A flexible method for estimating tip diversification rates across a range of speciation and extinction scenarios

Supplementary Material 3

Simulation study

Our simulation study encompasses nine simulation scenarios retrieved from previous publications. Each of the simulation scenarios make different assumptions for the underlying model generating the tree:

- (1) "Speciation rate evolves via diffusion process" – trees in this set were simulated in a model where speciation rates evolve as a continuous trait under a Brownian motion model, where higher σ^2 leads to higher heterogeneity of rates in the tree; extinction-fraction is constant through time (Rabosky, 2010; Beaulieu and O'Meara, 2015; Rabosky, 2016l; Title and Rabosky, 2019).

- (2) "Single- and multi-regime, constant-rate birth–death" – trees in this set were simulated in a model assuming stochastic rate shifts. Shifts to new rate regimes are allowed in lineages that do not survive to the present, but trees are constrained to have at least one rate shift in the observed tree (Michell et al., 2019).

- (3) "Single-regime, constant-rate birth–death, lambda uniform" – trees in this set were simulated under a constant birth death-model across a range of extinction fractions (from pure birth to extinction-fraction = 1). Speciation rates for the generating model were sampled uniformly in respect to extinction-fraction (Title and Rabosky, 2019).

- (4) "Single-regime, constant-rate birth–death" – trees in this set were simulated under a constant birth death-model with no shifts to new rate regimes. Values for lambda were drawn from an exponential distribution with rate parameter of 1; values for extinction were drawn such that extinction fraction was uniformly distributed (Mitchell and Rabosky, 2017)

- (5) "Single- and multi-regime, constant-rate birth–death" – trees in this set were simulated under a model assuming stochastic rate shifts. Speciation and extinction rates for the generating model were sampled from the posterior distribution of parameters estimated from the "whales" tree using a Bayesian approach. New diversification parameters for rate shifts were drawn from an exponential distribution centered near the posterior distribution of parameters from the "whales" tree. Trees were simulated to have one or more diversification shifts and an approximate size as the "whales" tree (50 to 150 tips) (Moore et al., 2016)

- (6) "Single-regime, constant-rate birth–death, net diversification uniform" – similar to "3", but both extinction-fraction and net-diversification are sampled from uniform distributions (Title and Rabosky, 2019).

- (7) "Pure birth root regime, 1–4 discrete shifts to diversity-dependent regimes" – trees in this set were simulated under a pure birth model at the root. One to four regimes are allowed to happen randomly in a specific time interval. Parameters for the new rate regimes followed a diversification-dependent model, where speciation rates are modelled in function of the number of lineages in the subclade or the clade-specific carrying capacity (Rabosky, 2014).

- (8) "Single- and multi-regime, constant-rate birth–death" – trees in this set were simulated under a forward-time Poisson process with rate shifts. Shifts are allowed to occur in lineages that subsequently become extinct. Rates within regime are constant through time (Rabosky et al., 2017).

- (9) "Rate changes at every speciation event, constant extinction-fraction" – trees in this set were simulated under a model where diversification rates are inherited at speciation, but with a shift. Daughter diversification rates are sampled from a joint probability distribution parametrized by the paternal rates. Rates are branch specific and extinction-fraction is constant through time (Maliet et al., 2019)

Besides the range of simulation conditions used for the trees, an advantage of using simulated trees and previous results from Title and Rabosky (2019) is that priors and other settings for BAMM were chosen by the authors of BAMM, and thus are likely good approximations to what a skilled user would do. When comparing Bayesian models, either with each other or with non-Bayesian models, priors and MCMC runs may matter greatly in the accuracy of results: one could set priors that exclude the truth and find terrible performance or set priors at the true values and find excellent performance even in the absence of data. Title and Rabosky (2019) used priors set by BAMMTools (Rabosky et al., 2014) and thus this is a fair test with no special foreknowledge of the true parameters given to BAMM.

We used the same function to set priors for the BAMM runs on the set of 35 trees simulated under the model "rate changes at every speciation event, constant extinction-fraction" from Maliet et al. (2019). For the ClaDS runs, we used the default parameters of the data augmentation inference method implemented in julia (Maliet and Morlon, 2021; following the manual at https://hmorlon.github.io/PANDA.jl/dev/), which implements the model ClaDS2 (constant extinction-fraction; Maliet et al. 2019). All code necessary to replicate this study are available in: [repo to be added upon acceptance due to double-blind review].

