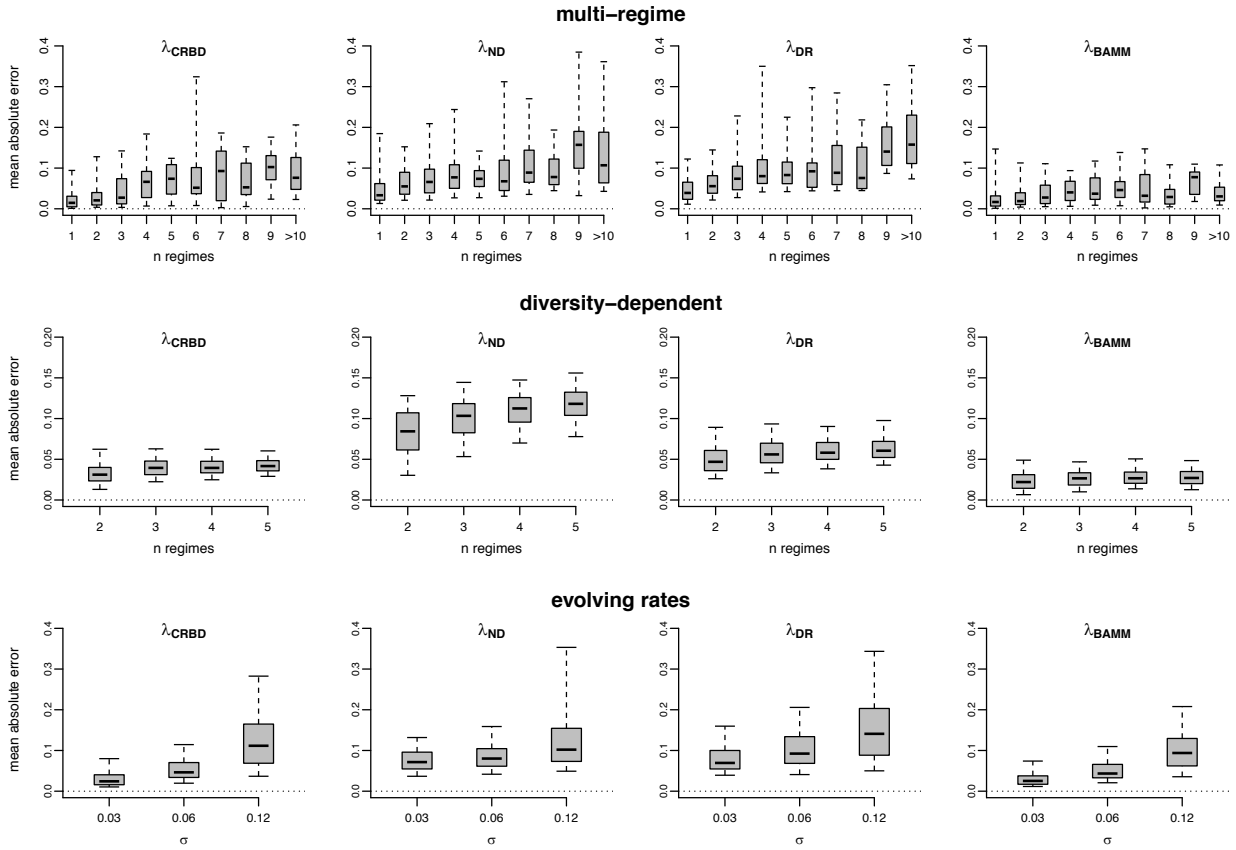


Figure 3. Mean per-tip absolute error in speciation rates as a function of the magnitude of rate heterogeneity in each simulated phylogeny. Results are presented separately for different categories of rate variation (Table 1); left column shows estimates from a constant-rate birth-death model for reference. The boxes and whiskers represent the 0.25 – 0.75, and the 0.05 – 0.95 quantile ranges, respectively. In some cases, λ_{ND} and λ_{DR} had more error than a simple CRBD model with no variation in tip rates. λ_{BAMM} had the least amount of error across all amounts of rate heterogeneity. See Figure S7 for λ_{TB}.

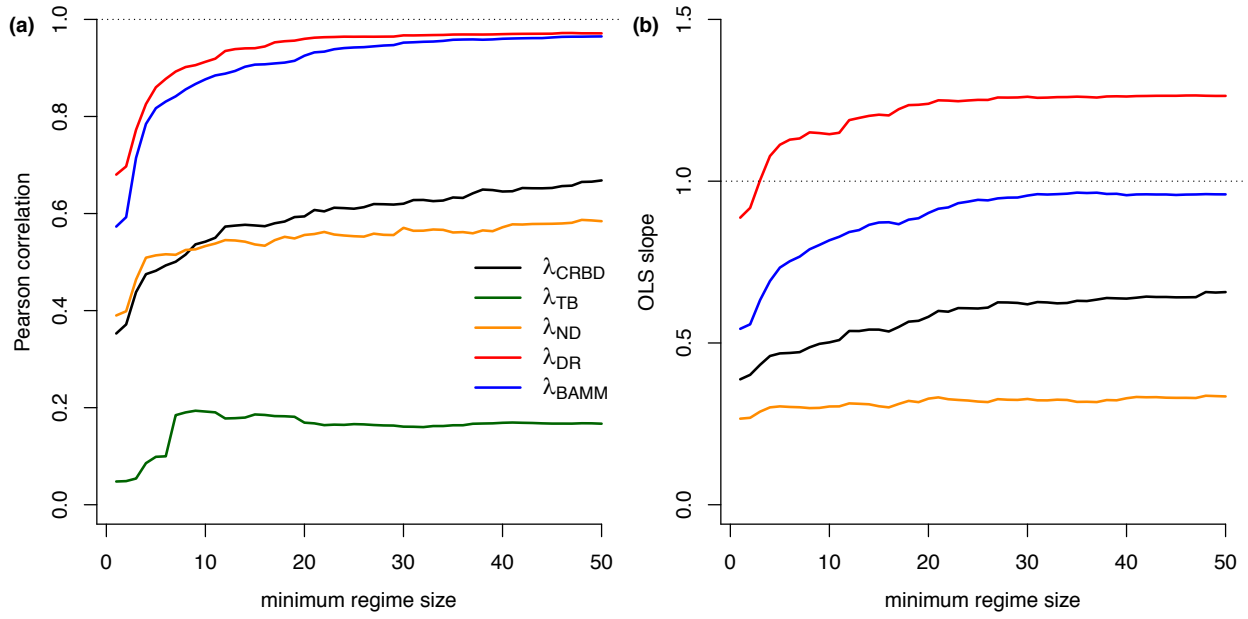


Figure 4. Performance of tip rate metrics as a function of regime size, including Pearson correlation (a) and OLS regression slope (b) for mean rates with respect to λ_{TRUE}. λ_{DR} and λ_{BAMM} outperform the other metrics when summarized in this fashion, although λ_{DR} tends to overestimate the rate of speciation. The x-axis denotes the minimum regime size across which performance was summarized. For example, x = 20 corresponds to the correlations and slopes computed for all regimes with 20 or more tips; a value of x = 1 is the corresponding results for all regimes. The OLS slope for λ_{TB} is not visible as it ranges between 15 and 23.

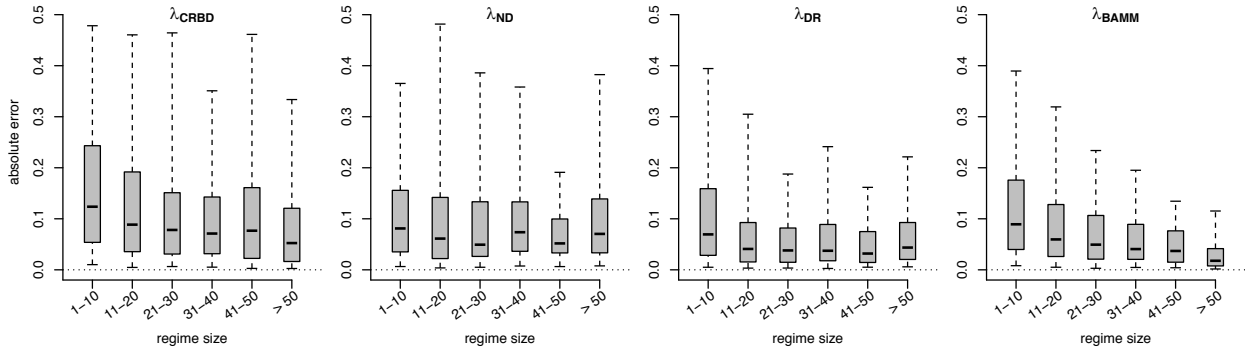
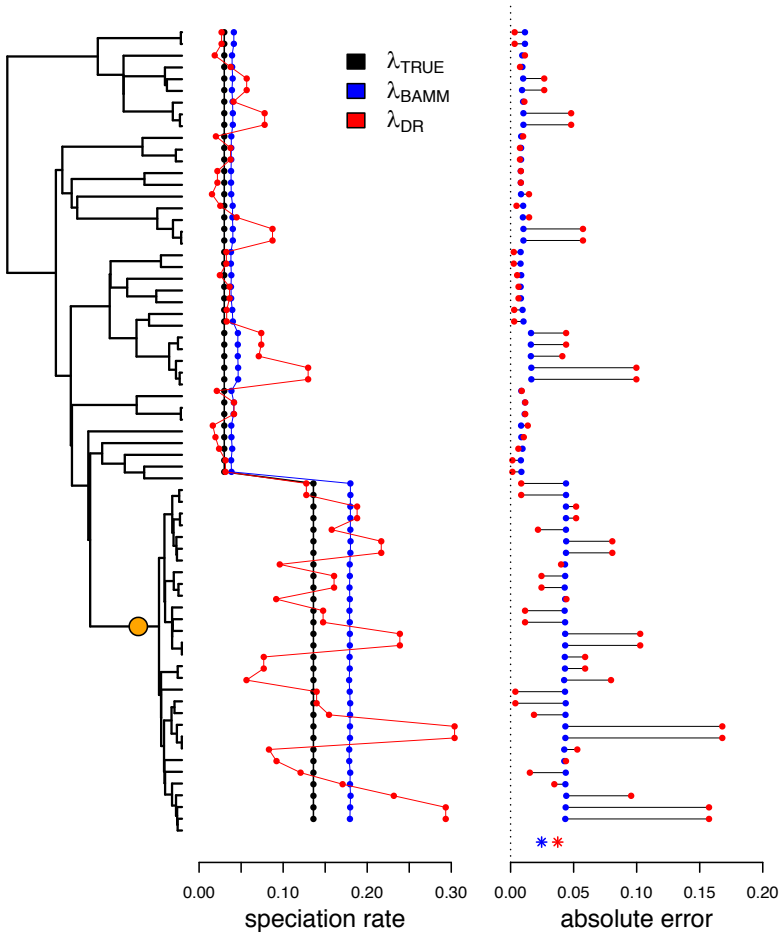


Figure 5. Mean per-regime absolute error in relation to true rate regime size, as binned into 10 size categories. The boxes and whiskers represent the 0.25 – 0.75, and the 0.05 – 0.95 quantile ranges, respectively. Perfectly estimated rates have an error of zero. λ_{DR} and λ_{BAMM} exhibit the least error when averaged by regimes, and λ_{DR} does slightly better for small clades (10-clade median error 0.07 for λ_{DR}, and 0.08 for λ_{BAMM}).