Note that four trees had to be excluded from the comparisons due to convergence problems or limitations in one of the methods. Those correspond to: (1) Three trees in the "rate changes at every speciation event, constant extinction-fraction" for which either BAMM did not recognize as ultrametric and/or having non-negative branch lengths or for which MiSSE did not stop fitting different rate combinations after a c. two weeks run. And (2) One tree in the "single- and multi-regime, constant-rate birth–death" scenario for which ClaDS did not reach a gelman below 1.5 after a c. two weeks run.

Comparison of computing time among three model-based approaches

A reviewer pointed out that MiSSE can be slower than other model-based approaches because it has to integrate over many pairs of ordinary differential equations. MiSSE can, indeed, be slower than other methods (Table SM3.1) and new developments will probably be necessary to improve its computational speed (much like ClaDS started as a slower software than the current version, Maliet et al. 2019; Maliet and Morlon 2022).

Table SM3.1: Computing times for full runs of three model-based approaches on Lupinus (Fabaceae), a tree of c. 120 tips.

Method	Computing time
BAMM (Rabosky 2014)	1:39
ClaDS (Maliet and Morlon 2022)	3:12
MiSSE (this study)	1:33:18

References

Beaulieu, J. M., & O'Meara, B. C. (2015). Extinction can be estimated from moderately sized molecular phylogenies. *Evolution*, *69*(4), 1036–1043. https://doi.org/10.1111/evo.12614

Maliet, O., Hartig, F., & Morlon, H. (2019). A model with many small shifts for estimating species-specific diversification rates. *Nature ecology & evolution,* 3(7), 1086-1092.

Maliet, O., & Morlon, H. (2022). Fast and accurate estimation of species-specific diversification rates using data augmentation. *Systematic Biology*, 71(2), 353-366.

Mitchell, J. S., & Rabosky, D. L. (2017). Bayesian model selection with BAMM: effects of the model prior on the inferred number of diversification shifts. *Methods in Ecology and Evolution*, *8*(1), 37–46. https://doi.org/10.1111/2041-210X.12626

Mitchell, J. S., Etienne, R. S., & Rabosky, D. L. (2019). Inferring diversification rate variation from phylogenies with fossils. Systematic Biology, 68(1), 1–18. https://doi.org/10.1093/sysbio/syy035

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*(34), 9569–9574. https://doi.org/10.1073/pnas.1518659113

Rabosky, D. L. (2010). Extinction rates should not be estimated from molecular phylogenies. *Evolution*, *64*(6), 1816–1824. https://doi.org/10.1111/j.1558-5646.2009.00926.x

Rabosky, D. L. (2016). Challenges in the estimation of extinction from molecular phylogenies: a response to Beaulieu and O'Meara. *Evolution*, *70*(1), 218–228. https://doi.org/10.1111/evo.12820

Rabosky, D. L., Grundler, M., Anderson, C., Title, P., Shi, J. J., Brown, J. W., Huang, H., & Larson, J. G. (2014). BAMM tools: an R package for the analysis of evolutionary dynamics on phylogenetic trees. *Methods in Ecology and Evolution*, *5*(7), 701–707. https://doi.org/10.1111/2041-210X.12199

Rabosky, D. L. (2014). Automatic detection of key innovations, rate shifts, and diversity-dependence on phylogenetic trees. *PloS one*, *9*(2), e89543. https://doi.org/10.1371/journal.pone.0089543

Rabosky, D. L., Mitchell, J. S., & Chang, J. (2017). Is BAMM flawed? Theoretical and practical concerns in the analysis of multi-rate diversification models. *Systematic biology*, *66*(4), 477–498. https://doi.org/10.1093/sysbio/syx037

Title, P. O., & Rabosky, D. L. (2019). Tip rates, phylogenies, and diversification: what are we estimating, and how good are the estimates? *Methods in Ecology and Evolution*, *10*(6), 821–834. https://doi.org/10.1111/2041-210X.13153