700

701 **Figure 6.** Relationship between λ_{TRUE}, λ_{BAMM}, and λ_{DR} for a simulated phylogeny containing a

702 single rate shift (orange circle). Subplots to the right of the tree illustrate true and estimated rates

703 for each tip (left) and corresponding absolute error (right). Asterisks at the bottom denote mean

704 per-tip error in tip rate metrics. Mean per-tip error is relatively low and similar between λ_{DR} and

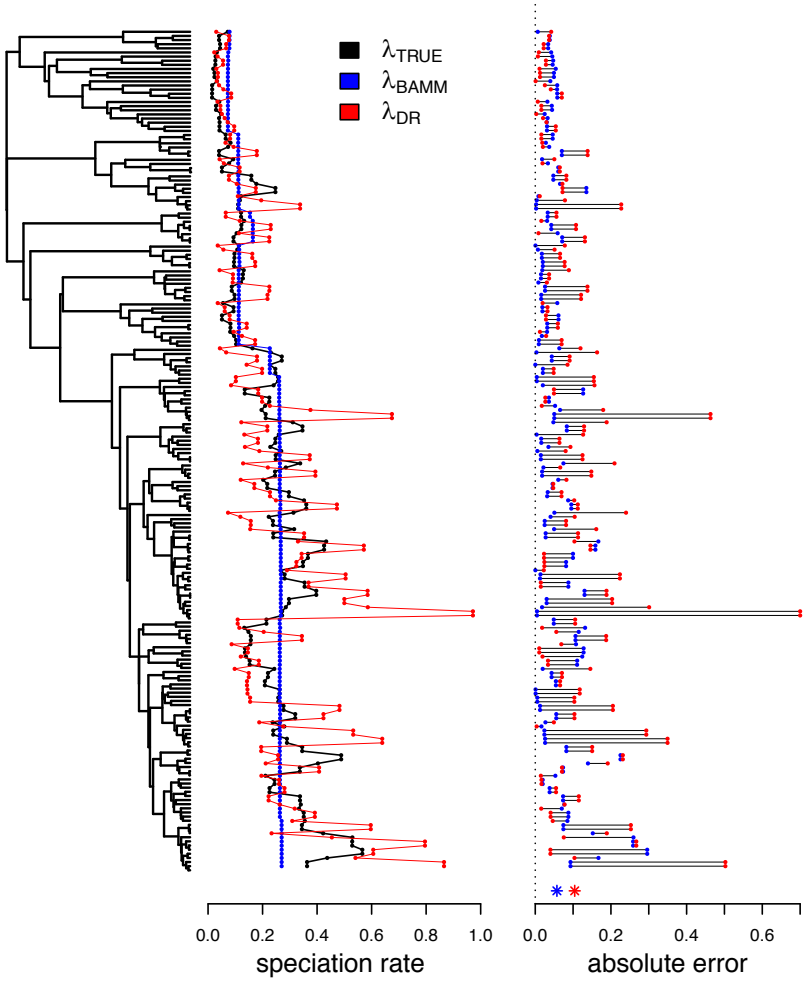
705 λ_{BAMM}, but the sample variance in λ_{DR} tip rates is high. In this example, the variance in absolute

706 per-tip error in λ_{DR} is 0.002 versus 0.0003 for λ_{BAMM}. On average, λ_{DR} tends to either

707 overestimate or underestimate rates relative to λ_{BAMM}, even if the mean per-tip error is relatively

708 low for both metrics.

709



710

711 **Figure 7.** Relationship between λ_{DR}, λ_{BAMM}, and λ_{DR} for a phylogeny simulated under an

712 "evolving rates" model, such that the speciation rate itself varies under a diffusion model. See

713 Figure 6 for additional details. Neither metric is particularly well equipped to infer the true rate

714 variation in this case. However λ_{BAMM}'s conservative estimates are still more accurate relative to

715 λ_{DR}, which is negatively impacted by high variance in tip rates. Here, variance in absolute per-tip

716 error in λ_{DR} is 0.012 versus 0.003 for λ_{BAMM}.

717

718

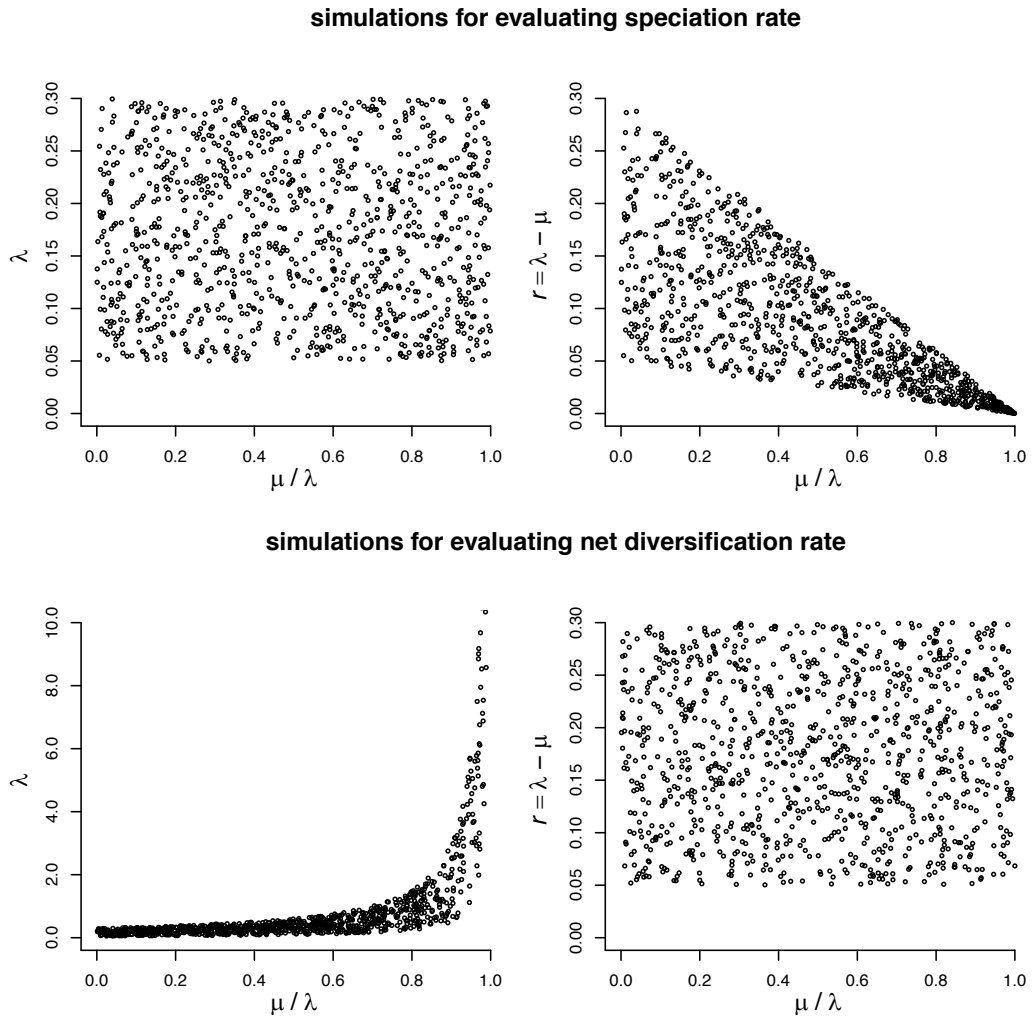


Figure S1. Details of the phylogeny simulations designed to evaluate the performance of the four tip metrics in terms of speciation rate and net diversification rate. From the top row, it is clear that when λ is sampled uniformly with respect to ε, the distribution of *r* is not uniform: the mean, range and variance in *r* decrease dramatically as ε increases. The reverse is true for the distribution of λ when *r* is sampled uniformly with respect to ε (bottom row). Our simulation design ensures that λ and *r* are sampled from identical uniform distributions with respect to ε and ensures comparability of the resulting error estimates.

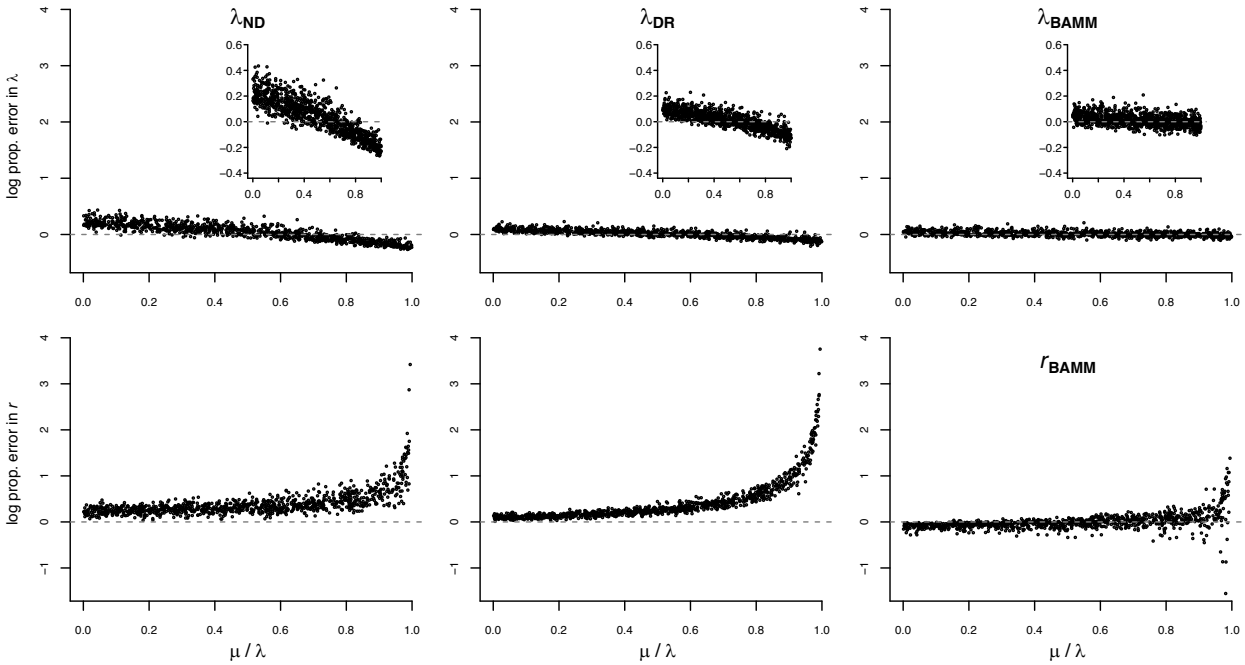


Figure S2. Log proportional accuracy in λ (top) and *r* (bottom) for different tip rate metrics, across a range of relative extinction rates. For BAMM, the estimated net diversification rate is presented. Proportional error of 0 implies perfect accuracy. Inset plots reveal greater detail in error for λ to ease metric comparison. All tip metrics track λ much more accurately than they track *r*, and λ_{BAMM} does so with the least amount of error. See Figure S3 for λ_{TB}.

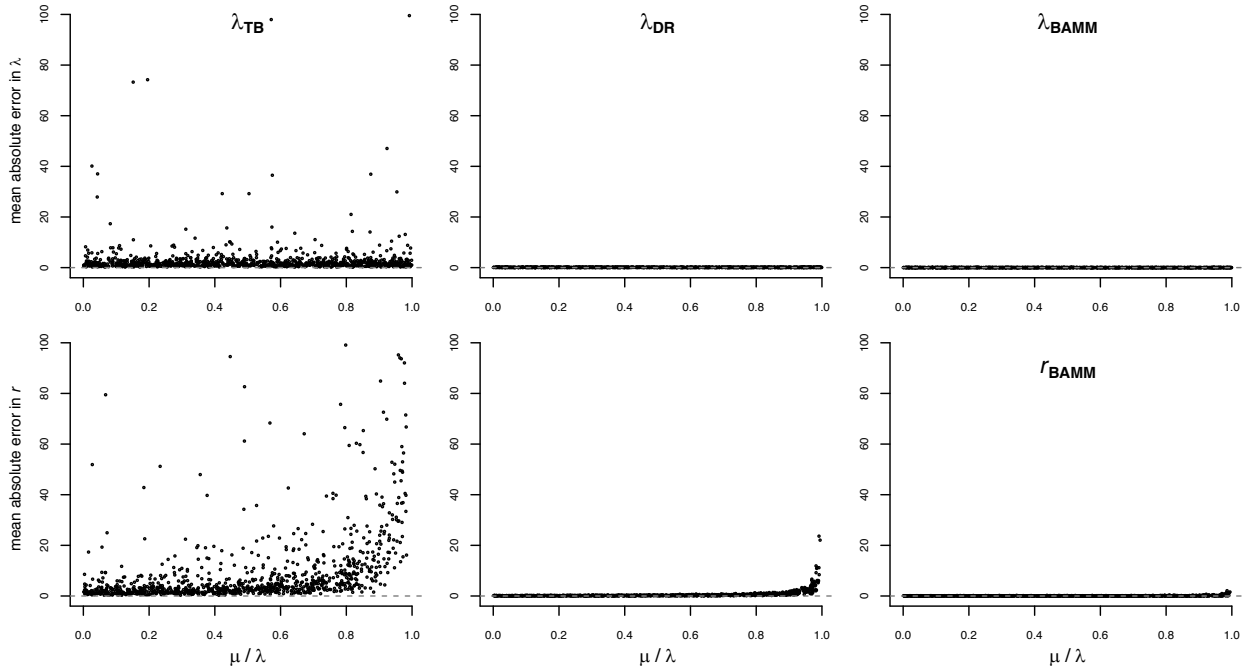


Figure S3. Mean absolute error in λ (top) and *r* (bottom) for λ_{TB}, with λ_{DR} and λ_{BAMM} on the same scale for comparison. For BAMM, the estimated net diversification rate is presented. λ_{TB} more accurately tracks λ and than *r*, but the amount of error is orders of magnitude greater than for other metrics.

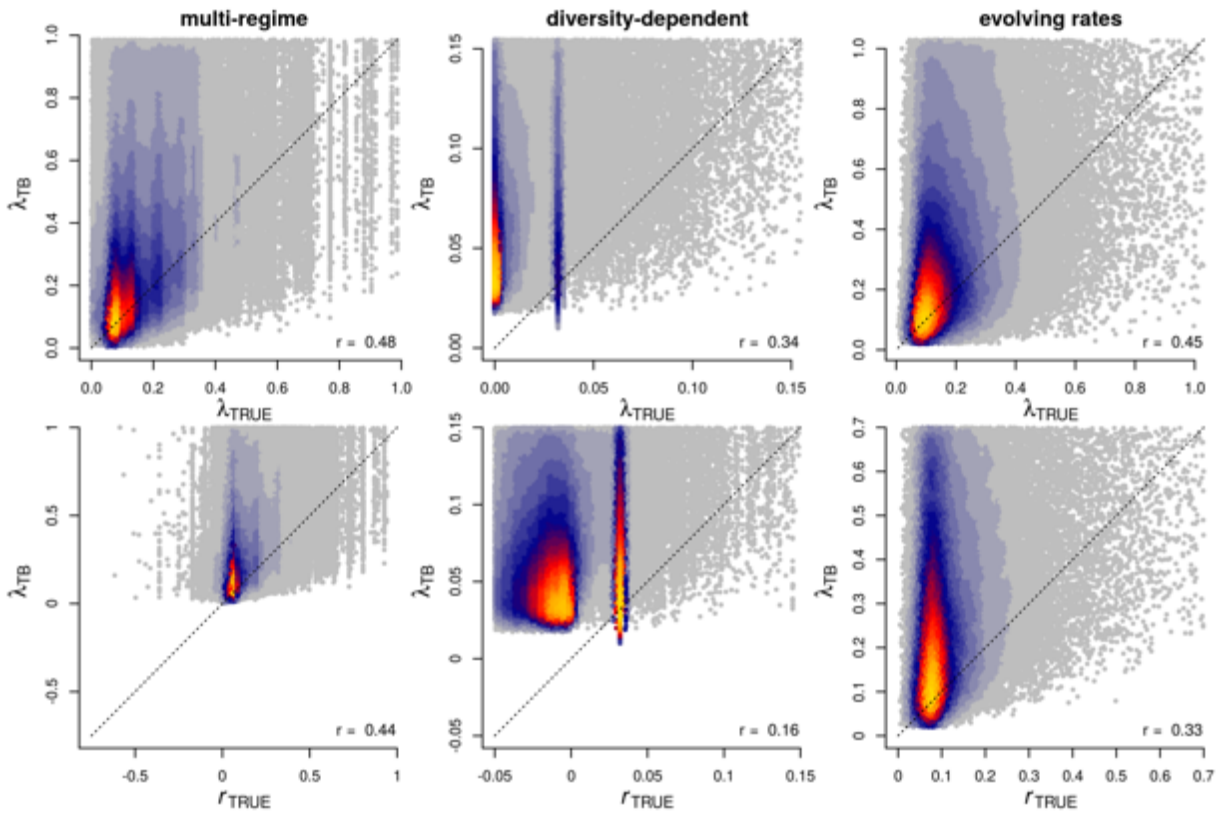


Figure S4. True tip rates (top row: λ_{TRUE}, bottom row: r_{TRUE}) in relation to λ_{TB}. Tip rates were compared separately for different major categories of phylogeny simulations (rows). Plotting region is restricted to the 99th percentile of true rates, but Spearman correlations between true and estimated rates (lower right of each figure panel) are based on the full range of the data. Colors indicate the density of points in the scatter plots. λ_{TB} is not particularly correlated with true tip rates.

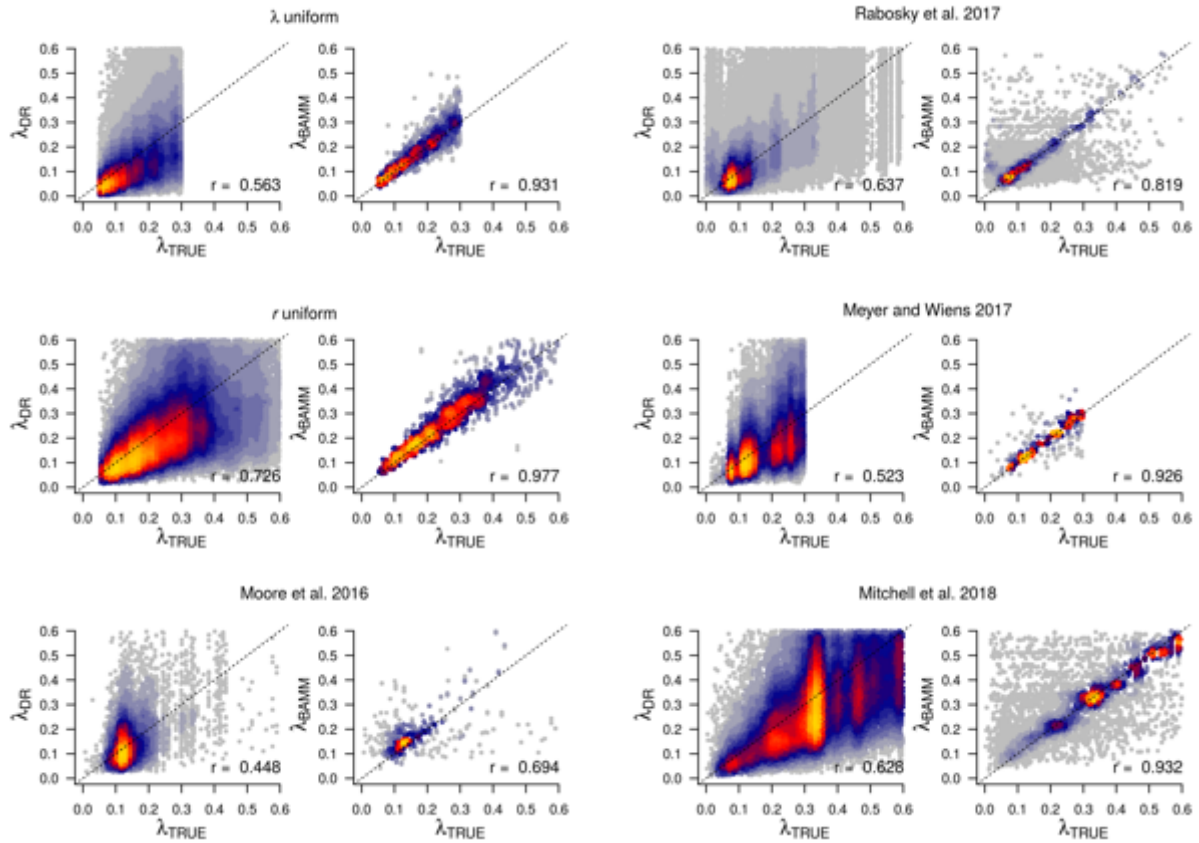
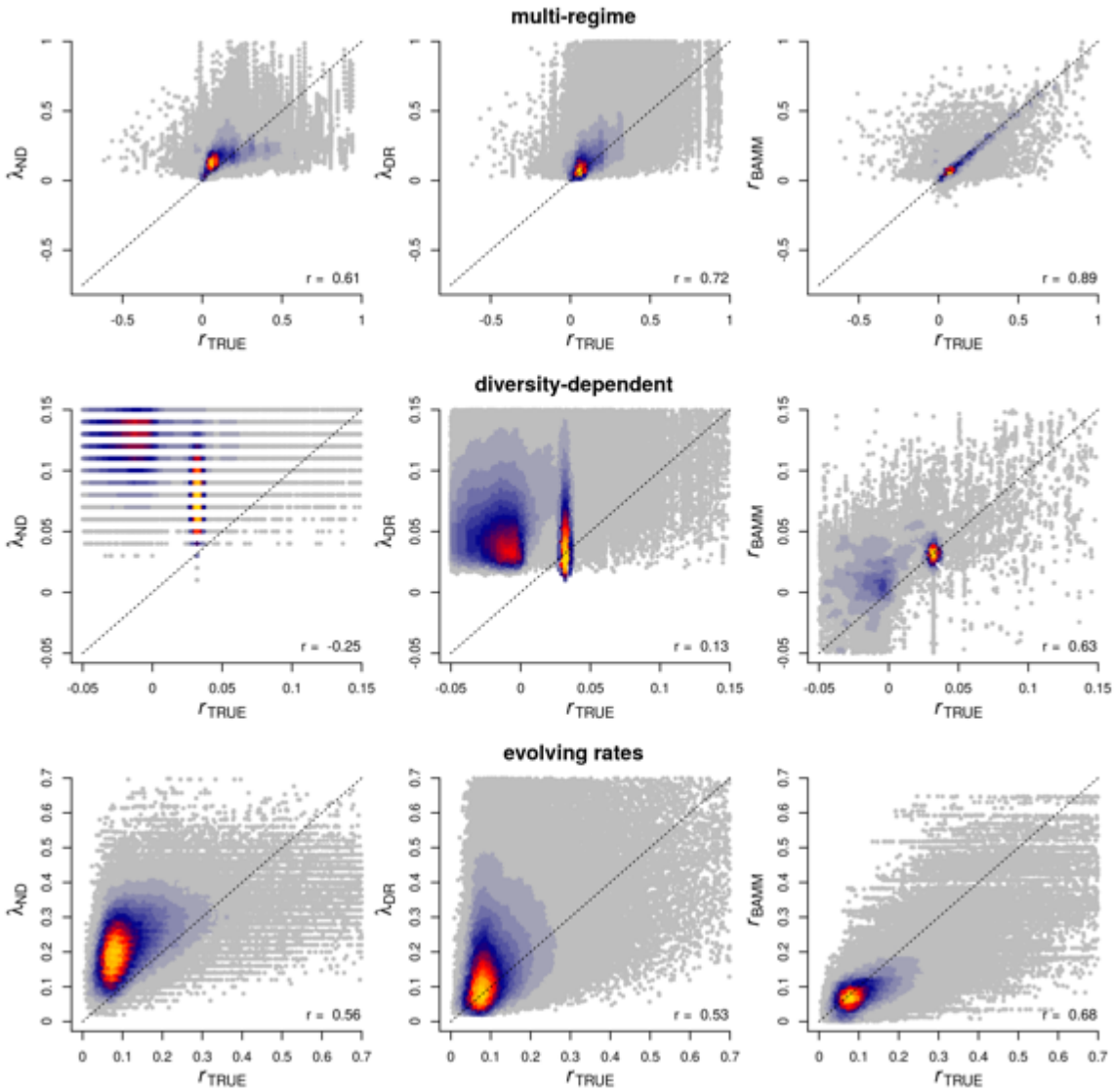
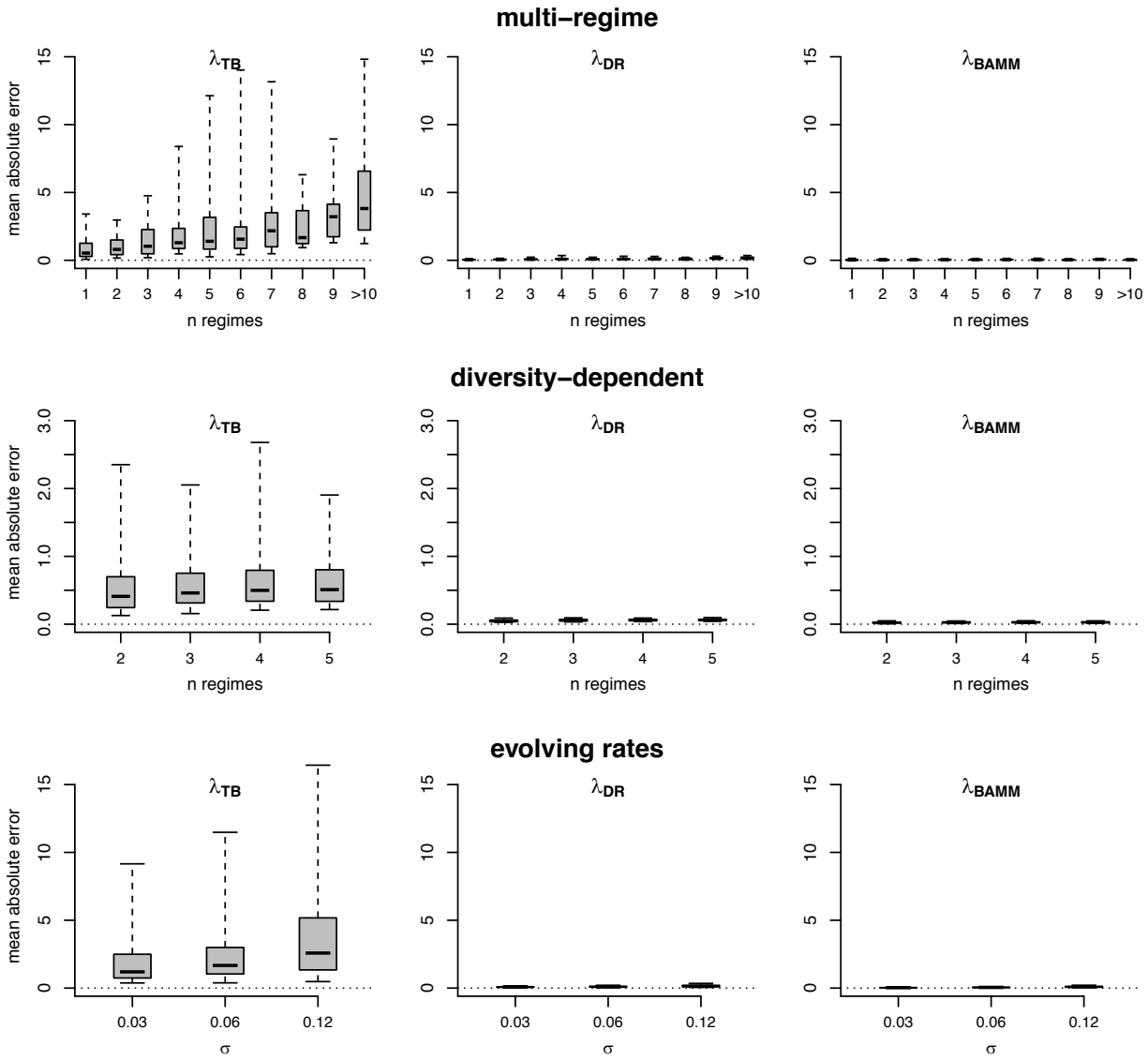


Figure S5. True tip rates (λ_{TRUE}) in relation to estimated tip rates from multi-regime phylogenies, as inferred from the best two tip rate metrics, λ_{DR} and λ_{BAMM}. Data are separated by source, to confirm that patterns described in the main text are not driven by any one simulation study. Spearman's correlation is presented in the bottom right corner. Colors indicate the density of points in the scatter plots. Regardless of the dataset, λ_{BAMM} performs noticeably better than λ_{DR}.



781

782 **Figure S6.** True net diversification tip rates (r_{TRUE}) in relation to estimated tip rates. Tip rates
783 were compared separately for different major categories of phylogeny simulations (rows).
784 Plotting region is restricted to the 99th percentile of true rates, but Spearman correlations
785 between true and estimated rates (lower right of each figure panel) are based on the full range of
786 the data. Colors indicate the density of points in the scatter plots. The horizontal gaps in λ_{ND} for
787 diversity-dependent trees are an artefact of all trees having the same crown age. Relative
788 performance comparison aside, correlations with r_{TRUE} are lower than with λ_{TRUE} (Figure 2).



789

790 **Figure S7.** Mean per-tip absolute error in λ_{TB} as a function of the magnitude of rate

791 heterogeneity in each simulated phylogeny. λ_{DR} and λ_{BAMM} are included on the same scale for

792 comparison. Results are presented separately for different categories of rate variation (Table 1).

793 The boxes and whiskers represent the 0.25 – 0.75, and the 0.05 – 0.95 quantile ranges,

794 respectively. Error in λ_{TB} generally increases with increasing rate heterogeneity, and this error is

795 almost two orders of magnitude greater than error in other tip rate metrics.

796

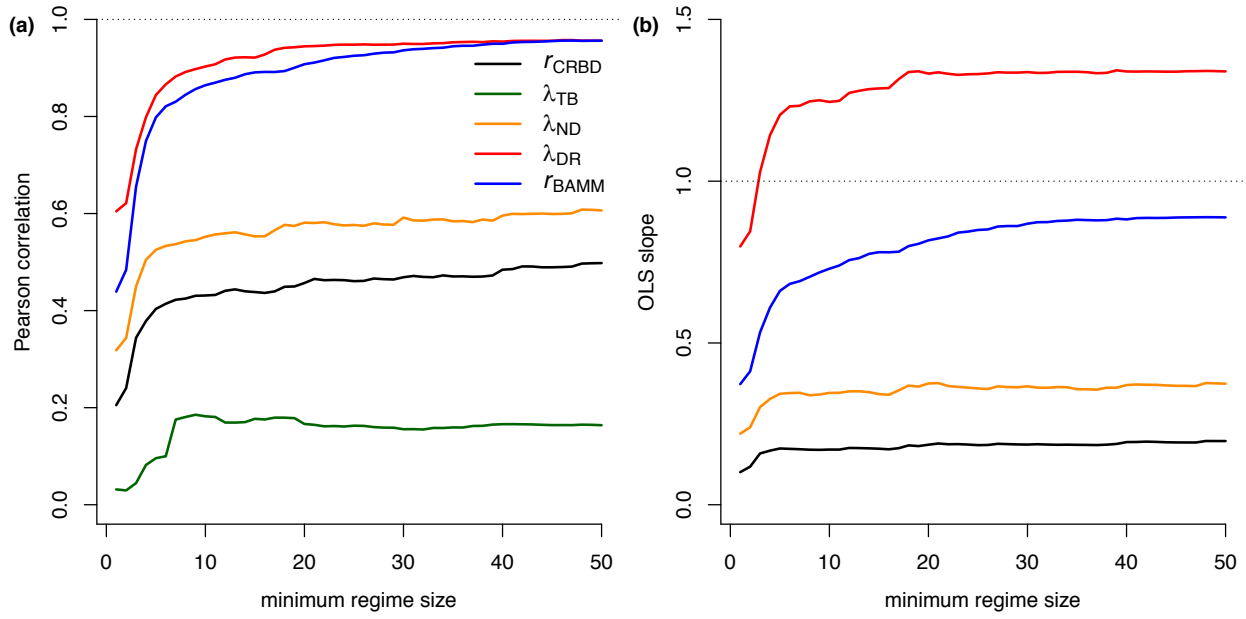


Figure S8. Performance of tip rate metrics as a function of regime size, including Pearson correlation (a) and OLS regression slope (b) for mean rates with respect to r_{TRUE}. λ_{DR} and r_{BAMM} outperform the other metrics when summarized in this fashion, although λ_{DR} overestimates the rate of net diversification (more so than it overestimated λ_{TRUE}, Figure 4). The x-axis denotes the minimum regime size across which performance was summarized. For example, x = 20 corresponds to the correlations and slopes computed for all regimes with 20 or more tips; a value of x = 1 is the corresponding results for all regimes. The OLS slope for λ_{TB} is not visible as it ranges between 10 and 25.